Lecture Notes in Networks and Systems 396

Yu-Dong Zhang
Tomonobu Senjyu
Chakchai So-In
Amit Joshi *Editors*

Smart Trends in Computing and Communications

Proceedings of SmartCom 2022

Springer

Lecture Notes in Networks and Systems

Volume 396

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Advisory Editors

Fernando Gomide, Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Turkey

Derong Liu, Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA
Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

More information about this series at https://link.springer.com/bookseries/15179

Yu-Dong Zhang · Tomonobu Senjyu ·
Chakchai So-In · Amit Joshi
Editors

Smart Trends in Computing and Communications

Proceedings of SmartCom 2022

Springer

Editors
Yu-Dong Zhang
University of Leicester
Leicester, UK

Chakchai So-In
Department of Computer Science
Khon Kaen University
Khon Kaen, Thailand

Tomonobu Senjyu
Faculty of Engineering
University of the Ryukyus
Nishihara, Okinawa, Japan

Amit Joshi
Global Knowledge Research Foundation
Ahmedabad, Gujarat, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-16-9966-5 ISBN 978-981-16-9967-2 (eBook)
https://doi.org/10.1007/978-981-16-9967-2

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023, corrected publication 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The Sixth Edition of the SmartCom 2022—Smart Trends in Computing and Communications was held during 17 and 18 December 2021 physically at Four Points by Sheraton Jaipur, India, and digitally on Zoom which is organised by Global Knowledge Research Foundation. The associated partners were Springer, KCCI and InterYIT IFIP. The conference will provide a useful and wide platform both for display of the latest research and for exchange of research results and thoughts. The participants of the conference will be from almost every part of the world, with background of either academia or industry, allowing a real multinational multicultural exchange of experiences and ideas.

A great pool of more than 480 papers were received papers for this conference from across seven countries among which around 75 papers were accepted with this Springer Series and were presented through digital platform during the two days. Due to overwhelming response, we had to drop many papers in hierarchy of the quality. A total of 12 technical sessions were organised in parallel in 2 days along with few keynotes and panel discussions. The conference will be involved in deep discussion and issues which will be intended to solve at global levels. New technologies will be proposed, experiences will be shared, and future solutions for enhancement in systems and security will also be discussed. The final papers will be published in proceedings by Springer LNNS Series.

Over the years, this conference has been organised and conceptualised with collective efforts of a large number of individuals. I would like to thank each of the committee members and the reviewers for their excellent work in reviewing the papers. Grateful acknowledgements are extended to the team of Global Knowledge Research Foundation for their valuable efforts and support.

I look forward to welcome you on the 7th Edition of this SmartCom Conference in 2023.

Leicester, UK	Yu-Dong Zhang
Nishihara, Japan	Tomonobu Senjyu
Khon Kaen, Thailand	Chakchai So-In
Ahmedabad, India	Amit Joshi

Contents

Implementation of Tweet Stream Summarization for Malicious Tweet Detection 1
Pranoti Mane, Shrikant Sonekar, and Sayema Kausar

Comparative Analysis of a Secure Authentication Protocol for 5G-Enabled IoT Network Using Public and Shared Secret Key 11
Shamshuddin K and Vishal B. Pattanashetty

DDoS Botnet Attack Detection in IoT Devices 21
Bandari Pranay Kumar, Gautham Rampalli, Pille Kamakshi, and T. Senthil Murugan

Text/Sign Board Reading Aid for Visually Challenged People 29
Srikhakollu Pandu Ranga Avinash, V. Krishna Sree, M. Saahith Reddy, K. Sravani Kumari, and A. Gaurav

Automatic Segmentation of Red Blood Cells from Microscopic Blood Smear Images Using Image Processing Techniques 39
K. T. Navya, Subhraneil Das, and Keerthana Prasad

Design and Simulation of Microstrip Bandpass Filter Techniques at X-Band for RADAR Applications 51
Shraddha S. Shinde, D. P. Rathod, and ArunKumar Heddallikar

A Novel Approach to Detect Plant Disease Using DenseNet-121 Neural Network 63
Nilesh Dubey, Esha Bhagat, Swapnil Rana, and Kirtan Pathak

Offline Handwritten Signature Forgery Verification Using Deep Learning Methods 75
Phan Duy Hung, Pham Son Bach, Bui Trong Vinh, Nguyen Huu Tien, and Vu Thu Diep

Potential Applications of Advanced Control System Strategies in a Process industry—A Review 85
Pradeep Kumar Juneja, Sandeep Kumar Sunori, Kavita Ajay Joshi, Shweta Arora, Somesh Sharma, Prakash Garia, Sudhanshu Maurya, and Amit Mittal

Analysis of Algorithms for Effective Cryptography for Enhancement of IoT Security 91
Valerie David, Harini Ragu, Vemu Nikhil, and P. Sasikumar

Relevance of Artificial Intelligence in the Hospitality and Tourism Industry 101
Smrutirekha, Priti Ranjan Sahoo, and Ravi Shankar Jha

Deep Learning-Based Smart Surveillance System 111
G. Sreenivasulu, N. Thulasi Chitra, S. Viswanadha Raju, and Venu Madhav Kuthadi

Pronunciation-Based Language Processing Tool for Tai Ahom 125
Dhrubajyoti Baruah, Joydip Sarmah, Kamal Gayan, and Satya Ranjan Phukon

Improving Ecommerce Performance by Dynamically Predicting the Purchased Items Using FUP Incremental Algorithm 137
K. Kalaiselvi, K. Deepa Thilak, S. Saranya, T. Rajeshkumar, M. Malathi, M. Vijay Anand, and K. Kumaresan

A Novel Approach to Privacy Preservation on E-Healthcare Data in a Cloud Environment 149
Kirtirajsinh Zala and Madhu Shukla

Temporal Reasoning of English Text Documents 159
Parul Patel

Design, Development, and Integration of DMA Controller for Open-Power-Based Processor SOC 169
Gannera Mamatha, M. N. Giriprasad, and Sharan Kumar

A Comparative Study Using Numerical Simulations for Designing an Effective Cooling System for Electric Vehicle Batteries 179
Akshat Maheshwari, Vineet Singhal, Ajay Kumar Sood, and Meenakshi Agarwal

Lexical Resource Creation and Evaluation: Sentiment Analysis in Marathi 187
Mahesh B. Shelke, Saleh Nagi Alsubari, D. S. Panchal, and Sachin N. Deshmukh

Artificial Neural Network Method for Appraising the Nephrotic Disease 197
K. Padmavathi, A. V. Senthilkumar, Ismail Bin Musirin, and Bınod Kumar

Artificial Neural Network and Math Behind It 205
Harshini Pothina and K. V. Nagaraja

Machine Learning-Based Multi-temporal Image Classification Using Object-Based Image Analysis and Supervised Classification 223
Swasti Patel, Priya Swaminarayan, Simranjitsingh Pabla, Mandeepsingh Mandla, and Hardik Narendra

A Model to Generate Benchmark Network with Community Structure 235
Shyam Sundar Meena and Vrinda Tokekar

Human Activity Recognition Using LSTM with Feature Extraction Through CNN 245
Rosepreet Kaur Bhogal and V. Devendran

Sentiment Classification of Higher Education Reviews to Analyze Students' Engagement and Psychology Interventions Using Deep Learning Techniques 257
K. R. Sowmia, S. Poonkuzhali, and J. Jeyalakshmi

Strategic Network Model for Real-Time Video Streaming and Interactive Applications in IoT-MANET 267
Atrayee Gupta, Bidisha Banerjee, Sriyanjana Adhikary, Himadri Shekhar Roy, Sarmistha Neogy, Nandini Mukherjee, and Samiran Chattyopadhyay

Limitation for Single-equation Dependency 277
Michael Shell, John Doe, and Jane Doe

Autocorrelation of an Econometric Model 289
Preeti Singh and Sarvpal Singh

Trends of Artificial Intelligence in Revenue Management of Hotels 299
Smrutirekha, Priti Ranjan Sahoo, and Ashtha Karki

Sentiment Analysis and Vector Embedding: A Comparative Study 311
Shila Jawale and S. D. Sawarkar

3G Cellular Network Fault Prediction Using LSTM-Conv1D Model 323
N. Geethu and M. Rajesh

Deep Learning for Part of Speech (PoS) Tagging: Konkani 337
Annie Rajan, Ambuja Salgaonkar, and Arshad Shaikh

***DAMBNFT*: Document Authentication Model through Blockchain and Non-fungible Tokens** 347
Uday Khokhariya, Kaushal Shah, Nidhay Pancholi, and Shambhavi Kumar

A Novel Hybrid Translator for Gujarati to Interlingual English MTS for Personage Idioms 355
Jatin C. Modh and Jatinderkumar R. Saini

Alleviation of Voltage Quality-Related Issues: A Case Study of Bahir Dar Textile Share Distribution System 369
Dessalegn Bitew Aeggegn, Yalew Werkie Gebru, Takele Ferede Agajie, Ayodeji Olalekan Salau, and Adedeji Tomide Akindadelo

Customer Perception, Expectation, and Experience Toward Services Provided by Public Sector Mutual Funds 381
R. Malavika and A. Suresh

DigiDrive: Making Driving School Management Effective 393
Manasi Khanvilkar, Atreya Rastradhipati, and Wricha Mishra

Revamping an E-Application for User Experience: A Case Study of eSanjeevaniOPD App 405
Remya Vivek Menon and G. Rejikumar

Secured Data Transmission in Low Power WSN with LoRA 417
Arabinda Rath, S. Q. Baig, Bisakha Biswal, and Gayatri Devi

Safeguarding Cloud Services Sustainability by Dynamic Virtual Machine Migration with Re-allocation Oriented Algorithmic Approach 425
Saumitra Vatsal and Shalini Agarwal

Survey on Cloud Auditing by Using Integrity Checking Algorithm and Key Validation Mechanism 437
M. Mageshwari and R. Naresh

Privacy-Preserving Min–Max Value Retrieval Using Data Obfuscation Algorithm in Distributed Environment 447
Bhople Yogesh Jagannath

Intelligent Virtual Research Environment for Natural Language Processing (IvrE-NLP) 453
Prashant Chaudhary, Pavan Kurariya, Shashi Pal Singh, Jahnavi Bodhankar, Lenali Singh, and Ajai Kumar

Analyzing the Performance of Object Detection and Tracking Techniques 467
Suchita A. Chavan, Nandini M. Chaudhari, and Rakesh J. Ramteke

A Novel Area Efficient GF(2^m) Multiplier 475
Asfak Ali, Ram Sarkar, and Debesh Kumar Das

Virtual Testbed of Vehicular Network for Collision Simulation and Detection on SUMO and OMNeT++ Simulators 485
Tatini Shanmuki Krishna Ratnam and C. Lakshmikanthan

Artificial Intelligence-Based IoT Clustered Smart Vending Machine 497
Ravivanshikumar Sangpal, Suhas Khot, Pratiksha Vallapure, Rajkumar Mali, Rasika Kumbhar, and Aparna Hambarde

Artificial Intelligence-based Vehicle In-Cabin Occupant Detection and Classification System 509
M. Tamizharasan and C. Lakshmikanthan

A Hybrid Approach Towards Machine Translation System for English–Hindi and Vice Versa 523
Shraddha Amit Kalele, Shashi Pal Singh, Prashant Chaudhary, Lenali Singh, Ajai Kumar, and Pulkit Joshi

Gender Issues and New Governance: Managing Menstrual Health and Hygiene with Emerging Multimedia Options for a Progressive Economy 533
Suparna Dutta and Ankita Das

An IoT-Based Smart Parking System for Smart Cities 543
Venkatesh Mane, Ashwin R. Kubasadgoudar, Raghavendra Shet, and Nalini C. Iyer

Agriculture Stakeholder's Information Need—The Survey in Gujarat, India 551
Axita Shah and Jyoti Pareek

System Design and Implementation of Assistive Device for Hearing Impaired People 565
P. Nikita, B. H. Shraddha, Venkatesh Mane, N. Shashidhar, P. Vishal, and Nalini C. Iyer

Internet of Things Security: A Blockchain Perspective 577
Mohammad Luqman and Arman Rasool Faridi

Fuzzy C-Means Clustering of Network for Multi Mobile Agent Itinerary Planning 589
Nidhi and Shuchita Upadhyaya

Acceptance of Eco-Friendly Substitute of White Skim Coat (Wall Putty) in Mumbai 599
Mandar Anil Chitre, Shivoham Singh, and Manuj Joshi

Economics of Immutability Preserving Streaming Healthcare Data Storage Using Aggregation in Blockchain Technology 611
Sachin Gupta and Babita Yadav

A Comparative Cost Analysis of Organizational Network Security Test Lab Setup on Cloud Versus Dedicated Virtual Machine 623
Sachin Gupta, Bhoomi Gupta, and Atul Rana

Insights into the Black Box Machine Learning Models Through Explainability and Interpretability 633
Sachin Gupta and Bhoomi Gupta

An Improved AODV Routing Algorithm for Detection of Wormhole and Sybil Attacks in MANET 645
Shreeya Mishra, Umesh Kumar, and Komal Mehta Bhagat

An Ensemble Model (Simple Average) for Malaria Cases in North India 655
Kumar Shashvat, Arshpreet Kaur, Ranjan, and Vartika

IoT-Based Smart City Management Using IAS 665
Pranali Nitnaware, Supriya Sawwashere, Shrikant V. Sonekar, M. M. Baig, and Snehal Tembhurne

Big Data Analytics and Machine Learning Approach for Smart Agriculture System Using Edge Computing 675
U. Sakthi, K. Thangaraj, T. Poongothai, and M. K. Kirubakaran

Real-Time Object Detection System with Voice Feedback for the Blind People 683
Harshal Shah, Meet Amin, Krish Dadwani, Nishant Desai, and Aliasgar Chatiwala

Bird Video Summarization Using Pre-trained CNN Model and Deep Learning 691
Rachit Adhvaryu, Viral Parekh, and Dipesh Kamdar

Generation of Seismic Fragility Curves for RC Highways Vulnerable to Earthquake-Induced Landslides Based on ICT 701
Aadityan Sridharan and Sundararaman Gopalan

Influence of Student–Teacher Interrelationship on Academic Achievements: A Logit Model 711
Monika Saini, Asha Choudhary, and Ashish Kumar

Smart Water Metering Implementation 721
Urja Mankad and Harshal Arolkar

Artificial Bee Colony Optimization-Based Load Balancing in Distributed Computing Systems—A Survey 733
Vidya S. Handur and Santosh L. Deshpande

Distributed and Scalable Healthcare Data Storage Using Blockchain and KNN Classification 741
Manjula K. Pawar, Prakashgoud Patil, and Amit Singh Patel

A Sustainable Green Approach to the Virtualized Environment in Cloud Computing 751
Anjani Gupta, Prashant Singh, Dhyanendra Jain, Anupam Kumar Sharma, Prashant Vats, and Ved Prakash Sharma

Threats and Challenges of Artificial Intelligence in the Healthcare Industry 761
Priti Ranjan Sahoo, Smrutirekha, and Mou Chatterjee

Exploiting Tacit Knowledge: A Review and Possible Research Directions 771
Pawankumar Saini and Pradnya Chitrao

An Efficient Approach to Stamp Verification 781
Ha Long Duy, Ha Minh Nghia, Bui Trong Vinh, and Phan Duy Hung

Fragmented Central Affinity Approach for Reducing Ambiguities in Dataset 791
Tanvi Trivedi and Mahipal Singh Deora

Correction to: Human Activity Recognition Using LSTM with Feature Extraction Through CNN C1
Rosepreet Kaur Bhogal and V. Devendran

Author Index 797

Editors and Contributors

About the Editors

Prof. Yu-Dong Zhang worked as a postdoc from 2010 to 2012 with Columbia University, USA, and as an Assistant Research Scientist from 2012 to 2013 with Research Foundation of Mental Hygiene (RFMH), USA. He served as a Full Professor from 2013 to 2017 with Nanjing Normal University. Now, he serves as Professor with School of Computing and Mathematical Sciences, University of Leicester, UK. His research interests include deep learning and medical image analysis. He is Fellow of IET (FIET) and Senior Members of IEEE, IES, and ACM. He was the 2019 recipient of "Web of Science Highly Cited Researcher." He is the author of over 300 peer-reviewed articles, including more than 50 ESI Highly Cited Papers, and four ESI Hot Papers. His citation reached 16626 in Google Scholar (h-index 72) and 9878 in Web of Science (h-index 55). He has conducted many successful industrial projects and academic grants from NIH, Royal Society, GCRF, EPSRC, MRC, British Council, and NSFC.

Tomonobu Senjyu received his B.S. and M.S. degrees in electrical engineering from the University of the Ryukyus in 1986 and 1988, and Ph.D. degree in electrical engineering from Nagoya University in 1994. Since 1988, he has been with the Department of Electrical and Electronics Engineering, University of the Ryukyus, where he is currently a professor. His research interests are in the areas of power system optimization and operation, advanced control, renewable energy, IoT for energy management, ZEH/ZEB, smart city, and power electronics. He is a member of the Institute of Electrical Engineers of Japan and Senior member of IEEE.

Dr. Chakchai So-In [SM, IEEE (14); SM, ACM (15)] is Professor of Computer Science at the Department of Computer Science at Khon Kaen University. He received B.Eng. and M.Eng. from KU (TH), in 1999 and 2001, respectively; and M.S. and Ph.D. from WUSTL (MO, USA) in 2006 and 2010; all in computer engineering. In 2003, Dr. So-In was Internet Working Trainee in a CNAP program

at NTU (SG) and obtained Cisco/Microsoft Certifications. In 2006 to 2010, Dr. So-In was an intern at mobile IP division, Cisco Systems, WiMAX Forum, and Bell Labs (USA). His research interests include mobile computing/sensor networks, Internet of things, computer/wireless/distributed networks, cybersecurity, and intelligent systems and future Internet. He is/was on Editor/Guest members in *IEEE Access*, *PLOS One*, *PeerJ (CS)*, *Wireless and Mobile Computing*, and *ECTI-CIT*. Dr. So-In has authored/co-authored over 100 publications and 10 books including *IEEE JSAC*, *IEEE Communications/Wireless Communications Magazine*, *Computer Networks*; and *(Advanced) Android Application Development*, *Windows Phone Application Development*, *Computer Network Lab*, *Network Security Lab*. He has served as a committee member and reviewer for many prestigious conferences and journals such as ICNP, WCNC, Globecom, ICC, ICNC, PIMRC; and IEEE Transactions (Wireless Communications, Computers, Vehicular Technology, Mobile Computing, Industrial Informatics); IEEE Communications Magazine, Letters, Systems Journal, Access; and Computer Communications, Computer Networks, Mobile and Network Applications, Wireless Network.

Dr. Amit Joshi is currently Director of Global Knowledge Research Foundation and also Entrepreneur and Researcher who has completed his masters and research in the areas of cloud computing and cryptography in medical imaging. Dr. Joshi has an experience of around ten years in academic and industry in prestigious organizations. Dr. Joshi is an active member of ACM, IEEE, CSI, AMIE, IACSIT-Singapore, IDES, ACEEE, NPA, and many other professional societies. Currently, Dr. Joshi is International Chair of InterYIT at International Federation of Information Processing (IFIP, Austria). He has presented and published more than 50 papers in national and international journals/conferences of IEEE and ACM. Dr. Joshi has also edited more than 40 books which are published by Springer, ACM, and other reputed publishers. Dr. Joshi has also organized more than 50 national and international conferences and programs in association with ACM, Springer, IEEE to name a few across different countries including India, UK, Europe, USA, Canada, Thailand, Vietnam, Egypt, and many more.

Contributors

Adhikary Sriyanjana Department of Information Technology, Jadavpur University, Jadavpur, India

Adhvaryu Rachit C. U. Shah University, Surendranagar, Gujarat, India

Aeggegn Dessalegn Bitew Department of Electrical and Computer Engineering, Debre Markos University, Debre Markos, Ethiopia

Agajie Takele Ferede Department of Electrical and Computer Engineering, Debre Markos University, Debre Markos, Ethiopia

Agarwal Meenakshi BML Munjal University, Gurugram, Haryana, India

Agarwal Shalini Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India

Akindadelo Adedeji Tomide Department of Basic Sciences, Babcock University, Ilishan Remo, Nigeria

Ali Asfak ETCE Department, Jadavpur University, Kolkatta, West Bengal, India

Alsubari Saleh Nagi Department of Computer Science & IT, Dr. Babasaheb Ambedkar, Marathwada University, Aurangabad, M.S., India

Amin Meet Parul University, Vadodara, Gujarat, India

Arolkar Harshal GLS University, Ahmedabad, India

Arora Shweta Graphic Era Hill University, Bhimtal Campus, India

Bach Pham Son FPT University, Hanoi, Vietnam

Baig M. M. J.D. College of Engineering and Management, Nagpur, Maharashtra, India

Baig S. Q. Department of Computer Science and Engineering, ABIT, Cuttack, Odisha, India

Banerjee Bidisha Department of Computer Science and Engineering, Jadavpur University, Jadavpur, India

Baruah Dhrubajyoti Department of Computer Application, Jorhat Engineering College, Jorhat, Assam, India

Bhagat Esha Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research, Faculty of Technology Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, India

Bhagat Komal Mehta Bhagwan Parshuram Institute of Technology, Delhi, India

Bhogal Rosepreet Kaur School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, India

Bin Musirin Ismail Faculty of Electrical Engineering, Centre of Electrical Power Engineering, Universiti Teknologi, Skudai, Malaysia

Biswal Bisakha Department of Computer Science and Engineering, ABIT, Cuttack, Odisha, India

Bodhankar Jahnavi Centre for Development of Advanced Computing, Pune, India

Chatiwala Aliasgar Parul University, Vadodara, Gujarat, India

Chatterjee Mou KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India

Chattyopadhyay Samiran Department of Information Technology, Jadavpur University, Jadavpur, India

Chaudhari Nandini M. KSET, DRS Kiran and Pallavi Patel Global University, Vadodara, Gujarat, India

Chaudhary Prashant AAI, Centre for Development of Advanced Computing, Pune, India

Chavan Suchita A. KBC North Maharashtra University, Jalgaon, Maharashtra, India

Chitrao Pradnya Symbiosis Institute of Management Studies, A Constituent of Symbiosis International University, Pune, Maharashtra, India

Chitre Mandar Anil Pacific Academy of Higher Education and Research Society, Udaipur, Rajasthan, India

Choudhary Asha Department of Mathematics and Statistics, Manipal University Jaipur, Jaipur, India

Dadwani Krish Parul University, Vadodara, Gujarat, India

Das Ankita Birla Institute of Technology, Noida, India

Das Debesh Kumar CSE Department, Jadavpur University, Kolkatta, West Bengal, India

Das Subhraneil Department of Electronics and Communication, Manipal Institute of Technology, Manipal, India;
Manipal Academy of Higher Education, Manipal, India

David Valerie School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

Deepa Thilak K. Department of Networking and Communications, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

Deora Mahipal Singh Bhupal Nobel's University, Udaipur, India

Desai Nishant Parul University, Vadodara, Gujarat, India

Deshmukh Sachin N. Department of Computer Science & IT, Dr. Babasaheb Ambedkar, Marathwada University, Aurangabad, M.S., India

Deshpande Santosh L. Visveswaraya Technological University, Belagavi, India

Devendran V. School of Computer Science and Engineering, Lovely Professional University, Phagwara, India

Devi Gayatri Department of Computer Science and Engineering, ABIT, Cuttack, Odisha, India

Diep Vu Thu Hanoi University of Science and Technology, Hanoi, Vietnam

Doe Jane Anonymous University, Atlanta, GA, USA

Doe John Anonymous University, Atlanta, GA, USA

Dubey Nilesh Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research, Faculty of Technology Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, India

Dutta Suparna Birla Institute of Technology, Noida, India

Duy Ha Long FPT University, Hanoi, Vietnam

Faridi Arman Rasool Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Garia Prakash Graphic Era Hill University, Bhimtal Campus, India

Gaurav A. Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Bachupally, Nizampet, Hyderabad, Telangana, India

Gayan Kamal Department of Computer Application, Jorhat Engineering College, Jorhat, Assam, India

Gebru Yalew Werkie Department of Electrical and Computer Engineering, Debre Markos University, Debre Markos, Ethiopia

Geethu N. Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

Giriprasad M. N. Department of ECE, JNTUA College of Engineering Anantapur, Anantapur, Andhra Pradesh, India

Gopalan Sundararaman Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

Gupta Anjani Dr. Akhilesh Das Gupta Institute of Technology & Management, GGSIP University, New Delhi, India

Gupta Atrayee Department of Computer Science and Engineering, Jadavpur University, Jadavpur, India

Gupta Bhoomi Maharaja Agrasen Institute of Technology, GGSIPU, Delhi, India

Gupta Sachin SoET, MVN University, Haryana, India

Hambarde Aparna Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, Maharashtra, India

Handur Vidya S. KLE Technological University, Hubballi, India

Heddallikar ArunKumar RADAR Division, SAMEER, IIT B CAMPUS, Mumbai, India

Hung Phan Duy FPT University, Hanoi, Vietnam

Iyer Nalini C. KLE Technological University, Hubballi, India

Jagannath Bhople Yogesh Department of Information Technology, Government Polytechnic Washim, Washim, India

Jain Dhyanendra Dr. Akhilesh Das Gupta Institute of Technology & Management, GGSIP University, New Delhi, India

Jawale Shila Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

Jeyalakshmi J. Department of CSE, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India

Jha Ravi Shankar KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India

Joshi Kavita Ajay Graphic Era Hill University, Bhimtal Campus, India

Joshi Manuj Pacific Academy of Higher Education and Research Society, Udaipur, Rajasthan, India

Joshi Pulkit AAI, Centre for Development of Advanced Computing, Pune, India

Juneja Pradeep Kumar Graphic Era University, Dehradun, India

K Shamshuddin School of Electronics and Communication Engineering, KLE Technological University, Hubballi, Karnataka, India

Kalaiselvi K. Department of Networking and Communications, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

Kalele Shraddha Amit AAI, Centre for Development of Advanced Computing, Pune, India

Kamakshi Pille Kakatiya Institute of Technology and Science, Warangal, India

Kamdar Dipesh V. V. P. Engineering College, Rajkot, Gujarat, India

Karki Ashtha KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India

Kaur Arshpreet Department of Computer Science and Engineering (ACED), Alliance University, Bangalore, Karnataka, India

Kausar Sayema JDCOEM, An Autonomous Institute, Nagpur, Maharashtra, India

Khanvilkar Manasi MIT Institute of Design, MIT-ADT University, Pune, MH, India

Khokhariya Uday Pandit Deendayal Energy University (PDEU), Gandhinagar, India

Khot Suhas Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, Maharashtra, India

Kirubakaran M. K. Department of Information Technology, St. Joseph's Institute of Technology, Chennai, Tamil Nadu, India

Krishna Sree V. Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Bachupally, Nizampet, Hyderabad, Telangana, India

Kubasadgoudar Ashwin R. KLE Technological University Hubballi, Hubballi, India

Kumar Ajai AAI, Centre for Development of Advanced Computing, Pune, India

Kumar Ashish Department of Mathematics and Statistics, Manipal University Jaipur, Jaipur, India

Kumar Bandari Pranay Kakatiya Institute of Technology and Science, Warangal, India

Kumar Bınod Department of MCA, Rajarshi Shahu College of Engineering, Pune, India

Kumar Shambhavi Pandit Deendayal Energy University (PDEU), Gandhinagar, India

Kumar Sharan Department of ECE, JNTUA College of Engineering Anantapur, Anantapur, Andhra Pradesh, India

Kumar Umesh Govt. Women Engineering College, Ajmer, India

Kumaresan K. Department of CSE, K.S.R. College of Engineering, Tiruchengode, Tamil Nadu, India

Kumbhar Rasika Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, Maharashtra, India

Kurariya Pavan Centre for Development of Advanced Computing, Pune, India

Kuthadi Venu Madhav Department of CS&IS, School of Sciences, BIUST, Palapye, Botswana

Lakshmikanthan C. Department of Electronics and Communication Engineering and Department of Mechanical Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Luqman Mohammad Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Mageshwari M. Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Chennai, Tamil Nadu, India

Maheshwari Akshat BML Munjal University, Gurugram, Haryana, India

Malathi M. Department of ECE, Vivekananda College of Engineering for Women (Autonomous), Namakkal, Tamil Nadu, India

Malavika R. Department of Management, Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, Kerala, India

Mali Rajkumar Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, Maharashtra, India

Mamatha Gannera Department of ECE, JNTUA College of Engineering Anantapur, Anantapur, Andhra Pradesh, India

Mandla Mandeepsingh Parul Institute of Engineering and Technology, Parul University, Vadodara, India

Mane Pranoti JDCOEM, An Autonomous Institute, Nagpur, Maharashtra, India

Mane Venkatesh KLE Technological University, Hubballi, India

Mankad Urja L.J. University, Ahmedabad, India

Maurya Sudhanshu Graphic Era Hill University, Bhimtal Campus, India

Meena Shyam Sundar Department of Computer Science and Engineering, Shri Vaishnav Institute of Information Technology, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore, India

Menon Remya Vivek Department of Management, Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, Kerala, India

Mishra Shreeya Govt. Women Engineering College, Ajmer, India

Mishra Wricha MIT Institute of Design, MIT-ADT University, Pune, MH, India

Mittal Amit Graphic Era Hill University, Bhimtal Campus, India

Modh Jatin C. Gujarat Technological University, Ahmedabad, India

Mukherjee Nandini Department of Computer Science and Engineering, Jadavpur University, Jadavpur, India

Nagaraja K. V. Department of Mathematics, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

Narendra Hardik Parul Institute of Engineering and Technology, Parul University, Vadodara, India

Naresh R. Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Chennai, Tamil Nadu, India

Navya K. T. Department of Electronics and Communication, Manipal Institute of Technology, Manipal, India;
Manipal Academy of Higher Education, Manipal, India

Neogy Sarmistha Department of Computer Science and Engineering, Jadavpur University, Jadavpur, India

Nghia Ha Minh FPT University, Hanoi, Vietnam

Nidhi Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, India

Nikhil Vemu School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

Nikita P. KLE Technological University, Hubballi, India

Nitnaware Pranali J.D. College of Engineering and Management, Nagpur, Maharashtra, India

Pabla Simranjitsingh Parul Institute of Engineering and Technology, Parul University, Vadodara, India

Padmavathi K. Computer Applications, Hindusthan College of Arts and Science, Coimbatore, India

Panchal D. S. Department of Computer Science & IT, Dr. Babasaheb Ambedkar, Marathwada University, Aurangabad, M.S., India

Pancholi Nidhay Pandit Deendayal Energy University (PDEU), Gandhinagar, India

Pandu Ranga Avinash Srikhakollu Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Bachupally, Nizampet, Hyderabad, Telangana, India

Pareek Jyoti Department of Computer Science, Gujarat University, Ahmedabad, India

Parekh Viral C. U. Shah University, Surendranagar, Gujarat, India

Patel Amit Singh KLE Technological University, Hubballi, Karnataka, India

Patel Parul Department of ICT, VNSGU, Surat, India

Patel Swasti Parul Institute of Technology, Parul University, Vadodara, India

Pathak Kirtan Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research, Faculty of Technology Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, India

Patil Prakashgoud KLE Technological University, Hubballi, Karnataka, India

Pattanashetty Vishal B. School of Electronics and Communication Engineering, KLE Technological University, Hubballi, Karnataka, India

Pawar Manjula K. KLE Technological University, Hubballi, Karnataka, India

Phukon Satya Ranjan Department of Computer Application, Jorhat Engineering College, Jorhat, Assam, India

Poongothai T. Department of Computer Science and Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India

Poonkuzhali S. Department of CSE, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India

Pothina Harshini Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

Prasad Keerthana Manipal School of Information Sciences, Manipal, India; Manipal Academy of Higher Education, Manipal, India

Ragu Harini School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

Rajan Annie Department of Computer Science, University of Mumbai, Mumbai, India;
DCT's Dhempe College, Panaji, India

Rajesh M. Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

Rajeshkumar T. Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India

Rampalli Gautham Kakatiya Institute of Technology and Science, Warangal, India

Ramteke Rakesh J. KBC North Maharashtra University, Jalgaon, Maharashtra, India

Rana Atul SoET, MVN University, Haryana, India;
Research Scholar, MVN University, Haryana, India

Rana Swapnil Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research, Faculty of Technology Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, India

Ranjan Department of Computer Science and Engineering (ACED), Alliance University, Bangalore, Karnataka, India

Rastradhipati Atreya MIT Institute of Design, MIT-ADT University, Pune, MH, India

Rath Arabinda Department of Computer Science and Engineering, ABIT, Cuttack, Odisha, India

Rathod D. P. Department EnTC, VJTI (Veermata Jijabai Institute of Technology), Mumbai, India

Ratnam Tatini Shanmuki Krishna Department of Electronics and Communication Engineering and Department of Mechanical Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Rejikumar G. Department of Management, Amrita Vishwa Vidyapeetham, Kochi, Kerala, India

Roy Himadri Shekhar Department of Computer Science and Engineering, Jadavpur University, Jadavpur, India

Saahith Reddy M. Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Bachupally, Nizampet, Hyderabad, Telangana, India

Sahoo Priti Ranjan KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India

Saini Jatinderkumar R. Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India

Saini Monika Department of Mathematics and Statistics, Manipal University Jaipur, Jaipur, India

Saini Pawankumar Symbiosis International University, Pune, Maharashtra, India

Sakthi U. Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India

Salau Ayodeji Olalekan Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti, Nigeria

Salgaonkar Ambuja Department of Computer Science, University of Mumbai, Mumbai, India

Sangpal Ravivanshikumar Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, Maharashtra, India

Saranya S. Department of ECE, SRM Easwari Engineering College, Chennai, India

Sarkar Ram CSE Department, Jadavpur University, Kolkatta, West Bengal, India

Sarmah Joydip Department of Computer Application, Jorhat Engineering College, Jorhat, Assam, India

Sasikumar P. School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

Sawarkar S. D. Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

Sawwashere Supriya J.D. College of Engineering and Management, Nagpur, Maharashtra, India

Senthil Murugan T. Kakatiya Institute of Technology and Science, Warangal, India

Senthilkumar A. V. Computer Applications, Hindusthan College of Arts and Science, Coimbatore, India

Shah Axita Department of Computer Science, Gujarat University, Ahmedabad, India

Shah Harshal Parul University, Vadodara, Gujarat, India

Shah Kaushal Pandit Deendayal Energy University (PDEU), Gandhinagar, India

Shaikh Arshad Learning and Personalization, BYJU'S, Bangalore, India

Sharma Anupam Kumar Dr. Akhilesh Das Gupta Institute of Technology & Management, GGSIP University, New Delhi, India

Sharma Somesh Graphic Era Hill University, Bhimtal Campus, India

Sharma Ved Prakash Dr. Akhilesh Das Gupta Institute of Technology & Management, GGSIP University, New Delhi, India

Shashidhar N. KLE Technological University, Hubballi, India

Shashvat Kumar Department of Computer Science and Engineering (ACED), Alliance University, Bangalore, Karnataka, India

Shelke Mahesh B. Department of Computer Science & IT, Dr. Babasaheb Ambedkar, Marathwada University, Aurangabad, M.S., India

Shell Michael Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Shet Raghavendra KLE Technological University Hubballi, Hubballi, India

Shinde Shraddha S. Department EnTC, VJTI (Veermata Jijabai Institute of Technology), Mumbai, India

Shraddha B. H. KLE Technological University, Hubballi, India

Shukla Madhu Department of Computer Engineering, Marwadi University, Rajkot, India

Singhal Vineet BML Munjal University, Gurugram, Haryana, India

Singh Lenali AAI, Centre for Development of Advanced Computing, Pune, India

Singh Prashant Dr. Akhilesh Das Gupta Institute of Technology & Management, GGSIP University, New Delhi, India

Singh Preeti MMM University of Technology, Gorakhpur, India

Singh Sarvpal MMM University of Technology, Gorakhpur, India

Singh Shashi Pal AAI, Centre for Development of Advanced Computing, Pune, India

Singh Shivoham Pacific Academy of Higher Education and Research Society, Udaipur, Rajasthan, India

Smrutirekha KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India

Sonekar Shrikant JDCOEM, An Autonomous Institute, Nagpur, Maharashtra, India

Sonekar Shrikant V. J.D. College of Engineering and Management, Nagpur, Maharashtra, India

Sood Ajay Kumar BML Munjal University, Gurugram, Haryana, India

Sowmia K. R. Department of IT, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India;
Department of CSE, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India

Sravani Kumari K. Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Bachupally, Nizampet, Hyderabad, Telangana, India

Sreenivasulu G. Computer Science and Engineering, ACE Engineering College, Hyderabad, India

Sridharan Aadityan Department of Physics, Amrita Vishwa Vidyapeetham, Amritapuri, India

Sunori Sandeep Kumar Graphic Era Hill University, Bhimtal Campus, India

Suresh A. Department of Management, Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, Kerala, India

Swaminarayan Priya Parul Institute of Computer Applications, Parul University, Vadodara, India

Tamizharasan M. Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Tembhurne Snehal J.D. College of Engineering and Management, Nagpur, Maharashtra, India

Thangaraj K. Department of Information Technology, Sona College of Technology, Salem, Tamil Nadu, India

Thulasi Chitra N. Department of CSE, MLR Institute of Technology, Secunderabad, Telangana, India

Tien Nguyen Huu FPT University, Hanoi, Vietnam

Tokekar Vrinda Department of Information Technology, Institute of Engineering and Technology, Devi Ahilya Vishwavidyalaya, Indore, India

Trivedi Tanvi Bhupal Nobel's University, Udaipur, India

Upadhyaya Shuchita Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, India

Vallapure Pratiksha Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, Maharashtra, India

Vartika Department of Computer Science and Engineering (ACED), Alliance University, Bangalore, Karnataka, India

Vats Prashant Dr. Akhilesh Das Gupta Institute of Technology & Management, GGSIP University, New Delhi, India

Vatsal Saumitra Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India

Vijay Anand M. Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India

Vinh Bui Trong Hanoi Procuratorate University, Hanoi, Vietnam

Vishal P. KLE Technological University, Hubballi, India

Viswanadha Raju S. CSE, JNTUHCEJ, Jagityal, Karimnagar, Telangana, India

Yadav Babita SoET, MVN University, Aurangabad, India

Zala Kirtirajsinh Department of Computer Engineering, Marwadi University, Rajkot, India

Implementation of Tweet Stream Summarization for Malicious Tweet Detection

Pranoti Mane, Shrikant Sonekar, and Sayema Kausar

Abstract In contrast to traditional media, social media is populated by anonymous individuals who have the freedom to broadcast whatever they choose. This online social media culture is dynamic in nature, and the move from traditional media to digital media is growing increasingly popular among people. While conventional media will continue to be used less frequently in the future, the increasing use of online social networks (OSNs) will obfuscate the real information provided by traditional media. Genuine users provide material that is beneficial to the broader public; on the other hand, spammers transmit irrelevant or misleading content that turns social media into a front for spreading false information. Consequently, undesirable text or susceptible links might be delivered to targeted users. These fabricated texts are anonymous, and they are occasionally linked to possible URLs. A precise statistical classification for a piece of news is not possible with the current systems because of data limitations and communication types. We will look at a variety of research publications that employ a variety of strategies for master training in the prediction and detection of harmful material on social media websites and networks. In this study, we attempted to identify spam tweets from a large collection of tweets by utilizing TVC algorithm to identify it. When a content paper is incorporated in this manner, a brief summary produced by utilizing the most important keywords from the first document is known as summarization. It is necessary to have a dynamic approach of dealing with the condensed information that is supplied through Twitter feeds. This research presents a novel way for producing a major substance-based summary in a shorter period of time than previously available. We also propose to detect harmful tweets both offline and online and to do so in real time. Most notably, when compared to the other current frameworks, the proposed framework performs multi-subject summarization on an online dataset, which results in a reduction in the amount of time required.

Keywords Tweet clustering · Tweet stream · Malicious tweet · Timeline · Summary · Continuous summarization

P. Mane (✉) · S. Sonekar · S. Kausar
JDCOEM, An Autonomous Institute, Nagpur, Maharashtra, India
e-mail: pranoti.mane23@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_1

1 Introduction

Due to technological advancements, an online community has evolved into a critical network for sharing and exchanging information. Twitter is among the most popular online social networking services because it allows users to send and read a large number of text-based messages. Thousands of millions of individuals share their information with their contacts on Twitter, and around 400 million tweets are sent every day. It is a frequent feature of social networks that multiple different individuals have a strong influence on one another. Because of the huge amount of available individuals and the volume of data shared on Twitter social networks, spammers engage in a slew of criminal operations aimed at spreading spam messages via connected website URLs.

Many e-commerce applications have made use of the social networking site Twitter to promote their products. Bot software automates functions in social media networks such as Twitter. Robotic process automation (RPA) is a type of software that automates the execution of operations over the online platform and in other web-based applications. In terms of computerization, Twitter is not particularly picky. Twitter just includes the definition of a CAPTCHA graphic as part of the registration procedure. Using Twitter APIs, bots conduct automation tasks for users after they have completed the sign-up procedure and received all of their personal information. Cyborgs are in charge of controlling the interface between humans and bots software.

In the case of large-scale tweets, conventional document summarizing algorithms are insufficient, and they are also insufficient in the case of tweets that are received quickly and continuously. To completely overcome this problem, Twitter summarization is required, which should have a new level of utility that is fundamentally different from ordinary summary. When summarizing tweets, it is important to consider the worldwide element of the tweets that have arrived.

Stemming is a phrase that is used in the field of etymological morphological and data recovery to describe the procedure of reducing curved (or occasionally determinate) words to their basic stem, base, or root frame, which is by and large a constructed word shape. The stem does not have to be identical to the grammatical underpinning of the word; it is often sufficient that related terms point to a comparable stem, independent of whether or not this stem is a legal root in and of itself.

During the registration process, collocations will be words that are sorted through prior to or after the preparation of ordinary dialect information is completed (content). Despite the way that stop words are frequently used to refer to the most well-known words in some kind of a language, there is no single widely used list of stop words that is utilized by all normal dialect preparation equipment, and to be sure, not all gadgets even use such a list. A few instruments, in particular, refrain from emptying particular stop phrases in order to aid in state seeking efforts.

2 Literature Survey

The process of summarizing a tweet is divided into two stages. The introduction advance necessitates the clustering of tweet information, followed by the summarizing of the information.

Twitter posts are limited to 140 characters or less in length, and we are only considering English posts in this case. All of the tweets are written in a casual manner with non-standard spelling and, as much as feasible, without the use of accentuation. Nonlinear and non-descriptions of Twitter posts were created using the crossovers TF-IDF-based calculation [1] as a starting point. The following are some ways to document summarizing that have been clarified. Using the arbitrary summarizer approach, k posts or even every subject is randomly selected for inclusion as a summary. This technique was beneficial in terms of achieving the final aim of providing the most exceedingly terrible execution and, in addition, setting the lowest bound of scenario execution. The latest summarizer technique selects the most recent k posts from the determination pool to serve as a summary. It has the ability to select the first portion of a news story to serve as a summary. This strategy is implemented in light of the fact that smart summarizers are unable to outperform a plain summarizer in terms of performance. This summarizer just makes use of the first segment of the document to provide a summary.

The computation for streaming informational aggregation has been extensively researched and written about by a variety of authors. BIRCH is an improved iterative decreasing and clustering algorithm that makes use of progressions' calculation. An unstructured machine learning technique [2] was used to perform this calculation.

An authentic and realistic result can be developed through mathematical approach for selecting the cluster based on distance and energy. Through this, the cluster can be easily identified and can take up the responsibility of running the entire network [3]. It is used to achieve multiple levels of clustering over large informational sets of varying sizes and shapes. The advantage of BIRCH is that it has the ability to form groups in response to incoming information and to respond in a dynamic manner to those groups.

People can share their opinions on Twitter, which is a good platform for doing so. Because of the sheer volume of tweets, it takes a lot of effort to keep up with what is going on inside the events. Presented here is a new technique for outlining events that makes use of excellent journalists and creates live game updates based on Twitter posts about the events. Great journalists selected illustrative tweets from the majority of non-educational tweets [4] to use as examples.

TextRank summarizer [5] is yet another positioning computation that is based on a chart. It is this strategy that makes use of the PageRank computation. This resulted in yet another chart-based summarizer that merges potentially even more data than LexRank. This occurs as a result of the fact that it modifies the weights of postings in a recursive manner. The final score assigned to each post is determined by how well it is associated with other posts that are swiftly affiliated with it, as well as the method in which distinct presentations are associated on different articles. Content

ranking does not necessitate any further semantic learning. TextRank computation is a diagram-based approach that is used to discover the sentences that are in the top place.

In the web mining industry, differential evolutionary timeline Summarization (ETS) [6] is a tool that builds timelines for large amounts of information. ETS provides developmental instructions for a given time period. ETS provides outlines that are based on a score of time characteristic attributes. The most popular point of view is that it stimulates speedy news perusal and learning grasp of the issues. Iterative substitution is used to transform the ETS errand into an adjusted enhanced issue.

This summarizer employs a chart-based method, as demonstrated in LexRank [7]. It distinguishes between pairwise similarities between two phrases either between two blog posts, respectively. It has an impact on the proximity to score, which is the weight of the edge between both the two sentences. According to the weights of the connections that are connected with one another, the last rating of postings is processed in the following way: In contrast to coordinate recurrence summarization, this summarizer is excellent for providing summarizing in light of standard to chart. In spite of the fact that it is predicated on repetition, this framework makes use of the links between phrases to contain more information. In comparison with the recurrence-based method, this is a more complicated computation.

It is also possible to use Twitter streams for event summaries, which allows you to communicate with data in real time. When it comes to occasion summarization, the participant technique is used. Summary tweet extraction is accomplished through the use of three essential segments: Participant detection, sub-occasion detection, and summary tweet extraction [8]. Members of an occasion are distinguished by their member identifier. Individuals who participate in events are referred to as members. Sub-occasion discovery is a feature that summarizing sub-occasions that have been identified with members. By applying the summary tweet extraction section [9], the tweets are extracted from the sub-occasions.

Zhenhua Wang and colleagues have developed a summarizing structure known as Sumblr. Continuous summarization using stream clustering is what we are talking about here. Because of the vast amount of noisy and excessive tweets that are contained in it, continuous summarization is a challenging task to accomplish. The notion in question was taken into consideration when the idea of continuous tweet streaming summarization was being discussed. A tweet stream clustering component, a high-level summarization module, and a timeline generation module are the three primary components of this system. The social media platform Sumblr can also be used to experiment with dynamic, rapidly incoming, and large-scale tweet streams [10], among several other things.

3 Proposed Approach

Realizing continuous Twitter stream summarization is not an easy task, given that countless people are excellent for nothing, immaterial, and raucous in nature as a result of the social concept of tweeting, among other factors. Tweets are inextricably linked to the time at which they are posted, and fresh tweets have such a tendency to reach their target audience at a rapid pace. Since the volume of tweets is always significant, the summary calculation must be extremely efficient. It should be able to provide Twitter summaries of conscience spans of time. It should be able to automatically recognize subtheme modifications as well as the minutes in which they occur. A multi-subject reproduction of a continuously tweet stream summarizing structure, to be particular Sumblr, is proposed in this paper. It is tested on more comprehensive and significant scale informative collections, and the results are discussed in this study.

Figure 1 depicts the multi-theme variant of the Sumblr structure, which is composed of three core components: the tweet stream clustering module, the abnormal state summary module, and the timeline age module. Figure 1 shows the multi-theme form of the Sumblr structure. The Twitter stream clustering module is responsible for keeping the online truthful information current. The incremental clustering technique is used to keep tweets in their web version current. When a theme-based Twitter stream is provided, it can be used to group tweets in a productive manner and maintain reduced group data. The different leveled summarization module provides two types of synopses: Online rundowns and rundowns that have been recorded. In a brief web summary, it is depicted what is now being investigated among the general public. As a result, the contribution for building online synopses is retrieved precisely from the current bunches that have been kept up in memory in this manner. A verifiable summary, on the other hand, helps individuals to perceive the most important events that occurred during a specific period, which means we must

Fig. 1 System overview

disregard the influence of tweet content that occurred from outside that period in order to do this. As a result, the recovery of the information essential for the creation of genuine designs is more difficult. The development recognition computation at the heart of the chronology age module gives both ongoing and runtime timelines in the same calculation.

4 Implementation

Algorithm 1: Multi-topic Version of Sumblr

Input: a cluster set CL_set

```
While! Stream.end() do
Tweet t = stream.next();
Choose Cpt in CL_set whose centroid closet to it.
If MaxSim(t) < MBS then
Create a new cluster CLnew = {t};
CL_set.add(CLnew);
Else
Update Cpt with t;
If TScurrent % (αi) == 0 then
Store CL_set into PTF;
```

Algorithm 2: Tweet Stream Clustering

Input: CL_set represent a set of cluster

```
While! stream.end () do
Tweet t = stream.next ();
Choose Cpt in CL_ set whose centroid closest to t;
If MaxSim (t) < MBS then
Generate new Cluster CLnew = {t}
CL_set.add (CLnew)
Else
update Cpt with t
If TScurrent%(αi) = = 0 then
Store CL-set into PTF.
```

Algorithm 3: TCV-Rank Summarization

Input: The set of clusters D(c)

Output: SS represent the set of summaries

```
SS = Φ, T = {represent number of tweets in set D(c)};
Comparability graph to be built on T;
```

3. Calculate LR i.e. LexRank scores;
4. Tc represent the tweets having highest LR score in each cluster;
5. while |SS| < L do
6. for each tweet ti in Tc_SS do
7. calculating vi in accordance the Equation given in (2);
8. select tmax with the highest vi;
9. SS.add (tmax);
10. while |SS| < L do
11. for each tweet t'i in T _ SS do
12. calculate v'i according to Equation
13. select t'max with the highest v'i;

    1. SS.add (t'max);
    2. returns SS;

Algorithm 4: Topic Evolution Detection

Input: tweet stream binned by time units

Output: timeline node set TLN

1. TLN = ∅;
2. Whilestream.end() do
3. Bin Ci = stream.next();
4. If hasLargeVariation() then
5. TLN.add(i);
6. Return TLN;

5 Result Analysis

Using Sumblr with a single topic against Sumblr with many topics, the time comparison in Fig. 2 is shown. In comparison with Sumblr with several topics, Sumblr with a single topic took more time to complete.

Fig. 2 Time comparison

Fig. 3 Memory comparison

Fig. 4 Comparison of malicious and non-malicious tweets

Figure 3 depicts a comparison of memory. When compared to the suggested solution, the existing system needed greater CPU consumption. The storage comparison between different systems is depicted in Fig. 4.

6 Conclusions

The investigation of several approaches for document summary, such as sifting and Twitter summarization, has come to an end. These approaches are used to monitor a large number of tweets at the same time. Due to the fact that Twitter information is raucous and excessive, sifting is not an effective strategy. As a result, the summarizing

technique is used to compress the information included in tweets. Traditional document summarizing algorithms are not practical for large-scale tweets, and they are also not adequately applicable for tweets that are received quickly and continually; in addition, they do not focus on static and small-scale informational indexes. An adjustment of a constant Twitter post live feed summarization system, specifically Sumbler, has been developed to address this issue by producing simple summaries and timelines with respect to streams and assessing it on more comprehensive and considerable magnitude information examination, which manage dynamic tweet streams with rapid arrival and a large scale. During the process of continuous summarization, our technique will identify the altering dates and timeframes in a progressive manner. Furthermore, evolutionary timeline summarization (ETS) somehow does not pay attention to concerns such as proficiency and adaptability, which are critical in our streaming-based environment.

References

1. S. Anjali, N.M. Meera, M.G. Thushara, A graph based approach for keyword extraction from documents. in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)* (2019), pp. 1–4. https://doi.org/10.1109/ICACCP.2019.8882946
2. R. Yan, X. Wan, J. Otterbacher, L. Kong, X. Li, Y. Zhang, Evolutionary timeline summarization: A balanced optimization framework via iterative substitution. in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2011), pp. 745–754
3. S.V. Sonekar, M.M. Kshirsagar, L. Malik, Cluster head selection and malicious node detection in wireless Ad Hoc networks. In: *Next-Generation Networks*, ed. by D. Lobiyal, V. Mansotra, U. Singh. Advances in Intelligent Systems and Computing, vol 638 (Springer, Singapore, 2018). https://doi.org/10.1007/978-981-10-6005-2_55
4. M. Kubo, R. Sasano, H. Takamura, M. Okumura, Generating live sports updates from Twitter by finding good reporters. in *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (2013), pp. 527–534. https://doi.org/10.1109/WI-IAT.2013.74
5. T. Chen, D. Miao, Y. Zhang, A graph-based keyphrase extraction model with three-way decision. in *Rough Sets. IJCRS 2020*, ed. by R. Bello, D. Miao, R. Falcon, M. Nakata, A. Rosete, D. Ciucci. Lecture Notes in Computer Science, vol 12179 (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-52705-1_8
6. S. Dutta, V. Chandra, K. Mehra, S. Ghatak, A.K. Das, S. Ghosh, *Summarizing Microblogs during Emergency Events: A Comparison of Extractive Summarization Algorithms* (2019)
7. C.-F. Tsai, Z.-C. Chen, C.-W. Tsai, MSGKA: an efficient clustering algorithm for large databases. in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 5 (2002), p. 6. https://doi.org/10.1109/ICSMC.2002.1176400
8. C. Shen, F. Liu, F. Weng, T. Li, A participant-based approach for event summarization using twitter streams. in *Proceedings of the Human Language Technology Annual Conferene North America Chapter Association for Computer Linguistics* (2013), pp. 1152–1162
9. P. Nerurkar, A. Shirke, M. Chandane, S.G. Bhirud, Empirical analysis of data clustering algorithms. Procedia Comput. Sci. **125**, 770–779 (2018)
10. D. Rudrapal, A. Das, B. Bhattacharya, A survey on automatic Twitter event summarization. J. Inf. Process. Syst. **14**, 79–100 (2018)

Comparative Analysis of a Secure Authentication Protocol for 5G-Enabled IoT Network Using Public and Shared Secret Key

Shamshuddin K and Vishal B. Pattanashetty

Abstract Due to advancements in technologies and to achieve high efficiency, the world is shifting from 4G to a 5G network. The advantages of 5G network are enormous in every field and thus are catching speed to be a better version of what exists in today's world. The features like increased bandwidth availability, massive network capacity, and ultra-low latency are keeping 5G out of the box and giving raw network engineers to explore and invent high standards. As the world goes with pros and cons hand in hand, so is the advent of 5G technology. Finding its applications in communication, massive IoT, and mobile broadband, it also faces some challenges which affects the day-to-day life of humans. 5G applications face security risks due to the new technology used and the performance requirements of the specific application scenario. To avail the 5G services, the IoT devices need to communicate with an intermediate network known as an access network, which is usually publicly accessible and thus prone to attacks. In this paper, implementation and results of a 5G-enabled secure authentication protocol using symmetric and asymmetric key are compared to identify the best technique for privacy preservation in 5G-enabled IoT network.

Keywords Cooja · Cryptography · Privacy · Security · IoT

1 Introduction

The Internet of Things (IoT) is evolving as a platform for Internet-based service and comprises smart devices, such as smart phones and sensor nodes [1]. The IoT devices consist of communication system, RFID, and complex computational system. The most challenging aspects in IoT network are security and privacy [2].

S. K (✉) · V. B. Pattanashetty
School of Electronics and Communication Engineering, KLE Technological University, Hubballi, Karnataka, India
e-mail: shamshuddin@kletech.ac.in

V. B. Pattanashetty
e-mail: vishalbps@kletech.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_2

In the present networking system, the corrupted devices can corrupt the functioning of network. If an IoT device is corrupted, then the entire connected physical world will get influenced dangerously. The advancements in the technologies in the communication sector occurring like a whirlwind, from first generation (1G) of cellular technologies to 5G in the expanse of about 40 years, has opened up so many different ways of communication and data transfer not only between two individuals but also between different devices like sensors, servers, IoT, and Internet of Vehicles (IoV). These provide platforms for the emanating technologies that are transforming into an essential component of our economy as well as our community. The 5G technology is becoming an indispensable technology as it is taking a humongous leap in performance when compared to the 4G networks. In the IoT network, the communication system will be broadcast in nature which makes wireless transmission system susceptible to eavesdropping and endangering the secrecy of data [3]. So there arises an increased need to protect and safeguard the data we send and receive. In today's world, due to increase in online data transfer such as online banking transaction and payment, security and privacy of data are gaining more importance and also area of concern.

2 Literature Survey

In 5G-enabled IoT network, the number IoT node will be enormous in number. Authentication of device is very important for providing privacy and security in 5G and IoT network. Liyanage et al. [2] describe the various privacy issues due to the new technologies involved in 5G network. The various 5G security and privacy issues, their defense, and also the road map for future research are described by the author of the paper [3]. Using Cooja simulator, the authors simulate IoE-based home automation system to analyze IEEE 802.15.4 protocol to explore IoT environment [1]. In another paper, authors show simulation results of energy consumption by a node using Cooja simulator [4]. The author has conducted experiment on energy constrained IoT devices using Cooja simulator and also carried out comparison of two protocols: constrained application protocol (CoAP) and hypertext transfer protocol (HTTP) [5].

Many researchers have proposed various authentication protocols. In the paper [6], authors proposed a secure authentication protocol for 5G-enabled IoT network. The proposed protocol is verified using Scyther tool [7] for various security parameters. In this paper, the proposed protocol in paper [6] is simulated using Cooja simulator [8] and further analyzed using symmetric and asymmetric key.

The proposed system is designed by using algorithms like the Diffie–Hellman for key exchange and RSA, and other cryptographic techniques [9] to authenticate the user equipment (UE) and to authenticate the access and mobility management (AMF) in 5G network [10], to check the integrity of the device that requests for the service, and finally to acknowledge the service request [11]. Typically, a user

device requests for service to core access and mobility management function and is responsible for service allocation after the verification of network slice selection association information (NSSAI) [12].

‘

3 System Design

The system design can be best explained with the help of Fig. 1. The black box on the Fig. 1 shows the focus on the area we are working on. The main goal is to authenticate and preserve security between the user equipment and AMF. User equipment such as mobile phones, various IoT devices like cars, or a fitness wearables, send a request to the 5G network [13], the request reaches to the AMF for the services, and the AMF is accessible by any IoT node, hence, it is prone to attacks by the hackers. That being the case, the proposed protocol presents a scenario where attackers try to attack, but attack failed, because of the authentication protocol checks for the integrity of the device requesting service, the attacks are stopped at the root itself.

3.1 Functional Block Diagram

Figures 2 and 3 represent the functional block diagram of the protocol [6]. The block diagram consists of the extreme ends between which the communication takes place that is user equipment and the AMF. The different blocks that are required to carry

Fig. 1 System design [6]

Fig. 2 Functional block diagram for symmetric key

Fig. 3 Functional block diagram for asymmetric key

on the protocol include a random number generator (nonce), a hashing function, encryption, decryption, and a simple logical XOR function. With the help of these functions and a UDP protocol running at the application layer in end devices, the protocol is implemented.

3.2 Algorithm

Algorithm 1:

```
STEP 1: Select RandomNo : rn and User id :
 UEid
Send (UEi to AMF) : UEid, rn
STEP 2: Ki = retrievekey from UEid
Choose RandomNo : rn2
Skey = rn2 exor Ki
e1 = Eki(rn, Skey)
h1 = H(Skey — e1 — rn)
send (AMF to UEid) : e1, h1
STEP 3: (rn, Skey) = Dki(e1)
rn2 = Skey exor Ki
e2 = Eki(rn, Skey)
h2 = H(Skey — e2 — rn)
if( e1 == e2 and h1 == h2 )
Authenticated AMF
else (Terminate the protocol and communication)
h2d= H(h1 — Skey — e2 — rn)
send (UE to AMF) : h2d
STEP 4: h3 = H(h1 — Skey — e1 — rn)
if(h3 == h2d)
authenticated User Equipment UEid
send (AMF to UE) : h3
STEP 5: e3 = Eki(ServiceRequest)
h4 = H(e3)
A = h4 ⊕ h2d
send (UEid to AMF) : e3, A
STEP 6: h4 = A exor h2d
h4= H(e3)
if(h4d== h4)
then integrity is preserved
ServiceRequest = Dki(e3)
execute the ServiceRequest
send (AMF to UE): Ack
```

4 Results and Discussion

The analysis of results is based on the following parameters, which are average power consumption, radio duty cycle, node information, network graph, and sensor map.

- Average power consumption: The sum of the power contributions of the node's electronic components, which are in turn dependent on the component state and

Fig. 4 Average power consumption

the actual operation executed, determines the node's power consumption as shown in Fig. 4.

- Radio duty cycle: The ratio of time when transmission of signal is ON to the time when transmission of signal is OFF, as shown in Fig. 5.
- Node information: Figures 6 and 7 give information about various parameters when simulated using symmetric key and asymmetric key.
- Network graph: Figure 9 shows interconnections between a group of elements that are depicted in network diagrams. A node is the representation of each entity. Links are used to represent connections between nodes.
- Sensor map: Figure 8 shows sensor map, and device manager allows users to examine real-time graphs of sensor data over several date and time periods using the logged sensor readings. The graphs can be useful in identifying power or cooling patterns, as well as possible trouble locations in the data center environment.

The comparison of symmetric key and asymmetric key is listed in Table 1. From node analysis, we found that asymmetric key encryption can be used to perform simple operations like exchanging passwords, etc., because it is more secure than symmetric key encryption. Public key generation will take more time, and its database has to be maintained. Symmetric key encryption is suitable if there is bulk data transfer. If more nodes are present, symmetric key encryption is the go-to method, and it is less secure.

Fig. 5 Radio duty cycle

Fig. 6 Node information for asymmetric key

Fig. 7 Node information for symmetric key

4.1 Conclusion

An efficient application layer security protocol between user equipment and AMF in 5G network is implemented and verified. The protocol is tested for real time parameters using Cooja simulator on Contiki OS. Based on the observation, asymmetric encryption can be used to perform operations like exchanging passwords, etc. Public key generation will take more time, its database has to be maintained, and it is more secure. Whereas symmetric encryption is suitable for bulk data transfer. If more nodes are present, symmetric encryption is the preferred method, and it is less secure compared to asymmetric encryption.

Fig. 8 Sensor map

Fig. 9 Network graph

Table 1 Comparison of encryption algorithms

Parameter	Asymmetric key	Symmetric key
Average power consumption (mW)	60.185	60.182
Listen duty cycle (%)	99.996	99.994
Average transmission time (ms)	600	450

4.2 Future Scope

The secured authentication protocol can be further analyzed by implementing a key management system for public and private key, also sharing secret or session key using public key, and analyzing the performance of the both techniques. In this paper, the results obtained for single user equipment (UE) node is described; in the future, the number of UE nodes can be increased, and performance can be analyzed.

References

1. G.P. Dave, N. Jaisankar, Analysis and design of energy efficient protocols for wireless sensor networks. in *2017 Innovative Power Advanced Computing Technology* (2017), pp. 1–7
2. M. Liyanage, J. Salo, A. Braeken, T. Kumar, S. Seneviratne, M. Ylianttila, 5G privacy: scenarios and solutions. in *IEEE 5G World Forum, 5GWF 2018—Conference Proceedings* (2018), pp. 197–203
3. E. Bertino, S.R. Hussain, O. Chowdhury, 5G Security and Privacy: A Research Roadmap. *arXiv*, (2020)
4. M. Tutunoić, P. Wuttidittachotti, Discovery of suitable node number for wireless sensor networks based on energy consumption using Cooja. in *2019 21st International Conference on Advanced Communication Technology* (2019), pp. 168–172
5. K.P. Naik, U.R. Joshi, Performance analysis of constrained application protocol using Cooja simulator in Contiki OS. in *2017 International Conference on Intelligence Computer Instrumentation and Control Technology* (2017), pp. 547–550
6. S. Sharma et al. Secure authentication protocol for 5G enabled IoT network. in *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (2018), pp. 621–626
7. C. Cremers, *Scyther: Semantics and Verification of Security Protocols* (2006)
8. A. Dunkels, B. Grönvall, T. Voigt, Contiki—a lightweight and flexible operating system for tiny networked sensors. in *Proceedings of Conference on Local Computer Networks, LCN* (2004), pp. 455–462
9. J.H. Anajemba, C. Iwendi, M. Mittal, T. Yue, *Improved Advance Encryption Standard with a Privacy Database Structure for IoT Nodes* (2020), pp. 201–206. https://doi.org/10.1109/csnt48778.2020.9115741
10. F. Pan, H. Wen, H. Song, T. Jie, L. Wang, 5G security architecture and light weight security authentication. in *2015 IEEE/CIC International Conference on Communication China—Work. CIC/ICCC 2015* (2017), pp. 94–98. https://doi.org/10.1109/ICCChinaW.2015.7961587
11. Y. Li, L. Cai, Security aware privacy preserving in D2D communication in 5G. IEEE Netw. **31**, 56–61 (2017)

12. D. Fang, Y. Qian, 5G wireless security and privacy: architecture and flexible mechanisms. IEEE Veh. Technol. Mag. **15**(2), 58–64 (2020). https://doi.org/10.1109/MVT.2020.2979261
13. N. Alsaffar, H. Ali, W. Elmedany, Smart transportation system: a review of security and privacy issues. in *2018 International Conference on Innovative Intelligence Informatics, Computer Technology 3ICT 2018* (2018), pp. 2018–2021. https://doi.org/10.1109/3ICT.2018.8855737

DDoS Botnet Attack Detection in IoT Devices

Bandari Pranay Kumar, Gautham Rampalli, Pille Kamakshi, and T. Senthil Murugan

Abstract Distributed denial of service (DDoS) attacks are most harmful threats on present Internet. The Internet of things (IoT) weaknesses give an optimal objective to botnets, making them a significant supporter of the expanded number of DDoS attacks. The expansion in DDoS attacks made it essential to address the results as it suggests toward the IoT business as one of the significant causes. The main motive of this paper is to provide an analysis of the attempts to prevent DDoS attacks, which mainly at the network level. The reasonableness of these arrangements is removed from their effect in settling IoT vulnerabilities. It is obvious from this survey that there is no ideal arrangement yet for IoT security, but this field has large chances for innovative work.

Keywords Distributed denial of service · Internet of things · Botnet

1 Introduction

DDoS attacks are subclass for denial of service (DoS) attacks. A DDoS attack involves multiple connected online devices, collectively referred to as a botnet, which are wont to flood a target Website with fake traffic. The most commonly DDoS attack involves the flooding of a huge amount of traffic to deplete network resources, bandwidth, target CPU time, etc. The DDoS most attacks are HTTP flood, NTP amplification, SYN flood, DDoS flood, hoping flood, UDP flood, and ICMP (PING) flood. In this paper, we use DDoS attacks-based network flow type of testing in IoT botnet a categorical specific dataset.

The IoT botnet works for launching DDoS attacks on target IoT structures to interrupt the services and operations of IoT. This type of malware can interrupt the functions of connected devices after executing the code. Sometimes hackers can create the botnet to target specific IoT devices. The botnet works as an attacker. Firstly, they will create the bots for getting their tools and scripts for the DDoS

B. P. Kumar (✉) · G. Rampalli · P. Kamakshi · T. Senthil Murugan
Kakatiya Institute of Technology and Science, Warangal 506015, India
e-mail: bandaripranay766@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_3

attack and they remotely send commands to active IoT devices to activate the botnet. After activation, the botnets are sending requests to many targeting IoT devices. After attacking all the target IoT devices, the attacked can crash the entire network.

The Internet of things also being called as the Internet of objects, which refers to the wireless networks present between the objects. The devices are connected through the Internet on stands on the IPV6 allows the unique address for them. There some various security services that are mandatory for IoT, and this build for information for security, they are

- Confidentiality: Messages passing from source to destination can be easily intercepted by the attacker, and the content can be hacked. So this message must be hidden from all relay nodes, meaning the message will pass securely end to end is required in the Internet of things. The same can also be applied to devise storage. A simple solution to this is the encryption/decryption mechanism.
- Integrity: The message passing from source to destination should not be changed; it must be received by the recipient. The same is sent by the sender. No medium must change the content of the message as it is being passed or passed device.
- Availability: For the continuous operation of the Internet of things and access to data when necessary, it is also important. The services provided by the devices shall always be available and continuously in working order. So it is important to intrusion detection and intrusion prevention to ensure availability.
- Authenticity: The end-user must be able to learn the identity of the other to ensure that he or she interacts with them. The same entities they claim.

2 Literature Survey

The process of IoT botnet attack detection by the acceptance of IoT technology is spread strongly because the aptitude to produce a far better service. Security additionally changing into a significant difficult role with this development, and IoT contains a resource low procedure and memory resource constraints of IoT. Ali et al. [1] performed a comprehensive systematic literature review on IoT-based botnet attacks. The prevailing state of the art among the realm of study was given and mentioned intimately. A systematic methodology was adopted to create positive the coverage refers of all vital studies. It have been terminated methodology and duplicable. The review made public the prevailing planned contributions, datasets handled, network rhetorical ways used, and analysis focus of the primary hand-picked studies. The results of this review unconcealed that analysis throughout this domain gains momentum, notably among past three years (2018–2020). Nine key contributions were additionally known, with analysis on system, model is that the foremost conducted. This method concluded solvation to detect the botnet activity corresponding to a consumer of IoT devices and networks related.

McDermott et al. [2] proposed deep learning method to detect model based on a bidirectional long short-term memory-based recurrent neural network (BLSTM-RNN). The Mirai botnet attack was detected by the BLSTM-RNN detection model

was helped by the four attack vectors. Results for DNS, UDP, and Mirai were highly supportive with 98, 98, 99% validation accuracy and validation loss metrics of 0.116453, 0.125630, 0.000809 accordingly.

IoTs known as a device-to-device communication, communicate with each other without human presence, such as a small room lamp to the remote controller. Sonar et al. [3] conduct a survey on DDoS attacks on IoTs. Authentic user restricted, network resources unavailable, and bandwidth are consumed, when DDoS is attacked. Upgraded and high-power capacity devices must be developed to find and shut out the attack. Sonar et al. [3] reveled by few DDoS attacks, which are capable of malfunctioning or complete shutdown of an entire IoT network.

Thousands of IoT devices are vulnerable to even a sings DDoS attacks and IoT botnets group of thousands of IoT devices causing huge threats to Internet security in health, home, social, defense, etc. Al-Duwairi et al. [4] proposed a security information and event management-based IoT botnet DDoS attack detection and helped system. Monitoring packets are TCP SYN, ICMP, DNS packets that were used by the system to detects and block DDoS attack traffic from the compromised IoT devices.

In late 2016, Internet was stormed by large high-profile targets with DDoS attacks. Antonakakis et al. [5] provided a quick analysis of the genesis of a botnet, and the type of classes related to the devices were being affected and the way its variants evolved to compete for vulnerable devices provided alongside a seven-month revised analysis done on the Mirai's growth to a peak of 600K infections and also analysis done on the history of its DDoS victims. Poor security practices were the most reason for the emergence of Mirai and have become a serious concern for safety, security, and the privacy of IoT devices.

The proposed algorithm experiment results expressed a good performance to strengthen the safety and security of the IoT with multiple unprotected devices. Yin et al. [6] proposed a structure for software-defined IoT with the help of an SDx pattern. Elaborated an algorithm for detection and interception of DDoS attacks based on SD-IoT. Boundary switches of SD-Iot are used to confirm whether the DDoS attack happened, finding the source of attack, and attacker behind the DDoS attack.

3 Methodology

In this paper, we are implementing our model on the DDoS dataset. The DDoS dataset consists of 1,048,555. We observed that attacks are used in a single dataset. In this paper, we consider only 1000 rows and 47 columns to implement our model. In our model building, we used Python language code implementation to specify IoT attacks of the dataset. We started our implementation by import libraries and defining functions for plotting. We using matplotlib for graphical representation of data as shown in Fig. 1.

Counting on the data, not all plots are getting to be made. An Axes3D object is formed a touch just like the other axes using the projection = "3D" keyword. Create

Fig. 1 Graphical representation of the column distribution

a replacement matplotlib. Figure and add replacement axes thereto of type Axes3D. Standardize the related type of dataset along any axis. Center to the mean and the component-wise scale to the unit variance the axis wants to compute the means and standard deviations along.

After importing the plot, they a graphical technique for knowledge of the set, usually as a graph showing the connection between either two or more variables. The design is usually drawn in both the cases either hand or by a computer. Within the past, they used sometimes in mechanical, or electronic plotters were commonly used. Reading the info from CSV could even be an indispensable essential in data science. Repeatedly, we get data from to varied sources which may we get exported to comma separated values (CSV) format so as that they are getting to be employed by the opposite systems. The Pandas library provides features using which we will read the CSV enter full also as in parts for fewer than a specific group of columns and rows. The read_csv province of the pandas library is employed by reading the content of a CSV file into the Python environment thanks to the pandas DataFrame. The function can read the files from the OS by using the proper path to the file. After showing, there is a 1 CSV enter the present version of the datasets. Subsequent hidden code cells define functions for plotting data after the code was connected to the importing libraries within the dataset. For the representation of the whole data within the graphical format, we use histogram representation.

IoT botnet is a network where devices are connected via the Internet of things (IoT). The DDoS attacks commonly attempt a malicious script to spoil the normal target on the target server. They have the types of attacks, and they are reflection attack and exploitation attack. The reflection attack has three types of attacks, such as TCP-based attack, TCP/UDP-based attack, and UDP-based attack. TCP-based attack has used Microsoft Structured Query Language MSSQL and simple service discovery protocol (SSDP). TCP/UDP-based attacks may happen with the help of

Fig. 2 DDoS taxonomy

Domain Name System (DNS), Lightweight Directory Access Protocol (LDAP), Network Basic Input/Output System (NetBIOS), Simple Network Management Protocol (SNMP), or PORTMAP. UDP-based attacks may be using network time protocol, the trivial file transfer protocol, the character generator protocol (CharGen) to attack the system. Exploitation attack is also divided into two classes—TCP-based attack which is using SYN flood and UDP-based attacks which is using UDP flood and UDP-lag. (Fig. 2).

4 Experiment and Result Analysis

Figure 3 shows results are detected dataset abnormal traffic and DDoS attack. We can also see the Fig. 3 anomaly detected the data correlation matrix. It calculates the Pearson coefficient of correlation (abbreviated the r), with each pair of quantitative features within the data receiving one number from -1 to 0 to $+1$, indicating that their linear dependence is strongly negative, absent, or strongly positive, respectively.

We can see that within the Fig. 4, each numerical attribute is plotted in rows and in columns. The diagonals are showing the histograms (if it is an equivalent column on the x and y-axis). The other charts are showing the scatter plots for every combination of the numerical variables within the dataset.

Fig. 3 Correlation matrix representation using DDoS attacks

5 Conclusion

In this paper, the DDoS botnet attack data enters detection algorithm based on the pre-processing of a dataset predicted method has been investigated. We implemented the plat graphs also can detecting an inconsistency caused any by a DDoS flooding attack. We achieving high detection on showing results.

Fig. 4 Scatter matrix representation of DDoS dataset

References

1. I. Ali, et al., Systematic literature review on IoT-based botnet attack. IEEE Access (2020)
2. C.D. McDermott, F. Majdani, A.V. Petrovski, Botnet detection in the internet of things using deep learning approaches. in *2018 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2018)
3. K. Sonar, H. Upadhyay, A survey: DDOS attack on Internet of things. Int. J. Eng. Res. Dev. **10**(11), 58–63 (2014)
4. B. Al-Duwairi et al., SIEM-based detection and mitigation of IoT-botnet DDoS attacks. Int. J. Electr. Comput. Eng. (2088-8708) **10**(2) (2020)
5. M. Antonakakis et al., Understanding the Mirai botnet. in *26th {USENIX} Security Symposium ({USENIX} Security 17)* (2017)
6. Da. Yin, L. Zhang, K. Yang, A DDoS attack detection and mitigation with software-defined Internet of things framework. IEEE Access **6**, 24694–24705 (2018)

Text/Sign Board Reading Aid for Visually Challenged People

Srikhakollu Pandu Ranga Avinash, V. Krishna Sree, M. Saahith Reddy, K. Sravani Kumari, and A. Gaurav

Abstract The technological advancements in today's world are focussing on designing smart devices that improve lifestyle of human beings by making things easier and simple. Based on this, a productive approach is proposed for detecting the text from the natural scenes and converting it into voice output. The aim of this project is to overcome the reading problems of visually impaired people and their complete dependence on Braille code. The project is developed using MATLAB software in which the algorithms used are maximally stable extremely regions (MSER), stroke width transform, optical character recognition (OCR), and speech synthesizer of MATLAB. By utilizing these techniques, the text from the images can be identified effectively, and it can be converted into speech output.

Keywords Maximally stable extremal regions · Optical character recognition · Stroke width transform · Bounding boxes · Speech synthesis

1 Introduction

Communication plays an important role in the life of a human being whether it is in audio or written format. Most importantly, text that is present visually carries important and valuable information that guides people when they go out. Hence, the advancements in the technology can be applied to digitalize the images to extract specific data from the images. As a part of this, we can extract the data from the images that contain useful information using specific text extraction techniques. Text extraction can be combined with text-to-speech synthesis (TTS) which can convert the visible text in audio output [1]. A model developed in this process can be utilized by the visually impaired people to read the text from the natural scenes.

S. Pandu Ranga Avinash (✉) · V. Krishna Sree · M. Saahith Reddy · K. Sravani Kumari · A. Gaurav
Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Bachupally, Nizampet, Hyderabad, Telangana 500090, India
e-mail: avinashkrishna37@gmail.com

V. Krishna Sree
e-mail: krishnasree_v@vnrvjiet.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_4

2 Literature Review

The literature survey for this project was carried out by studying and analyzing a variety of paper publications published by different authors at national and international level based on the preferred subject. A detailed survey of the papers includes

(1) Zafar Khan, Ismail Taibani, and Soheil Sayed M. H. Sabo of Siddiq College of Engineering have proposed a finger reader system where the text will be scanned by the device, and the person will be able to hear those words as synthesized speech. This device lacks in providing accuracy under low-light conditions.
(2) Professor Priya U. Thakre, Kote Shubham, and Shelke Om developed an assistance system for the blind people that puts forward a system which allows blind people to detect and avoid hurdles/obstacles was implemented as an android app. The limitations of the system are that the stability of glasses and using an android phone was weak.
(3) Rupali D. Dharmale proposed a text recognition model for visually challenged person in Int. Journal of Engineering Research and Applications. Through this paper, they briefly discussed about the object identification techniques and image processing that are used in fields of projects for visually impaired people.
(4) K.G. Krishnan, C.M. Propodi, and K. Kanimozhi have proposed an image to speech converter in which they have described a model using which a blind person can get information about the shape of a given image input.
(5) Hangrong Pan, Chucai Yi, and Yingli Tian have developed a AI model that is capable to recognize the information of an incoming bus. This system notifies the blind person about the bus details with the help of a speech signal.

3 Proposed System

The proposed model makes use of text which is present in physically available images and converts it into speech that can be heard by visually impaired person. In order to achieve this goal, various objectives are needed to be accomplished as listed below.

1. Taking the input image and converting the image into greyscale format from RGB format.
2. Detection of the text regions from the greyscale image by implementing maximally stable external regions (MSER).
3. Removal of non-text regions based on geometrical properties.
4. Removal of the non-text regions present in the image by using stroke width transform (SWT).
5. Expanding the bounding boxes around the identified letter candidates.
6. Applying optical character recognition (OCR) to recognize the identified areas.

Fig. 1 Block diagram of proposed method

7. Implement the speech synthesizer to receive a speech signal as voice output of the identified text.

Various stages that are involved in the implementation of the proposed model are shown in Fig. 1.

4 Implementation

The implementation of the above model starts by giving an image from the natural scene as the input to the system and converting the image to a greyscale format for increasing the processing efficiency. After the RGB image is converted to greyscale image, character identification techniques are applied for identifying and detecting the text from the given image. As a part of this, maximally stable extremal regions and stroke width transform are applied.

4.1 Maximally Stable Extremal Region

Maximally stable extremal regions or MSERs are a blob detection technique that is used to identify the text regions from the given input image. This technique uses the

strategy of identifying specific locales in the input image that vary in certain properties when compared with encompassing district locations. It is based on identifying all the regions that stay constant or nearly same throughout a wide range of threshold values [2], i.e. all the pixels present inside the extremal regions have an intensity value either low or higher than the pixel intensities outside the given region.

4.2 Text Region Localization

In general, images have different components that resemble with a text or alphabetical format which the model can identify. This leads to an invalid input when compared with the original output. To eliminate this kind of errors, there is a necessary to remove the non-text regions that appear to be as a text region. These non-text regions can be eliminated based on few geometrical properties which measures the maximally stable extremal region (MSER) properties for each connected component or object in the binary image and removes them. In MATLAB, by using the command "regionprops", we can measure the properties of the image region. This command measures a set of properties for each connected component in the binary image. The properties used in this thesis are "bounding box", "eccentricity", "solidity", "extent", "Euler", and "image".

4.3 Stroke Width Transform

Stroke width transform (SWT) is an algorithm that is designed to compare the stroke of a regions with the strokes of the other regions contained in the background. Here, a stroke refers to group of values that show uniform width throughout. The image that is produced as an output of the stroke width transform has a size which is equal to the input image where each element has the value of the width of the pixel associated. The main purpose of using stroke width transform is to eliminate all the non-text regions that are detected as a part of text detection [3].

The calculation gets greyscale picture as input and returns comparable picture of a similar size, where the areas of suspected content are checked. It has 3 significant advances: the stroke width change, gathering the pixels into letter competitor's dependent on their stroke width, and lastly, gathering letter applicants into districts of text. Here, Fig. 2a shows the input image. The stroke width of the input image is calculated as shown in the Fig. 2b.

Fig. 2 Example of stroke width transform

4.4 Expanding Bounding Boxes

After recognizing the text and elimination of non-text regions, we consolidate the text regions for definite detection result. For this a rectangular box is plotted around each character for object identification [9]. This is done by expanding bounding boxes around each letter candidates and grouping them into words and sentences. By defining the X and Y coordinates, a bounding box is drawn around the letter which acts as a point of reference for the process of text detection. After creating the bounding boxes, the individual letter candidates have to be merged together to form group of words or a meaningful sentence. This is done by calculating the overlap ratio. In MATLB, we have a function “bboxOverlapRatio” which gives the overlap ratio that is present between every two bounding boxes that are considered. The overlap ratio has a value between 0 and 1 where 1 represents that the overlap is perfectly done. In the next step, we combine the bounding boxes according to the overlap ratio as well as maximum and minimum dimensions.

4.5 Optical Character Recognition

After identifying the text areas and removing the non-text areas, the next step is to detect the text characters that are to be received as voice output. This done by implementation of optical character recognition (OCR). It is a process of recognizing printed or hand-written text and converting them into machine-encoded text. In the proposed technique, we utilize an optical character recognition function that accepts the input as text boxes that are previously made through bounding boxes. This process produces the extracted text from the input image as a text document [4].

4.6 Speech Synthesis

After the text output is received, we need to implement a speech synthesizer to convert the identified text into voice output. Text-to-speech (TTS) is an algorithm that is designed to provide a speech signal from the given text data. By implementing text to speech, we can easily convert the text into speech and audio signal. MATLAB consists of an audio tool box which can used to implement the speech synthesizer that can convert the input text data to audio format.

5 Results

The system takes a maximum time of 4–5 s in processing all the output images and giving an audio output. The results of the above implemented algorithms in MATLAB are as follows:

1. The output image after transforming the input image to greyscale format is given in the Fig. 3.
2. The output image of the second stage is an image with maximum stable extremal regions (MSER) identified. Figure 4 shows maximum stable extremal regions that are identified in the given input image.

Fig. 3 Input image and greyscale image

Fig. 4 Image with identified MSER

3. Figure 5 shows the image with non-text regions removed based on geometrical properties. This image had removed maximum number of non-text regions from the set of regions identified by MSER.
4. Few of the non-text regions that are still visible in the input image are further removed by employing stroke width variation. This output related to stroke with variation is shown in the Fig. 6.
5. The output image with bounding boxes expanded around the letter candidates is shown the Fig. 7.
6. The output image with detected text after the letter candidates are merged into words and sentences is shown in Fig. 8.

Fig. 5 Removing non-text regions based on geometrical properties

Fig. 6 Removing non-text regions based on stroke width variation (SWT)

Fig. 7 Image with expanded bounding boxes

Fig. 8 Image with detected text after merging of bounding boxes

Fig. 9 Identified text output

7. The recognized text from the input image by using optical character recognition is shown in the Fig. 9.
8. Also, with the help to speech synthesizer used in the programme, the identified text output is converted into a speech signal, and an audio output is generated.

6 Conclusion

In this project, we developed a system which can convert text or image into speech by using MATLAB software. To the best of researcher's knowledge, the first step was to

take an input image and perform image pre-processing. Segmentation of text images is done using maximally stable extremal regions. They are made unique by taking various thresholds. Then, the individual letter candidates are recognized. Later on, these individual letters are grouped to form the words. For word recognition, optical character recognition (OCR) is implemented, and speech processing toolboxes are used to convert text into speech.

As a part of future scope, this concept of text-to-speech conversion for blind people can be extended by integrating the model with Internet of things (IoT) so that it can be connected to various devices such as a smart watch.

References

1. C. Yi, Y. Tian, Scene text recognition in mobile applications by character descriptor and structure configuration. IEEE Trans. Image Process. 7 (2014)
2. S.M. Qaisar, A computationally efficient EEG signals segmentation and de-noising based on an adaptive rate acquisition and processing. In: *2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP)* (IEEE, 2018), pp. 182–186
3. A. Kumar, S. Gupta, Detection and recognition of text from image using contrast and edge enhanced segmentation and Ocr. IJOSCIENCE (Int. J. Online Sci.) Impact Factor **3**(3), 3 (2017)
4. Y. Zhu, C. Yao, X. Bai, Scene Text Detection in Technological Advancements, 10(1), 19–36 (2016)

Automatic Segmentation of Red Blood Cells from Microscopic Blood Smear Images Using Image Processing Techniques

K. T. Navya, Subhraneil Das, and Keerthana Prasad

Abstract Human blood is a very effective parameter to detect, diagnose and rectify ailments of the human body. Complete blood count (CBC) is a method to clinically obtain a statistical measure of blood and its related parameters, i.e., red blood cells (RBCs), white blood cells (WBCs), platelets, hemoglobin concentration to name a few. This helps to determine the physical state of the subject. For further diagnosis, peripheral blood smear, a thin layer of blood smeared on a microscope slide and stained using various staining methods is examined for the morphology of the cells by the pathologists. However, manual inspection of smear images is tedious, time-consuming, and laboratorian-dependent. Although there are certain software-based approaches to tackle the problem, most of them are not robust for all staining methods. Thus, the need is to create an automated algorithm that will work for different staining types, thereby alleviating both the aforementioned drawbacks. This work aims to create an automatic method of segmenting and counting RBCs from blood smear images using image processing techniques to help diagnose RBC-related disorders. In the proposed method, the images are first preprocessed, i.e., standardized to a uniform color and illumination profile using contrast enhancement, adaptive histogram equalization followed by Reinhard stain normalization algorithms. WBCs and platelets are extracted in HSI color space and subtracted from the original image to retain only RBCs. Thereafter using morphological operations and active contour segmentation algorithms, a count of total RBCs were obtained even for overlapped cells in the microscopic blood smear image. The proposed method achieved counting accuracy of 89.6% for 150 images.

K. T. Navya · S. Das
Department of Electronics and Communication, Manipal Institute of Technology, Manipal, India

K. Prasad
Manipal School of Information Sciences, Manipal, India
e-mail: keerthana.prasad@manipal.edu

K. T. Navya (✉) · S. Das · K. Prasad
Manipal Academy of Higher Education, Manipal 576104, India
e-mail: navya.kt@manipal.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_5

Keywords Red blood cells · Color normalization · Image processing · Segmentation · Active contour

1 Introduction

Blood sampling and diagnosis of diseases based on the blood count and characteristics are an important aspect of medicinal science and engineering. Human blood is an essential source for developing diagnosis patterns and antigens for numerous diseases. Blood is a specialized body fluid and has four main components [1].

- *Plasma*, liquid component produced in the bone marrow to transport blood cells throughout the body
- *Erythrocytes or red blood cells* is a biconcave disk with a central pallor, 6–8μm in diameter with pink stain helps carry oxygen from the lungs to the rest of the body.
- *Leukocytes or white blood cells*, largest cells, varies between 8 and 20μm in diameter, a dark purple nucleus with pale cytoplasm produce antibodies to fight off infection.
- *Thrombocytes or platelets*, small cell fragments without nuclei, 2–3μm in diameter, bluish in color helps in blood clotting.

Figure 1 shows a typical blood film or peripheral blood smear stained using the Romanowsky staining method to examine the blood cells microscopically. Staining plays a very important role in the analysis of blood smear images as the color profile, contrast, and feature details in a microscopic blood smear image are determined initially by this aspect. Popular staining methods under Romanowsky staining include Giemsa, Wright-Giemsa, Leishman stain, etc. The blood smear thus taken is analyzed in the feather edge of the slide, and hence, subsequent conclusion helps in knowing the current state of health of the patient as well as diagnosing a specific ailment. Considering the manual techniques followed previously for analyzing blood smear images, it was very tedious and not reliable. Thus, this work is aimed at developing

Fig. 1 Blood smear images [2] **a** microscopic view of blood smear **b** blood smear slide

an autonomous and robust system for image preprocessing followed by segmentation of red blood cells in a given blood smear image using image processing techniques in MATLAB to help diagnose RBC-related disorders.

2 Related Work

This section attempts at giving a brief overview of various past developments in the field of automated blood counting and its related parameters such as color normalization, preprocessing, RBC segmentation, background, and foreground differentiation using various segmentation methods based on underlying principles of image processing. Sharif et al. [3] elaborated a robust methodology for RBC segmentation using masking and morphological operation. This paper has a combination of pixel-based, region-based, and morphological segmentation. YCbCr color has been chosen for illumination issues. A marker-controlled watershed algorithm was used to handle overlapping cells from 20 images. Mazalan et al.[4] proposed an automated method to count RBCs in microscopic images using circular Hough transform (CHT) for 10 sample images and obtained 91.9% accuracy. This method is cost-effective and provides an alternative way to recognize and count circular cells. Alomari et al. [5] proposed an iterative structured circle method to segment WBC and RBC, and average accuracy of 95.3% for RBCs and 98.4% for WBCs was achieved for 100 images. However, this method depends on the number of iterations. Abbas et al. [6] presented a method for RBC segmentation in YCbCr color space. K-means clustering was used to identify cells from 90 Giemsa stained images. Tomari et al. [7] proposed a Hough transform (HT) method to count overlapped RBCs and obtained 94% average accuracy for four sample images. However, this method involved many parameters. Alam et al. [8] presented the YOLO algorithm to identify and count blood cells and obtained accuracy 96.09% RBC, 86.89% WBC, and 96.36% for 364 annotated images. Wei et al. [9] proposed K-means clustering-based method to segment and count overlapped RBCs. The method used the H and S components to differentiate between WBC and RBC. The author of the work obtained 92.9% accuracy for 100 Wright-Giemsa stained images. Acharya et al. [10] presented a method to identify and count RBC using K-Medoids and geometric features. This method achieved 98% accuracy for 1000 Wright stain images. Ejaz et al. [11] proposed HT-based method to segment and count RBCs and obtained 94.9% accuracy for 500 subjects. Berge et al. [12] proposed RBC segmentation method using boundary extraction and curvature calculation. The Delaunay triangulation method was used to split overlapped RBCs of any shapes. This method obtained 2.8% absolute error for 49 Giemsa stained images. Hegde et al. [13] proposed active contour method to segment WBCs. G'G'B channel representation for handling illumination and color variations and obtained 96% sensitivity for 54 images from the ALL-IDB2 dataset. Adagale et al. [14] proposed an overlapped red blood cell counting algorithm using template matching and pulse coupled neural network and achieved 90% average accuracy for 40 images. However, accuracy decreases due to overlapped RBCs. Cruz et al. [15] presented RBC count-

ing method using blob analysis and watershed transform in the HSV component and obtained 96% average accuracy for 10 blood samples. Loddo et al. [16] proposed a blood cell counting method using nearest neighbor and SVM techniques by cropping each cell manually. Clumped cells were counted using CHT. This method used 368 images from the ALL-IDB dataset and obtained an average accuracy of 99.2% for WBCs and 98% for RBCs. Yeldhos et al. [17] implemented FPGA-based RBC counting system. Watershed transform and CHT segmentation method were used in YCbCr color space and obtained 90.98% accuracy for 108 blood smear images from the ALL-IDB database. Tran et al. [18] presented deep learning semantic segmentation method for RBC and WBC segmentation and counting. SegNet architecture was utilized to segment blood cells by labeling each pixel. The segmentation accuracy for WBCs and RBCs was 94.93% and 91.11%, respectively. For cell counting, Euclidean distance transform and binary dilation are used and obtained 93.3% for RBC and 97.29% for WBC for 42 ALL-IDB database images. From the past studies, different methods such as circular Hough transform, watershed transform, morphological operations, thresholding-based methods, K-means clustering, ANN, DNN have been used for RBC segmentation and counting. However, these methods lack robustness in handling blood smear images with multiple stains. Hence, there is a need for developing a robust system to handle images taken from various laboratory settings.

3 Methodology

This section gives a sequence of steps that are to be followed to obtain the desired results. A methodology of the work is depicted in Fig. 2.

3.1 Image Acquisition

The required Leishman stained blood smear images are acquired from Kasturba Medical College (KMC), Hematology department, Manipal, with a 100x lens objective. Also, smear images from Isfahan MISP online database[19] are gathered for the process.

Fig. 2 Methodology of proposed work

3.2 Preprocessing

Due to the varied color profile and contrast of the microscopic smear images, there is a need to standardize the image to obtain a consistent color profile. To make an image ready for the automatic segmentation under various image settings and stains, illumination correction and color normalization methods have been implemented. The various preprocessing methods used are discussed further [20].

- *Linear Contrast Enhancement using Adaptive Histogram Equalization*
  This method is helpful to reduce illumination variation. The images are first converted from RGB to LAB color space where a consistent normalization of the L parameter (Luminosity) is performed followed by grayscale conversion and subsequent adaptive histogram equalization.

- *Gamma Correction*

  is a nonlinear method used to correct the image's luminance; i.e., it amplifies the shadows or the bright regions of the image as per the requirement. It is used to correct uneven illumination by encoding luminance in video or still image systems. Gamma correction is defined as:

$$V_{out} = AV_{in}^{\Gamma} \tag{1}$$

  Powers larger than 1 make the shadows darker and smaller than 1 make dark regions lighter.

- *Gray World Assumption method*

  It is a color correction function based on the principle that on average the scene is neutral gray. The algorithm produces an estimate of illumination by computing the mean of each channel of the image. The average pixel value of an unsigned 8-bit integer image is 127.5, so by calculating the real average pixel value, the scaling value is computed to scale the entire image linearly. In practice, the average of each individual channel is used to calculate a separate scaling value for each channel.

- *Histogram Matching*

  It adjusts histogram of the 2D image to match histogram of reference image in order to normalize the color. The reference RGB image is converted into HSV, and hue component is extracted and used for histogram matching with the input image to segment WBC and platelets. The HSV image is then converted into RGB image for further processing.

- *Reinhard's Stain Normalization*

  The algorithm thus maps the color distribution of an image to that of a well-stained target image thereby solving the problem of inconsistency. Technically it matches the mean and the standard deviation of each color channel in the two images in that color space [21]. The statistical approach to color mapping can be shown as in equations

$$l_{mapped} = \frac{l_{original} - \tilde{l}_{original}}{\hat{l}_{original}} \hat{l}_{target} + \tilde{l}_{target}$$

$$\alpha_{mapped} = \frac{\alpha_{original} - \tilde{\alpha}_{original}}{\hat{\alpha}_{original}} \hat{\alpha}_{target} + \tilde{\alpha}_{target}$$

$$\beta_{mapped} = \frac{\beta_{original} - \tilde{\beta}_{original}}{\hat{\beta}_{original}} \hat{\beta}_{target} + \tilde{\beta}_{target} \tag{2}$$

  where the superscripted variables $\tilde{l}$ and $\hat{l}$ represent the mean and standard deviation of luminosity parameter, respectively, in the LAB color space. Similarly, the subscripted variables $l_{original}$, l_{target}, and l_{mapped} indicate whether the parameter concerned is either of the original, target or the mapped image of the LAB color space parameters. An image with acceptable color profile, contrast, and illumination was chosen as the target image, and the source image was subjected to Reinhard staining normalization.

- *Noise Removal*

  The blood smear image post color normalization is subjected to noise removal and deblurring using Wiener filter. The filter processes a two-dimensional image and eliminates granular noise from the image. It is also useful in removing motion blur from the image, thereby making the image ready for further processing.

The image enhancement algorithms on being implemented on a set of images gave a series of results based on their resolution, contrast standardization, and processing method as stated in this section. In Fig. 3, the original faded Leishman stained microscopic image of a normal blood smear is shown which is almost devoid of sharpness, contrast, and color and the reference image for color normalization. The contrast-enhanced reference image is exposed to gamma correction for illumination correction and then passed through a Wiener filter with a filtering window of [10]. The resultant images are smoothened and exhibit a much less extent of jagged edges in the features of the image. The input image is subjected to grayworld color normalization in order to correct uniformity of color, contrast, and illumination variation. However, results obtained show consistency in image regions with proper contrast and deviation for other regions thereby proving to be inefficient for the required purpose. The images are then subjected to the statistical means of color matching in order to achieve a fixed level of uniform tone and contrast which can then be applied and fixated to any image which is fed as an input to the algorithm. The target image ideally should possess uniformity in regard to the factors of illumination, color tone, contrast, warmth,

Fig. 3 Preprocessed blood smear images **a** Leishman reference image **b** contrast-enhanced image **c** gamma correction for image **d** Wiener filtered image **e** Leishman stained input image **f** grayworld normalized image

Fig. 4 **a** H component of input image **b** H component of preprocessed reference image **c** histogram matched RGB image

and sharpness. Histogram matching technique is used for color normalization and illumination correction where hue component of the reference image is matched to hue component of the source image to extract WBCs and platelets from blood smear images, and the matched RGB image is as shown in Fig. 4. In order to compare the histogram matched image with another color normalization method, Reinhard stain normalization was applied for source image with reference to the target image. The normalized results obtained with different stain input and target images taken in 100x

Fig. 5 **a** Leishman and Wright stain input image **b** 100× and 40× reference image **c** color normalized image

and 40X magnification are shown in Fig. 5. On visualization, Reinhard normalized image combat both illumination and color variation and show a prominent difference between the blood cells.

3.3 Segmentation of Blood Cells

The normalized image is processed further for WBC, platelets, and RBC differentiation. Thus, the Reinhard normalized image is subjected to unsharp masking for sharpening the object boundaries to distinct with respect to the background, especially at the edges of the RBCs. WBCs are extracted using the S-channel of HSV that highlights the nucleus and green channel of RGB color space which highlights RBCs and cytoplasm of WBC. Then by combining both the channel outputs, WBCs are removed. Further, traces of WBCs are removed using morphological opening via a disk-shaped morphological element with the radius range of 80–90, and then similarly, platelets are removed, subtracted, and smoothened to obtain the image with only RBCs. The appropriate results are shown in Fig. 6.

We can observe from the RBC segmented image that due to morphological operations, boundaries of RBCs are altered. For further RBC disorder analysis, the morphology of the cell is very important. So to improve on this diverging active contour segmentation algorithm is used. The basic premise of active contours is energy minimizing models; i.e., the snake tends to minimize its energy, thereby shrinking onto the boundaries of the image objects[22]. The entire process of fitting onto the bound-

Fig. 6 **a** Input image **b** reinhard normalized image **c** sharpened image **d** S-channel of original image **e** green channel extraction **f** extracted WBC image **g** subtracted image **h** image smoothened by morphological dilation **i** resultant image after erosion

aries occurs in several iterations to get the best fit possible. Now, considering the snake as a continuous parametric variable, we can define its position in the image as

$$v(t) = (x(t), y(t)) \tag{2}$$

where $v(t)$ is the active contour and $x(t)$, $y(t)$ are the continuous contour coordinates. Thus, the energy equation associated with active contours is as follows:

$$E = \int_0^1 [E_{\text{int}}(v(t)) + E_{\text{img}}(v(t)) + E_{\text{con}}(v(t))]\mathrm{d}t \tag{3}$$

where E_{int} is internal energy, E_{img} is image forces and E_{con} is external constraint force. The divergence of the contour was achieved by exercising a certain degree of *contraction bias*. This parameter ranges from -1 to 1, and a negative value indicates expanding contour, whereas a positive value indicates a shrinking contour. For the initial contour mask, centroids of the morphologically segmented RBCs are taken as a seed point to diverge. Figure 7 depicts an initial mask and contour detected RBC binary image as an overlay mask to present the detected cells clearly.

Fig. 7 **a** Initial contour mask **b** RBC segmented image **c** Detected RBCs along with their shape enclosed within binary mask

4 Results and Discussion

In the case of any algorithm, a reference or ground truth is required to measure the accuracy of the used algorithm. The segmented RBCs are counted manually and compared with the ground truth. Here for ground truth formation, we make use of ImageJ Tool to segment the microscopic blood smear images to get a clear count of the RBCs present in the image. Clustered or ambiguous cell boundaries are drawn using a freehand line tool to obtain clear cell boundaries and count. Now using the count obtained from the ground truth of the images and the count obtained via our proposed algorithm we determine the accuracy of the method. In this work, accuracy is determined as the percentage of detected count divided by the actual count of RBCs present in that image. A total of 150 images, 75 Leishman stained images from KMC and 75 images from the online database [19] were used in the study. The active contour algorithm yields an overall accuracy of 89.6% for 150 images. Though it resolves a significant amount of overlapped cells present in an image, some overlapped cells and clusters still remain unresolved. Still, there is a need for a robust segmentation method and accurate ground truth for RBC count to achieve higher accuracy.

5 Conclusion

The methods implemented to meet the objective of the work has been achieved so far. Preprocessing of the images proved to be a cumbersome task and after several trials, Reinhard's method demonstrated to be the most effective method of color normalization. Similarly, post-WBC and platelet removal, the active contour model provided an RBC count closest to the actual values as obtained from the ground truth and achieved an overall accuracy of 89.6%. However, to overcome overlapping and densely clustered RBCs, a robust segmentation algorithm has to be implemented to get a much more accurate count. Further, segmentation processes using convolutional

neural networks and artificial intelligence toolboxes can be used to create a shape classifier for identifying cell edges, overlapped cells, and cell clusters accurately in order to detect specific diseases and anomalies.

References

1. H. Mohan, *Textbook of Pathology* (Medical Publishers Pvt. Limited, Jaypee Brothers, 2018)
2. K.W. Jones, Evaluation of cell morphology and introduction to platelet and white blood cell morphology. Clin. Hematol. Fundam. Hemost. 93–116 (2009)
3. J.M. Sharif, M. Miswan, M. Ngadi, M.S.H. Salam, M.M. bin Abdul Jamil, Red blood cell segmentation using masking and watershed algorithm: a preliminary study, in *2012 International Conference on Biomedical Engineering (ICoBE)* (IEEE, 2012), pp. 258–262
4. S.M. Mazalan, N.H. Mahmood, M.A.A. Razak, Automated red blood cells counting in peripheral blood smear image using circular Hough transform, in *2013 1st International Conference on Artificial Intelligence, Modelling and Simulation* (IEEE, 2013), pp. 320–324
5. Y.M. Alomari, S. Abdullah, S.N. Huda, R. Zaharatul Azma, K. Omar, Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm. Comput. Math. Methods Med. **2014** (2014)
6. N. Abbas, D. Mohamad et al., Microscopic RGB color images enhancement for blood cells segmentation in YCBCR color space for k-means clustering. J. Theor. Appl. Inf. Technol. **55**(1), 117–125 (2013)
7. R. Tomari, W.N.W. Zakaria, R. Ngadengon, M.H.A. Wahab, Red blood cell counting analysis by considering an overlapping constraint 2006–2015. Asian Res. Publishing Netw. (ARPN) **10**(3) (2015)
8. M.M. Alam, M.T. Islam, Machine learning approach of automatic identification and counting of blood cells. Healthcare Technol. Lett. **6**(4), 103–108 (2019)
9. X. Wei, Y. Cao, G. Fu, Y. Wang, A counting method for complex overlapping erythrocytes-based microscopic imaging. J. Innov. Optical Health Sci. **8**(06), 1550033 (2015)
10. V. Acharya, P. Kumar, Identification and red blood cell automated counting from blood smear images using computer-aided system. Med. Biol. Eng. Comput. **56**(3), 483–489 (2018)
11. Z. Ejaz, A. Hassan, H. Aslam, Automatic red blood cell detection and counting system using Hough transform. Indo Am. J. Pharm. Sci. **5**(7), 7104–7110 (2018)
12. H. Berge, D. Taylor, S. Krishnan, T.S. Douglas, Improved red blood cell counting in thin blood smears, in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (IEEE, 2011), pp. 204–207
13. R.B. Hegde, K. Prasad, H. Hebbar, B.M.K. Singh, Image processing approach for detection of leukocytes in peripheral blood smears. J. Med. Syst. **43**(5), 114 (2019)
14. S. Adagale, S. Pawar: Image segmentation using PCNN and template matching for blood cell counting, in *2013 IEEE International Conference on Computational Intelligence and Computing Research* (IEEE, 2013), pp. 1–5
15. D. Cruz, C. Jennifer, L.C. Castor, C.M.T. Mendoza, B.A. Jay, L.S.C. Jane, P.T.B. Brian et al., Determination of blood components (WBCs, RBCs, and platelets) count in microscopic images using image processing and analysis, in *2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)* (IEEE, 2017), pp. 1–7
16. A. Loddo, L. Putzu, C. Di Ruberto, G. Fenu, A computer-aided system for differential count from peripheral blood cell images, in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (IEEE, 2016), pp. 112–118
17. M. Yeldhos, Red blood cell counter using embedded image processing techniques. Res. Rep. **2** (2018)

18. T. Tran, O.H. Kwon, K.R. Kwon, S.H., Lee, K.W. Kang, Blood cell images segmentation using deep learning semantic segmentation, in *2018 IEEE International Conference on Electronics and Communication Engineering (ICECE)* (IEEE, 2018), pp. 13–16
19. M. Kashefpur, R. Kafieh, S. Jorjandi, H. Golmohammadi, Z. Khodabande, M. Abbasi, N. Teifuri, A.A. Fakharzadeh, M. Kashefpoor, H. Rabbani, Isfahan MISP dataset. J. Med. Signals Sens. **7**(1), 43 (2017)
20. Image enhancement techniques. https://in.mathworks.com. Accessed May 2020
21. E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, Color transfer between images. IEEE Comput. Graph. Appl. **21**(5), 34–41 (2001)
22. M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models. Int. J. Comput. Vis. **1**(4), 321–331 (1988)

Design and Simulation of Microstrip Bandpass Filter Techniques at X-Band for RADAR Applications

Shraddha S. Shinde, D. P. Rathod, and ArunKumar Heddallikar

Abstract A new approach to compact bandpass filter techniques such as end-coupled, parallel-coupled, hairpin, interdigital, and combline filters with their enhancement in size and frequency performance is introduced in this paper. Designed fifth-order filters that are simulated in the compact narrowband frequency is targeted for high selectivity, better return loss, and insertion loss with a passband ripple using Chebyshev response. To minimize the dimension of filters, we implemented an optimized design by determining its parameters, coupling resonators, and connecting elements. The results are simulated using Genesys Keysight software and iterated for the desired specimen. The prototype filter consists of Rogers RT/duroid 5880 substrate with a thickness of 0.508 mm. Experimental result of hairpin filter structure shows a reduced design (20.2 mm × 10 mm) and improvises the stopband characteristics with better harmonic suppression and narrower frequency bandwidth than the other conventional filter for future development in RADAR applications.

Keywords Bandpass filter · Microstrip · End coupled · Parallel coupled · Hairpin · Interdigital · Combline · Chebyshev response · RADAR

1 Introduction

Evolution in communication took place from the telegraphy in the Second World War toward Radio Frequencies (RF) in remote sensing and navigation. RF filters are required in the telecommunication system to reduce noise and interference of signal

S. S. Shinde (✉) · D. P. Rathod
Department EnTC, VJTI (Veermata Jijabai Institute of Technology), Mumbai, India
e-mail: ssshinde_m18@et.vjti.ac.in

D. P. Rathod
e-mail: dprathod@el.vjti.ac.in

A. Heddallikar
RADAR Division, SAMEER, IIT B CAMPUS, Mumbai, India
e-mail: arunkumar@sameer.gov.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_6

[1]. X-band (8–10 GHz) frequencies are used because it allows smaller RF components compared to other bands of frequencies. Microstrip is unique and contains a Radio Frequency (RF) filter used in many applications such as RADAR, satellite, and wireless communications [5, 7]. A filter is made of two-port networks that are used to regulate frequency response according to the requirement by transmission lines at frequencies within passband and attenuate frequency in stopband [1, 8]. This paper illustrates the X-band filter (8–12 GHz), designed by combining two or more resonators to make a more compact size structure terminated with 50 Ω characteristics impedance [9, 10]. Chebyshev response is selected in this paper as compared to Butterworth response as it is having better selectivity and steeper roll-off with flatter passband magnitude response [3]. Some of the X-band filters such as end-coupled, parallel-coupled, hairpin, interdigital, and combline filters were reported [3, 4]. The end-coupled filter is half-wavelength resonators at mid-frequency of band-pass filter design, a gap between the two adjacent resonators. In a parallel-coupled filter, parallel arrangements of resonators along the half of wavelength resonators are specified [11]. To minimize its dimension of the half-wavelength filter, it required a hairpin microstrip line structure. A hairpin filter is compact and can be folded from the open line $\frac{\lambda}{2}$ wavelength microstrip resonator having stepped impedance to form 'U'-shaped small structure that consists of parallel-coupled lines on both sides ends for miniaturization, which is optimal space utilization [12, 13]. The interdigital filter has properties that consist of arithmetical symmetry, considered as advantageous as it correlates with combline and other filters [14–16]. The drawback of a filter is that the frequency response gets affected by coupling among non-adjacent resonators [14]. Combline filters having large lumped capacitance and shorter resonator lines form a compact structure, and the filter size is relatively reduced. To achieve this further, all five filters are compared with analysis and simulation to determine their performance [14, 15]. To miniaturize the dimension, our work focused on the conventional bandpass filter technique. In this way, we have designed filters on RT/duroid 5880 substrates of thickness 0.508 mm, (dielectric constant) $\varepsilon_{\mathrm{re}} = 2.2$, and loss tangent = 0.0009. The design of filter techniques is performed on Genesys Keysight software. The filter design is chosen in the X-band frequency such that it can be used for various RADAR applications.

2 Material and Method

The Chebyshev filter illustrates a good performance compared to the Butterworth filter for frequency response and its cutoff [3, 17]. We considered Chebyshev prototype, and element parameters are given as $g_0 = g_6 = 1$, $g_2 = g_4 = 1.3712$,$g_1 = g_5 = 1.1468$, and$g_3 = 1.9750$ [2, 3].

Fig. 1 Structure of end-coupled bandpass filter

Fig. 2 Structure of parallel-coupled bandpass filter

2.1 Design of End-Coupled Bandpass Filter

End-coupled bandpass filter Fig. 1 is half guided wavelength resonator designed in Genesys Keysight software. Figure 1 shows width = 12 mm, length = 80.5 mm, and l = 10 mm, where 'l' is the length of the resonator. SubMiniature version A connector (SMA) is connected at both ends of the microstrip line [3, 4]. Using Genesys Keysight software, width, length, and line spacing can be determined [3, 11].

2.2 Design of Parallel-Coupled Bandpass Filter

Figure 2 shows a parallel-coupled filter having a width of 27 mm and a length of 36 mm, t = 6 mm, l = 5.25 mm, and w = 1.5 mm, where 't' is the tapping length, 'l' is the length of a resonator, and 'w' width of each resonator. A parallel-coupled filter is a cascaded-line section with open ends. Electrical length is specified by even and odd mode characteristics impedance [6, 11].

2.3 Design of Hairpin Filter

The tapped input and output hairpin structure is shown in Fig. 3, with width = 10 mm, length = 20.2 mm, w = 1.5 mm, t = 3.6 mm, S1 = 0.2 mm, and S2 = 0.4 mm, where

Fig. 3 Structure of hairpin bandpass filter

'w' is the width of each resonator element, 't' is the tapping length of the hairpin filter, 'S1' and 'S2'are spacing between the two resonators. The input transmission line is not directly attached to the first resonator [12]. Decreasing the tapping length 't' increases the unloaded quality of resonators, hence narrowing the 3 dB bandwidth as well as changing the response characteristics [3, 12].

2.4 Design of Interdigital Filter

Figure 4 shows the structure of the interdigital bandpass filter modeled in Genesys software of width = 10 mm, length = 21.5 mm, l = 6 mm, w = 1.7 mm, S1 = 0.2 mm, and S2 = 0.3 mm, where 'l' is tapped line, 'w' is the width of the resonator, and 'S1' and 'S2' are spacing between the two resonators having resonator length of $\frac{\lambda}{4}$ long with tapped line at first and last resonator [15, 16]. Each resonator has an electrical length of 90° at the mid-band frequency which is short-circuited at one end and open-circuited at another end. It is shorted to ground through the via-holes [3, 15].

Fig. 4 Structure of interdigital bandpass filter

Fig. 5 Structure of combline bandpass filter

2.5 Design of Combline Filter

Figure 5 shows combline filter with width = 12 mm, length = 17.4 mm, l = 5 mm, C1 = C2 = C3 = C4 = C5 = 0.25pf, w = 1.8 mm, where 'l' is the length of the tapped line, 'C1….Cn' are the capacitance values, and 'w' is the width of the resonator. It consists of lumped capacitance at one end and short-circuited at another end. The lumped capacitors are usually used for narrowband filters [17].

3 Proposed Design

The microstrip bandpass filter design techniques designed in this paper are based on miniaturized structure and better frequency response [3, 4]. Chebyshev response is chosen for its passband characteristics. Considering these values, the dimensions of the structure are modeled based on the capacitances and inductances of the equivalent circuits [3]. The step-by-step discussion structure and simulation and its tuning of filter techniques are presented in Fig. 6. After determining the specifications of the filter, mathematical calculations are carried out using empirical formulae [2, 3]. Electromagnetic simulation is done in Genesys Keysight software. Genesys software is used for accurate frequency response of parameters S11 (return loss) and S21 (insertion loss) with good comparison performance of all microstrip bandpass filter techniques. In this context, some modifications were performed such as tunning and spacing between the resonators to get the exact required response. The required filter frequency response is done in two iterations. After iterations, the fractional bandwidth of the filter is in the range of 7.3–15.8%. After determining the prototype, the filter design is finalized.

4 Result and Discussion

Figure 7 above shows the simulated result in Genesys software. The filter structure is realized using a dielectric material RT/duroid 5880. Using Genesys software, the design is simulated, and the length, width, and spacing were calculated. At the

Fig. 6 Flowchart for microstrip bandpass filter techniques

center frequency, the insertion loss (S21) is −0.73 dB, and the return loss (S11) is −39.523 dB.

The characteristics of the proposed parallel-coupled microwave bandpass filter are shown in Fig. 8. The bandpass filter is having an insertion loss of −1.496 dB across the passband, while the passband return loss is −42.837 dB across the required band 9.3–10.7 GHz.

Figure 9 shows the simulated result of the fifth-order hairpin bandpass filter. Comparing these results with end coupled and parallel coupled above in Fig. 7 and Fig. 8, it is noticeable that results are better for the following parameters. The insertion loss (S21) is −0.676 dB, and the return loss (S11) is −47.986 dB with a dimension reduction. Improved response characteristics are the reason why tapped structures are used. Hence, the hairpin bandpass filter is more preferred as compared to other filters.

Based on the above simulation, there is a minimum insertion loss of −0.795 dB and return loss of −41.275 dB at 10 GHz at the center frequency. It is observed that the

Fig. 7 Simulated result of fifth-order end-coupled bandpass filter

Fig. 8 Simulated result of fifth-order parallel-coupled bandpass filter

filter dimension is a bit larger than the hairpin filter. The frequency response increases slightly juxtapose to the hairpin filter, but it is much better than the end-coupled and parallel-coupled filter with a minimized dimension.

Figure 11 shows the simulation result that presents good S21 and S11 curves for the bandpass filter response. The center frequency is 10 GHz with an insertion loss of −0.597 dB and a return loss of −33.431 dB. It is observed that return loss and insertion loss have improved considerably with the miniaturized design when compared to other filters.

Fig. 9 Simulated result of fifth-order hairpin bandpass filter

Fig. 10 Simulated result of fifth-order interdigital bandpass filter

Table 1 shows the comparison of techniques such as end-coupled, hairpin, interdigital, and combline bandpass filters. It is observed that hairpin filter structure and combline filter are more compact with a great improvement in frequency response and dimensions. This shows that the design can be easily developed to handle and permit configurability which can be easily integrated with various applications.

Fig. 11 Simulated result of fifth-order combline bandpass filter

Table 1 Comparison of end-coupled, parallel-coupled, hairpin, interdigital, and combline bandpass filter techniques from Genesys software simulation

Type of filter	Center frequency (f0) GHz	Fractional bandwidth (FBW)(%)	Insertion loss (S_{21}) In dB	Return loss (S_{11}) In dB	Dimension (mm) L × W
End coupled	10	15.8	0.73	39.523	80.5 × 12
Parallel coupled	10	7.3	1.496	42.837	36 × 27
Hairpin	10	15	0.676	47.986	20.2 × 10
Interdigital	10	15.5	0.795	41.275	21.5 × 10
Combline	10	15.7	0.597	33.431	17.4 × 12

5 Conclusion

This paper presents novel techniques to design microstrip bandpass filters for the X-band with RT/duroid 5880 substrate. The simulated results from Genesys Keysight software show the best outcome with improved fractional bandwidth, good narrow-band response, lower insertion loss, and higher return loss. In this paper, we have introduced simulated results of different types of microstrip bandpass filters with improved steeper roll-off and high selectivity. We observed that fractional bandwidths of the filters in the range from 7.3% to 15.8% with average insertion loss and return loss of 0.8588 dB and 41.010 dB, respectively. Comparative analysis of the existing microstrip bandpass filter with the designed filter shows that combline filter with dimension 17.4 mm × 12 mm is the smallest among other filters with the least insertion loss of 0.597 dB and fractional bandwidth of 15.7%, but it requires lumped

capacitance to reduce the resonator lines. However, hairpin filters have fractional bandwidth of 15% with an insertion loss of 0.676 dB without the requirement of via-holes. So, hairpin can be easily integrated, has relatively simple design, and is of low cost. Such a filter can be used for possible future RADAR-based applications.

Acknowledgements The author acknowledges the Keysight technology for providing student trial versions of Genesys Keysight software and SAMEER for precious support and guidance.

References

1. J. J. Jijesh, Shivashankar, S. K. Anusha, K. M. Guna, Rashmi, "Design and development of bandpass filter for X-band RADAR receiver system," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) (2017), pp. 2065–2070. https://doi.org/10.1109/RTEICT.2017.8256963.
2. D.M. Pozar, *Microwave Engineering* (Wiley, 2012)
3. W. Jia-Sheng, M. J. *Lancaster, Microstrip Filters for RF/Microwave Applications* (Wiley, 2001)
4. I.C. Hunter, *Theory and Design of Microwave Filter*. IET (2001)
5. I. Hunter, L. Billonet, B. Jarry, P. Guillon, Microwave filters-applications and technology. Microwave Theory Tech. IEEE Trans **50**, 794–805 (2002). https://doi.org/10.1109/22.989963
6. A. Grebennikov, *RF and Microwave Transmitter Design* (Wiley, 2011)
7. B. Adli, R. Mardiati, Y. Y. Maulana, "Design of microstrip hairpin bandpass filter for x-band radar navigation," 2018 4th International Conference on Wireless and Telematics (ICWT) (2018), pp. 1–6, DOI: https://doi.org/10.1109/ICWT.2018.8527781
8. C. Ping, Y. T. Li, Z. Y. Chuan, Y. W. Sheng, "Design of X-band receiver on airborne SAR/GMTI multi-model reconnaissance radar," 2011 3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR) (2011), pp. 1–4
9. H. Xu, W. Sheng, "The X-band microstrip filter design," 2017 7th IEEE International Symposium on Microwave, Antenna, Propagation, and EMC Technologies (MAPE) (2017), pp. 351–355. https://doi.org/10.1109/MAPE.2017.8250872
10. F.T. Ladani, S. Jam, R. Safian, A novel X-band bandpass filter using substrate integrated waveguide resonators. IEEE Asia-Pacific Conf. Appl. Electromagn. (APACE) **2010**, 1–5 (2010). https://doi.org/10.1109/APACE.2010.5720093
11. A. O. Lindo et al., Parallel and end coupled microstrip bandpass filters at W-band. 2009 Asia Pacific Microwave Conference (Singapore, 2009), pp. 345–348. https://doi.org/10.1109/APMC.2009.5385376
12. R. Ahmed, S. Emiri, Ş. T. İmeci, Design and analysis of a bandpass hairpin filter. 2018 International Applied Computational Electromagnetics Society Symposium (ACES) (Denver, CO, 2018), pp.1–2. https://doi.org/10.23919/ROPACES.2018.8364310
13. S. M. Kayser Azam, M. I. Ibrahimy, S. M. A. Motakabber, A. K. M. Zakir Hossain, "A Compact Bandpass Filter Using Microstrip Hairpin Resonator for WLAN Applications," 2018 7th International Conference on Computer and Communication Engineering (ICCCE), (2018), pp. 313–316. https://doi.org/10.1109/ICCCE.2018.8539247
14. B. Mohajer-Iravani, M. A. El Sabbagh, Ultra-wideband and compact novel combline filters. 2009 IEEE International Symposium on Electromagnetic Compatibility (Austin, TX, 2009), pp. 176–179. https://doi.org/10.1109/ISEMC.2009.5284644
15. D. R. Basavaraju, H. V. Kumaraswamy, M. Kothari, S. Kamat, In-terdigital bandpass filters for duplexer realization for LTE band 28. 2017 International Conference on Intelligent Computing and Control (I2C2). Coimbatore (2017), pp. 1–4. https://doi.org/10.1109/I2C2.2017.8321919

16. A. Makrariya, P. K. Khare, Design and analysis of a 5-pole inter-digital band-pass filter at 2.4GHz. 2015 International Conference on Communication Networks (ICCN) (Gwalior, 2015), pp. 93–96, https://doi.org/10.1109/ICCN.2015.19
17. M. Yuceer, A reconfigurable microwave combline filter. IEEE Trans. Circuits Syst. II Express Briefs **63**(1), 84–88 (2016). https://doi.org/10.1109/TCSII.2015.2504010

A Novel Approach to Detect Plant Disease Using DenseNet-121 Neural Network

Nilesh Dubey, Esha Bhagat, Swapnil Rana, and Kirtan Pathak

Abstract The disease of crops is a major risk to food security and can incur a makeable loss to the people. But, the latest development in deep learning for solving this problem surpasses all the traditional methods in terms of efficiency, time period for detection and accuracy. In this paper, we came up with a rapid identification of leaf image and classify the image to correct class by using classical deep neural network architecture, DenseNet-121. This deep learning model has the ability to recognize 15 types of different plant disease, three of which are healthy ones, for better accurate results. The algorithm is highly optimized to produce results in less than 5 s after being fed into the system. The model's total testing accuracy for plant disease detection is 99%.

Keywords Plant disease · DenseNet-121 · Neural network · CNN · Agriculture

1 Introduction

India is the world's second largest agricultural producer, producing more than 280 million tons and accounting for more than 15% of India's GDP. Though the use of new technologies and equipment replaced nearly all the traditional methods of farming, but there are plenty of problems that farmers are been faced in his field of work like unpredictable climatic conditions and lack of support and awareness. Due to all of these issues, farmers' usage of chemical pesticides has increased, resulting in chemical build up in our soil, water, air, environment, animals, and even our own bodies.

One of the most important aspects of plant maintenance is disease detection because if disease is not discovered, the plant will not grow as it should. Farmers often do this by sight, but due to the rapid increase in complexity and variance in crop cultivation, it has become difficult for farmers, as well as agricultural professionals

N. Dubey (✉) · E. Bhagat · S. Rana · K. Pathak
Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research, Faculty of Technology Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, India
e-mail: nileshdubey.ce@charusat.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_7

and pathologists, to diagnose the disease, and it is also time-consuming. Tomatoes, potatoes, and peppers are the veggies that everyone eats because no cuisine can be imagined without them. However, these crops are now threatened by a variety of diseases, including early blight, septoria leaf spot in potatoes and tomatoes, and bacterial spot in peppers. Moreover, early detection of these agricultural diseases has the potential to reduce waste and control expenses, as well as improve product quality. Many classical technologies [1] such as biosensors have been proposed to enable simultaneous detection of many DNA or RNA sequences in a single reaction. Similarly, another thermography system catches emitted infrared radiation with thermo-graphic cameras and analyzes color differences. However, these systems have several drawbacks, such as the fact that it takes a long time to collect plant samples and test them physiologically in order to precisely diagnose the illness. This is where deep learning comes into play. It teaches a computer to filter inputs such as images, video, text, and so on through layers in order to learn how to predict and classify information without the need for human interaction. Though specific network topologies are utilized in deep learning, convolutional neural networks (CNNs) are commonly used in image recognition. CNNs are made up of four hidden layers:

(1) Convolutional
(2) Rectified linear unit (ReLU)
(3) Pooling
(4) Fully connected.

The principle and crucial features of images are extracted via convolution and then passed to the ReLU layer, where element-wise operations are conducted. The rectified feature map is then sent via a pooling layer, which decreases the dimensionality of the feature map and extracts prominent features, before the model learns and classifies the picture into the label using a fully connected layer and outputs a label. DenseNet-121 outperforms them all for image classification (leaf) because it is pretrained on the ImageNet dataset, which is the best training dataset for image classification.

2 Literature Review

The technology, which is currently used, includes spectroscopic, image-based, and volatile segregation-based methods for the detection of plant disease for the objective of generating a ground-based sensor system that helps to monitor the plant conditions [2].

Some procedure includes features retrieved from particle swarm optimization (PSO) [3, 4] and determine the affected cotton leaf spot and providing better final accuracy of 95%.

The apple (Malus domestica) disease detection [5] technique is implemented by incorporating k-means clustering and the Bayes classifying algorithm. The breakdown of texture is accomplished easily by using k-means clustering algorithm and

co-occurrence matrix. The author carried out the PCA analysis, and the Bayes classification algorithm does the disease detection. PNN algorithms are used for the classification of plant diseases that use layers like competitive and radial. The leaves images are preprocessed and converted to gray scale before feeding into the system [6]. Another paper [7] addresses specific thresholding methodologies. The threshold techniques such as triangle and simple are used to differentiate the affected area and the leaf segment. The disease is classified in the final step using percentage calculation of the area of the lesion and leaf.

Soft computing techniques [8] are used to present generic image segmentation algorithms. The grayscale leaf image, SVM classifier, and minimal distance criteria are used for image classification. One paper [9] proposes a framework for rapid and precise recognition of plant diseases using artificial neural networks and with other techniques involving image processing. The end system provides better, accurate results of up to 91%. The detection of plant disease by large deep neural network structures, such as AlexNet and GoogleNet [10], is carried out on a dataset with 38 class labels. In the case of AlexNet trained from scratch, the accuracy varies considerably from 85.53% to that of GoogLeNet by transfer learning, which has accuracy 99.34%, obtained from the Plant Village dataset.

The PlantVillage image dataset was used to train the model to classify 14 species of plants and 26 diseases. They carried out an evaluation of two new sets of data of 121 and 119 images, found on the Internet, after getting a testing accuracy of 99.35% on the PlantVillage dataset [11]. They got 31.40% accuracy for the first dataset and 31.69% for the second dataset. Challenges and inadequacies of works that used convolutional neural networks (CNNs) to detect specific crop diseases [12].

CNN architectures such as InceptionV3, VGG16, DenseNet-121, and ResNet50 for image processing are grounded on various characteristics that are present mainly on the leaves infected with anthracnose compared to the non-diseased leaves. The accuracy ranges from 92.4% to 98.7% for identifying brown spots on leaves [13]. One proposed approach implemented the k-means clustering algorithm for grape disease detection. It configured an SVM classifier based on 31 viable particular features to recognize grape powdery and grape downy mildew diseases with 90 and 93.33% as testing accuracies [14].

3 Methodology

3.1 Experimental and Technical Design

The full process of developing the model for plant disease detection using CNN is detailed. The entire operation is divided into various necessary stages in the subsections below, beginning with obtaining photographs from the Internet for classification.

Dataset. Proper datasets are required at all stages of object detection and recognition research, from the training phase to the evaluation of recognition algorithms' performance. All of the photos from the dataset were obtained from the GitHub repository [15]. The images in the dataset are categorized into 21 different classes, and we have chosen 15 of them. To make the model function, we included these 15 classes in our dataset. There are two types of classes in the dataset: 13 for disease classification and 2 for healthy detection. As a result, 20,638 photos have been collected for the model.

Next step was to make the images in the dataset randomize using NumPy and pandas with a SEED of 42 and adding two more attributes "Disease Type" and "Disease ID." After this, every class has given an ID for better classification and ease of programming. By this, every image will have two attributes.

(i) Disease ID.
(ii) Disease Type.

The dataset can be visualized using a histogram (Fig. 1). Here is the diagrammatic representation of whole dataset is shown.

Displaying the images of particular class using Python script (Fig. 2). The script will detect all the images of specific class and disease present in that crop. Here, we have displayed "Tomato Plant," and the disease selected was "Tomato Bacterial Spot."

Image Processing. Image processing is always required for every model in machine learning and deep learning. It depends on the system which one is using; our system can handle above size very smoothly and efficiently. Following will be classes for the disease detection:

Fig. 1 X-axis = Disease ID, Y-axis = Frequency of disease. Bins of the chart represent the number of times a particular disease ID occurs in the dataset

Fig. 2 Plant tomato having tomato bacterial spot disease

(1) Pepper bell bacterial spot.
(2) Pepper bell healthy.
(3) Potato early blight.
(4) Potato late blight.
(5) Tomato bacterial spot.
(6) Potato healthy.
(7) Tomato Early Blight.
(8) Tomato Late Blight.
(9) Tomato Leaf Mold.
(10) Tomato Spetoria Leaf Spot.
(11) Tomato Spider mites two Spotted spider mites.
(12) Tomato Target Spot.
(13) Tomato Tomato YellowLeaf Curl Virus.
(14) Tomato mosaic Virus.
(15) Tomato Healthy.

Fig. 3 Images converted to 64 × 64 resolution using image processing techniques

In Fig 3, the images of resolution are 64 × 64 that are to be fitted in the model. We can see that the images are quite blurry which might affect the training of model, but due to limited computation resource, we decided to make it 64 × 64 (Fig. 3).

4 Experimental Results

4.1 Neural Network Training

Dataset pictures were divided into two sets, training and validation datasets with a ratio of 8:2. Using the image instances, we developed a convolution neural network (CNN) based on DenseNet-121 which accepts a three channels input image of 64 × 64 × 3 resolution and returns a 15-dimensional vector where this vector is generated using soft-max activation function. Also, we had added global max pooling which extracts best possible features of the image. Using the global max pooling costs us less in computation level as it only takes the maximum values from a square matrix. Batch normalization is being used because it solves the problem called "internal covariate shift" and to the speed of training. We had used three kernels of size 1 × 1 to convolute the input layer image. We decided to use this network architecture because its rationalize the pattern of relationship between layers which was present in introduced architectures and pretrained on the ImageNet dataset which is one of the most preferable dataset when it comes to object detection and recognition:

(1) Highway networks
(2) Residual networks
(3) Fractal networks.

We solved the problem by assuring maximal information (and gradient) flow. DenseNets utilizes the network's potential through feature reuse by logically connecting every layer directly with each other. Instead of pulling representational

power from exceptionally deep or wide topologies, we rationally connect every layer directly with each other. It provides benefits such as strong gradient flow, parameter and computational efficiency, maintains low complexity features, and so on. The architecture of DenseNet-121 is depicted in the image below.

Dense block and transition block. Due to heavy number of connection in the DenseNets, the conceive gets less simple compared to other CNN like VGG and ResNets. The below figure demonstrates a straightforward schematics on the detailed architecture of DenseNet-121. It is the simplest DenseNet in terms of complexity among those designed over the ImageNet dataset. In Fig. 4, the measures under each volume represent the size of the width and depth. Figure 5 will tell more about upcoming structure in the architecture.

Fig. 4 Schematic diagram of DenseNet-121 architecture with dense block, convolution block, and transition layer

Fig. 5 Detailed structure of DenseNet-121

In Fig. 5 D1, D2 up to D4 represents dense block, and T1, T2 up to T3 represents transition block. There are bypass connections all over the network.

Moreover, image augmentation was used for better classify the unseen image and to increase the training accuracy. Convolution training was done using two Python libraries, namely Keras and TensorFlow back end, which is a framework, especially designed for deep learning purposes. Pixel values of every input images were normalized by dividing each and every value by 255 so that they lie within [0.0,1.0]. Moreover, the network was initialized with weights derived from ImageNet dataset which is world's largest collection of images. However, with the help of a categorical cross-entropy, loss metric model is compiled, and weights of the layers were optimized using optimizer called the Adam optimization algorithm having a learning rate of 0.02. A collection of 64 images, each having size of 64×64, were provide to the network as single batch per iteration. Considering timing, it was approximately 1 to 2 min for each epoch out of 70 in our experimental condition (Kaggle kernel with GPU-enabled mode). These hyper-parameters are decided after tuning the parameters with range of values and comparisons of accuracy and precision. After a successful training of the CNN model, the feature extraction layers which are from convolute to global max pooling layer were improvised to detect major features from the image in order to diagnose plant disease. The summary of model is shown in Fig. 6.

As the Fig. 6 elicits the information regarding the different layers in the model with respective activation functions, the input layer is of shape $64 \times 64 \times 3$; after that hidden layer is consisting of multiple layers of neurons including convolution layer of shape $64 \times 64 \times 3$ with 12 parameters. The principle layer in the hidden layer is the DenseNet-121 layer which can be say as the backbone of the whole model consists of 7,037,504 number of parameters. Then, a pooling and normalization layer to increase the efficiency and last part is of fully dense layer consists of 300 neurons and 307,500 number of neurons. The output layer is preceded by a batch normalization layer with 1200 parameters, and itself has a shape of 15 neurons and 4515 parameters.

As Fig. 7 suggests that final validation accuracy and loss are 99.00% and 0.02%, respectively. We achieved this with 70 epochs.

Performance metrics and testing. Essentially, the matrix was used to assess the performance of a classifier that had been pretrained with new images from the testing dataset. These results are then used to assess the performance metrics that are commonly employed in AI classification algorithms. Table 1 shows the performance indicators utilized in our study to evaluate the performance of our classifier, along with their descriptions and mathematical formula.

The confusion matrix is a table that displays the abovementioned numbers in such a way that the number of correctly categorized cases, as well as false positives and false negatives, can be easily seen. In our example, the matrix will correctly and incorrectly identify diseases. Figure 8 depicts the matrix created using performance metric approaches.

Some more details about the model insights are given in the following two graphs (Fig. 9).

```
Downloading data from https://github.com/keras-team/keras-applicat
29089792/29084464 [==============================] - 0s 0us/step
Model: "model_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         (None, 64, 64, 3)         0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 64, 64, 3)         12
_________________________________________________________________
densenet121 (Model)          multiple                  7037504
_________________________________________________________________
global_max_pooling2d_1 (Glob (None, 1024)              0
_________________________________________________________________
batch_normalization_1 (Batch (None, 1024)              4096
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_1 (Dense)              (None, 300)               307500
_________________________________________________________________
batch_normalization_2 (Batch (None, 300)               1200
_________________________________________________________________
dropout_2 (Dropout)          (None, 300)               0
_________________________________________________________________
root (Dense)                 (None, 15)                4515
=================================================================
Total params: 7,354,827
Trainable params: 7,268,531
Non-trainable params: 86,296
```

Fig. 6 Model summary with parameter evaluation and segmentation

The left graph shows representation of epoch's vs loss in training as well as testing phase. It can be clearly see that the loss during training was very high and gets decreased constantly with increase in epochs. At the maximum epochs of 70, losses are very near to 0.0. Considering the testing phase, loss was more than loss occurred in training phase, and it also gets decreased by oscillating as number of epochs increase. The decrement of loss is because of the Adam optimizer we have applied; it tries to improve the network better and better while learning it. Similarly, with the right graph, it can be clearly see that the accuracy during training was very low, approx. 0.4 and gets increased constantly with increase in epochs. At the maximum epochs of 70, the accuracy is very near to 0.0. Same optimizer was used in this phase.

5 Conclusion

Plant diseases are major food threats that ought to be overcome before it results in further loss of the whole field. But, often farmers confuse between diseases and that leads to destruction of the whole field. Due to this dilemma, farmers apply small

```
Downloading data from https://github.com/keras-team/keras-applicat
29089792/29084464 [==============================] - 0s 0us/step
Model: "model_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         (None, 64, 64, 3)         0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 64, 64, 3)         12
_________________________________________________________________
densenet121 (Model)          multiple                  7037504
_________________________________________________________________
global_max_pooling2d_1 (Glob (None, 1024)              0
_________________________________________________________________
batch_normalization_1 (Batch (None, 1024)              4096
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_1 (Dense)              (None, 300)               307500
_________________________________________________________________
batch_normalization_2 (Batch (None, 300)               1200
_________________________________________________________________
dropout_2 (Dropout)          (None, 300)               0
_________________________________________________________________
root (Dense)                 (None, 15)                4515
=================================================================
Total params: 7,354,827
Trainable params: 7,268,531
Non-trainable params: 86,296
```

Fig. 7 Final loss and validation accuracy

Table 1 Performance matrix formula

Measure	Formula
Accuracy	(TP + TN)/(TP + TN + FP + FN)
Misclassification rate (1—accuracy)	(FP + FN)/(TP + TN + FP + FN)
Sensitivity (or recall)	TP/(TP + FN)
Specificity	TN/(TN + FP)
Precision (or positive predictive value	TP/(TP + FP)

or excess amounts of pesticides and other chemicals. Here, we employ a convolution neural network (CNN) having multiple layers of ANN called deep learning algorithms to scale back this loss and guide farmers with suggestions.

Fig. 8 Confusion matrix of model

Fig. 9 Model loss (left) and model accuracy (right) produced at each epoch in training as well as testing

References

1. Y. Fang, R.P. Ramasamy, Current and prospective methods for plant disease detection. Biosensors **5**(3), 537–561 (2015)
2. S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, "A review of advanced techniques for detecting plant diseases," Comput. Electron. Agricult. **72**(1), 1–13 (2010); F. Author, S. Author, T. Author, Book Title. 2nd edn. Publisher, Location (1999)
3. P. Revathi, M. Hemalatha, Identification of cotton diseases based on cross information gain deep forward neural network classifier with PSO feature selection. Int. J. Eng. Technol. **5**(6), 4637–4642 (2014)
4. C. Zhou, H.B. Gao, L. Gao, W.G. Zhang, Particles warm optimization (PSO) algorithm. Appl. Res. Comput. **12**, 7–11 (2003)
5. B. Sabah, N. Sharma, "Remote area plant disease detection using image processing". IOSR J. Electron. Commun. Eng. **2**(6), 31–34 (2012), ISSN: 2278–2834
6. S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network". 2007 IEEE International Symposium on Signal Processing and Information Technology (2007)
7. S.B. Patil et al., Leaf disease severity measurement using image processinG. Int. J. Eng. Technol. **3**(5), 297–301 (2011)
8. V. Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques. Inf. Process. Agricult. **4**(1), 41–49 (2017)
9. H. Kulkarni Anand, R.K. Ashwin Patil, Applying image processing technique to detect plant diseases. Int. J. Mod. Eng. Res. **2**(5), 3661–3664 (2012)
10. S. P. Mohanty, D. Hughes, M. Salathe, "*Using Deep Learning for Image-Based Plant Disease Detection*," in arXiv:1604.03169[cs.CV]
11. S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection. Front. Plant Sci. **7**, 1419 (2016)
12. J. Boulent, S. Foucher, J. Théau, P.-L. St-Charles, Convolutional neural networks for the automatic identification of plant diseases. Front. Plant Sci. **10**, 941 (2019)
13. A. Anagnostis, G. Asiminari, E. Papageorgiou, D. Bochtis, A convolutional neural networks based method for anthracnose infected walnut tree leaves identification. Appl. Sci. **10**, 469 (2020)
14. M. Ji, L. Zhang, Q. Wu, "Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks", Inf. Process. Agricult.
15. https://github.com/spMohanty/PlantVillage-Dataset/tree/master/raw/color

Offline Handwritten Signature Forgery Verification Using Deep Learning Methods

Phan Duy Hung, Pham Son Bach, Bui Trong Vinh, Nguyen Huu Tien, and Vu Thu Diep

Abstract Offline signature verification is one of the most challenging tasks in biometric authentication. Despite recent advances in this field using image recognition and deep learning, there are many remaining things to be explored. The most recent technique, which is Siamese convolutional neural network, has been used a lot in this field and has achieved great results. This paper presents an architecture that combines the power of Siamese Triplet CNN and a fully-connected neural network for binary classification to automatically verify genuine and forgery signatures even if the forged signature is highly skilled. On the challenging public dataset for signature verification BHSig260, the proposed model can achieve a low False Acceptance Rate = 13.66, which is slightly better than the reference model. Based on this approach, the one-shot learning should make it possible to determine if the input image is genuine or fraudulent just from one base image. Therefore, our model is expected to be extremely suitable for practical problems, such as banking systems or mobile authentication applications, in which the amount of data for each identity is limited in quantity and variety.

Keywords One-shot learning · Offline signature verification · Siamese convolutional neural network · Triplet loss

P. D. Hung · P. S. Bach · N. H. Tien
FPT University, Hanoi, Vietnam
e-mail: hungpd2@fe.edu.vn

P. S. Bach
e-mail: bachpshe130246@fpt.edu.vn

N. H. Tien
e-mail: tien19mse13047@fsb.edu.vn

B. T. Vinh
Hanoi Procuratorate University, Hanoi, Vietnam
e-mail: vinhbt@tks.edu.vn

V. T. Diep (✉)
Hanoi University of Science and Technology, Hanoi, Vietnam
e-mail: diep.vuthu@hust.edu.vn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_8

1 Introduction

A signature is a handwritten depiction of a person's name, nick name or symbol. The traditional function of signature is to permanently affix to a document, which plays a role as a physical evidence of the author's personal witness and certification of the content [1]. One of the explanations for its widespread use is that it is simple, fast, non-invasive and other people are acquainted with it in everyday life. Signature verification system aims to automatically discriminate if a biometrics sample is indeed of a claimed individual. In other words, it is wont to classify if a signature is genius or fraud. The fraud signatures, which are called forgeries, are commonly classified in three types: random (blind) forgery, simple forgery and skilled forgery. Based on the signature acquisition method, most recent signature verification systems are categorized in two types: online (dynamic) signature verification and offline (static) signature verification [2]. In the online method, the user's signature is acquired by using an acquisition device like a digitizing table. The online signatures are collected as a sequence over time and contain numerous information, such as pen position, pen inclination, pressure. On the other hand, signatures acquired from the offline method are the digital images of the user's signatures signed in the document after the writing process has completed. When the online method can contribute more diverse features for the verified process, it requires significant devices and techniques to acquire data, which makes the system more expensive and cumbersome. On the other hand, the data from offline methods are easier to obtain, and the image digital signatures are more suitable for practical problems. However, the lack of informative features compared to the other method and the quality instability of image data make the verification process more challenging.

The problem of offline signature verification is often modeled as follows: given a group of genuine signatures of users, a model is trained to extract meaningful features from them. Then, the model is employed for verification: a user claims their identity and provides one or some new signatures, which can be employed by the model to classify those signatures as genuine (belong to the claimed individual) or forgery (created by someone else).

In signature verification, one of the most important and familiar challenges for this task is that handwritten signatures have high intra-class variability, which means there are always differences between each handwritten signature from the same person [2]. This issue leads to an aggravation when considering skilled forgeries, especially with the presence of low inter-class variability in the dataset. In short, when there are large differences between each signature sample (high intra-class variability) and the forgeries are very skillful that they are nearly the same with the genuine (low inter-class variability), the possibility of confusing the forgeries with the genuine is higher.

The second challenge when training an automatic signature verification system comes from the presence of particle knowledge during training. In the realistic scenario, we have got only to access genuine signatures for users enrolled within the system during the training phase, in line with [3]. During operations, however,

we would like the system to not only be able to accept the real, but also to reject forgeries. To solve that the forgery signatures for each user should be required for a better classification. But in general, it is not reasonable to require users to provide signatures that forge their own. Even if they can be collected and created from the service provider or from a third party, it would be a challenge to create a good forgeries data which is good enough to be absolutely distinguishable from the genuine signatures.

Thirdly, the amount of training data is always being considered. During the enrollment process in the real application, users are often required to produce only some samples of their signatures. In the meantime, some approaches like the Writer Independent approach need a large enough amount of data to perform well. That's why it is hard to create an efficient system with the very limited data collected from enrolled users in a very real case scenario. Whether or not there's an outsized number of users supplying their signatures to the system during the training phase, the performance of the classifier has to be superb for the new user, for whom supply only few samples of signatures.

Last but not least, like the face recognition system, an effective real-world signature verification system should be able to deal with the one-shot learning problem. Which means the verification application should instantly classify the genuine and fraudulent signatures of a user from just one genuine signature of that individual. However, deep learning algorithms do not work well with a very small dataset, especially if you only have one training example.

In general, signature verification systems have two phases: feature extraction and classification. In the first phase, the feature obtained from the extraction must be meaningful enough to represent the distinctive features of each individual (Writer Depend approaches) or show the clear difference between the genuine and forged signatures (Writer Independent approaches). Then, based on those features, a classifier will be used to make the final prediction.

In 2008, Impedovo et al. [1] provided a summary in automatic signature verification, with specific attention to most outstanding advancements in machine learning. Then, Shah et al. [2] presented a survey about the critical evaluation of 15 techniques applied on offline signature verification systems, which classify each work according to the feature extraction methods, classifiers and overall strengths and limitations of the systems. However, these reviews do not update to capture more recent trends, specifically the usage of deep learning methods, which have demonstrated superior results in multiple benchmarks.

Comprehensive surveys and state-of-the-art reviews of the recent literature are often found within the works of [3, 4] in step with these reviews, most deep learning methods for this task can archive results with high accuracy on an outsized dataset. Beside, among numerous researchers, a growing number of researchers have realized that "learning with small datasets" are a key factor for the success of signature verification in real practical scenarios. Authors like Bouamra et al. [5] attempt to train the model with a tiny low number of signature samples through a one-class support vector machine (OC-SVM) classifier. Hafemann et al. [6] propose an answer supported a meta-learning approach.

In recent years, many interesting techniques do not depend on hand-engineered feature extraction anymore. With the increase of the many deep learning approaches, the power to execute more meaningful feature representation from data (like pixels, just in case of images) by itself can promote faster and efficient learning, which helps researchers to save lots of significant amounts of work. In this field, Khalajzadeh et al. [7] applied CNNs for Persian signature verification, but only considered random forgeries in their tests.

Koch et al. [8] firstly propose an approach about Siamese network and one-shot learning for image recognition. This approach outperformed many available methods by previous authors. Moreover, they have argued that this approach can extend in other domains, especially for image classification.

Schoff et al. [9] came up with the same idea of applying the Siamese network, but in a different way. In this research, they propose a new system called FaceNet, which learns a way to map face images to a compact Euclidean space where distances directly correspond to a measure of face similarity. They achieved high accuracy ($95.12\% \pm 0.39$) in their published dataset, and also introduced new concepts about harmonic embeddings and harmonic triplet loss to deal with the face classification task.

The above approaches [8, 9] both apply Siamese structure including deep CNNs for getting a new representation of image feature, which will be used to classify tasks by comparing them over their distance between each other in new vector space. Those techniques are very flexible and do not need a dataset with millions of samples to make them work.

Dey et al. [10] applied convolutional Siamese network with contrastive loss function for writer independent offline signature verification and their system performs very well on cross dataset for this task. In addition, their approach has surpassed the state-of-the-art result on most of the benchmark signature datasets, which is encouraging this technique for further research. However, when performing evaluation across different datasets, the accuracy of their system gradually decreases on the datasets which have more distinctive features compared to the training dataset.

Chhabra et al. [11] have constructed an interesting method based on the convolutional Siamese net architecture from FaceNet [9] and triplet loss concept to solve the one-shot problem for offline signature verification. Their model has high accuracy and generalizability on the public signature database. However, they do not show the evaluation result in various databases and the performance comparison with other state-of-the-art methods.

Inspired by the work of Chhabra et al. [11], this study improves the network architecture of Deep Triplet Ranking CNN Network by applying more robust transfer learning models and more efficient classification methods. The work focuses mainly on forgery detection in signature verification systems with one-shot learning strategy, which means our model may not perform well on recognition identity from signature but to eliminate forgeries as much as possible. The model is evaluated on the dataset BHSig260, the challenging public dataset for signature verification.

The rest of this paper is organized as follows: In Sect. 2, we are going to give an outline of the general public dataset used, and present the steps of preprocessing.

Section 3 describes the research methodology. Section 4 includes experimental results and comparing them with other approaches. Finally, in Sect. 5, we summarize the result and define the longer term work.

2 Data Preparation

In this work, the signature dataset BHSig260 is used. These data are public, free to access, and contain skilled forgeries. BHSig260 dataset contains the offline signatures of 260 persons, in which 100 persons were signed in Bengali and the remaining were signed in Hindi. For each of the signers, 24 genuine and 30 forged signatures are available. This results in 2400 genuine signatures and 3000 forged signatures in Bengali, and 3840 genuine and 4800 forged signatures in Hindi. Our experiment is only in the Bengali signatures datas from this dataset.

Through analyzing the characteristics of the signature image, we have carried out the following preprocessing steps to improve the efficiency of the recognition process:

- The color factor is not important when comparing the dissimilarity of genuine and fraudulent signatures. That is why we convert all signature's image to grayscale.
- Using a median filter with kernel size 3×3, most salt and pepper noise are easily eliminated.
- After removing noise, using the Otsu threshold to convert the gray images to be the binary image. In addition, invert to the negative image so the background become zero and only the signature's area has pixel value 255.
- Finally, resizing all signature images to a fixed size for feeding into the network 128×128.

3 Research Methodology

The architecture proposed in Fig. 1 is modified from the architecture introduced by Chhabra et al. [11]. We uses 3 sysmetrics deep convolutional neural networks played as the subnetworks in the Siamese architecture. *Instead of building the new deep CNNs and training it from scratch, the Xception model, a pre-trained model in ImageNet is used for three branches*. The transfer learning strategy is used to reduce time for the training process: firstly, the model with the weights trained on ImageNet dataset will be removed the fully-connected layers on top, then all layers are freezed except the exit flow for train on the new prepared dataset. This strategy helps our model have the ability to extract the high feature representation of the image (from the freezed layers) while still being able to learn the important features from the signature image in new dataset. In this case, the Xception was chosen because of its superiority over models VGG-16, ResNet and Inception V3 and has been summarized

Fig. 1 Our Deep triplet ranking CNN architecture with Xception

in [12]. Moreover, it has fewer parameters than Inception V3, which can save more memories on the training and deployment processes.

In this design, our model accepts the input as three signature images categorized as: anchor (an individual's genuine signature), positive (another genuine signature of the same individual) and negative (a forged signature of the same individual). After the preprocessing, those triplet images are fed through the triple subnetworks in which each is constructed by transfer learning Xception top-up with a stack of a Batch Normalization, ReLU activation function and a global max pooling layer.

After being concatenated, our network uses a Lambda layer to compute two distances between embedding vectors: "positive-anchor", "anchor-negative". And the last calculates the triplet loss. Of course triplet loss optimization will continue to update the weights of the sequential part so that the embedding images have adjusted distance to become closer if they are genuine, or become further away from the forgeries.

In the next phase of genuine-forgery classification, we continue to use the idea of using distance vectors for the classifiers of weighted authors in [11] (Fig. 2), but replace the logistic regression classifier with a new classifier stack. The new binary classification includes a fully-connected layer and ReLU activation function, which is top-up with sigmoid function to calculate the probability that a signature is forged or genuine.

Fig. 2 Classification flow illustration

Table 1 Pairs labeling method for train and test

Set(Train/Test)	Pair	Label
Train users	Genuine—Genuine	0
	Genuine—Forgery	1
Test users	Base genuine—Genuine	0
	Base genuine—Forgery	1

4 Research Experiment

4.1 Experimental Setup

First of all, we split 20% of users and take their signature images because the test data. When forming the triplet combinations to train the triplet net, we take an anchor image (genuine signature of a person) and placing it in conjunction with both a positive sample (another genuine signature of the identical person) and a negative sample (a forged signature by some other person of the identical person). By that strategy, from dataset BHSig260 Bengali, we have 654,120 triplet combinations for training. The training pairs for the classifier are formed after step of embedding of all signatures using the train triplet net. We have the vector pairs, label them (see Table 1) and feed them to the classifier.

4.2 Results

The triplet net is trained by using Adam Optimizer with the learning rate equal to 1e-5 and mean absolute error for backpropagation to get the final encoding. The triplet combination is randomly split into a train set and validation set with ratio 2:1, respectively.

In this work, the positive class is the forged signature and the negative class is the genuine signature. We assess or compare the performance of system based on the following values:

- Area Under the Curve (AUC): the measure of the power of a classifier to differentiate between classes.
- False Acceptance Rate (FAR): the percentage of identification instances within which unauthorized persons are incorrectly accepted. FAR = FNR = FN/(FN + TP) = 1—recall. This parameter is that the most important during this study, the smaller the better.
- False Rejection Rate (FRR): the percentage of identification instances within which authorized persons are incorrectly rejected. FRR = FPR = FP / (FP + TN).
- Equal Error Rate (EER): The intersect of lines FAR, FRR.

The Tables 2 and 3 show the performance of the final model for detecting forgery and genuine signatures:

Beside, we have the performance on the test set in case of the pair Base Genuine—Forgeries is no longer depend on identity anymore, which means we construct the pair from the base genuine signature with all forgery signature in the dataset to get more positive label (the forged signature) for testing (Table 4).

The result from Table 4 can be used to compare with some state-of-the-art and the recent studies with Siamese network in the following Table 5:

When the solution is used in real applications, only one genuine signature image is necessary for new user. The image will be preprocessed and feed through the encoder to get the base embedding vector. This vector will be stored in the database and be marked by an id to query in future.

When a new signature image is attempted on the system, if the owner of the new signature claims to be anyone in the system by promoting the id, the system will use that id to track and query the base embedding out for comparison. The system then uses the encoder to get the embedding of the new signature image, and get

Table 2 Performance on validation set

AUC (%)	ERR	FAR	FRR
99.99	0.13	0.00	0.00

Table 3 Performance on test set

AUC (%)	ERR	FAR	FRR
86.16	22.75	18.57	27.97

Table 4 Performance on the test set with random forgeries by other user forged signatures

AUC (%)	ERR	FAR	FRR
99.34	14.18	13.66	27.97

Table 5 State-of-the-art performance on BHSig260 dataset (WD = Writer Dependent, WI = Writer Independent)

Language	Type	Features and algorithm	FRR	FAR
Bengali	WI [10]	SigNet	13.89	13.89
	WI [13]	Dutta et al	14.43	15.78
	WI [14]	Pal et al	33.82	33.82
	–	Our model	27.97	13.66

the vector's difference between the embedding of the base signature and the new signature. Lastly, the classifier will use this vector's difference as an input to predict the probability if the new image is a forged signature or not. The higher the value of this predicted result, the higher the probability that the new signature is a fake one.

5 Conclusions and Perspectives

The paper proposes an effective method for verifying the offline handwritten forgery signatures based on the improvement of a highly applicable study. The article has performed image preprocessing to improve recognition efficiency, improve in processing steps such as using transfer learning strategy and choosing a suitable machine learning architecture model for the problem.

From the result, we can achieve a good FAR result on the BHSig260 dataset which is slightly better than the reference results.

In the future, the study need to find more suitable preprocessing methods for different kinds of datasets. The model also need to be tuned hyper parameters of more powerful triplet ranking loss, and may need to apply triplet online mining techniques to let the learning process be better.

References

1. D. Impedovo, D. Pirlo, Automatic signature verification: the state of the art. IEEE Trans. Syst. Man Cybernetics Part C (Appl. Rev.), **38**(5), 609–635 (2008)
2. A.S. Shah, M.N.A. Khan, A. Shah, An appraisal of off-line signature verification techniques. Int. J. Modern Educ. Comput. Sci. **4**, 67–75 (2015)
3. L.G. Hafemann, R. Sabourin, L.S. Oliveira, Offline handwritten signature verification - Literature review. In: Proceedings of the Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Institute of Electrical and Electronics Engineers (IEEE) (Montreal, QC, Canada, 2017), pp. 1–8
4. M. Diaz, M.A. Ferrer, D. Impedovo, M.I. Malik, G. Pirlo, R.A. Plamondon, Perspective analysis of handwritten signature technology. ACM Comput. Surv. **51**, 1–39 (2019). https://doi.org/10.1145/3274658

5. W. Bouamra, C. Djeddi, B. Nini, M. Diaz, I. Siddiqi, Towards the design of an offline signature verifier based on a small number of genuine samples for training. Expert Syst. Appl. **107**, 182–195 (2018). https://doi.org/10.1016/j.eswa.2018.04.035
6. L.G. Hafemann, R. Sabourin, L.S. Oliveira, Meta-Learning for fast classifier adaptation to new users of signature verification systems. IEEE Trans. Inf. Forensics Secur. **15**, 1735–1745 (2020). https://doi.org/10.1109/tifs.2019.2949425
7. H. Khalajzadeh, M. Mansouri, M. Teshnehlab, Persian signature verification using convolutional neural networks. Int. J. Eng. Res. Technol. **1** (2012)
8. G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for oneshot image recognition. In: ICML, pp. 1–8 (2015)
9. F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015), pp. 815–823
10. S. Dey, A. Dutta, J.I. Toledo, S.K. Ghosh, J. Llados, U. Pal, *SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification*. arXiv:1707.02131v2 (2017)
11. O. Chhabra, S. Chakraborty, Siamese triple ranking convolution network in signature forgery detection. In: Proceedings of the Alliance International Conference on Artificial Intelligence and Machine Learning (AICAAM) (2019)
12. F. Chollet, *Xception: Deep learning with Depthwise Separable Convolutions*. arXiv preprint arXiv:1610.02357 (2016)
13. A. Dutta, U. Pal, J. Llados, Compact correlated features for writer independent signature verification. In: Proceedings of the 23rd International Conference on Pattern Recognition (ICPR) (2016), pp. 3411–3416
14. S. Pal, A. Alaei, U. Pal, M. Blumenstein, Performance of an off-line signature verification method based on texture features on a large indic-script signature dataset. In: Proceedings of the 12th IAPR Workshop on Document Analysis Systems (DAS) (2016), pp. 72–77

Potential Applications of Advanced Control System Strategies in a Process industry—A Review

Pradeep Kumar Juneja, Sandeep Kumar Sunori, Kavita Ajay Joshi, Shweta Arora, Somesh Sharma, Prakash Garia, Sudhanshu Maurya, and Amit Mittal

Abstract Present paper attempts to review the applications of advanced control strategies based on artificial intelligence techniques and its hybrid counterparts applicable in process industry. This chemical process industry may be textile, paper, water purification plant, sugar mill, leather, steel, or any sub-process which may be common in all these industries. It covers an exhaustive literature review.

Keywords Process control · Control system · Controller tuning · Time delayed process

1 Introduction

The process modeling involves the development of dynamic mathematic models with state space forms describing the thermal state in the heated and cooled objects. Dynamic model-based optimal control strategies for both batch and continuous thermal processes in terms of the maximum principle, dynamic programming, heuristic search, etc. are also proposed in this study. The target of the development of optimal control for the heating processes is to provide the optimal heating patterns based on the given criteria and constraints associated with dynamic models and others. A complete hierarchical computer control structure for heating processes is proposed [1]. Nixon discussed recent advances in the use of programmable logic controllers for batch house control systems [2]. Jaggers explored that more accurate sensors make possible tighter control loops [3]. Taniguchi et al., applied advanced technology for cold strip mill installed in Japan that improved gage accuracy. The specifications and features of the mill are set forth, and the gage control system is examined in more detail [4] (Table 1).

P. K. Juneja (✉)
Graphic Era University, Dehradun, India
e-mail: mailjuneja@gmail.com

S. K. Sunori · K. A. Joshi · S. Arora · S. Sharma · P. Garia · S. Maurya · A. Mittal
Graphic Era Hill University, Bhimtal Campus, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_9

Table 1 Literature review

Sr. no	Authors	Year	Technology	Application
1	S. W. Nixon	1988	PLC	Batch house control systems
2	T. Taniguchi et al	1989	Advanced gage control system	Cold strip mill
3	C. Constantinescu et al	1990	DCS	Industrial applications
4	P. P. Aslin	1992	ACS	pH control in sugar mill
5	E. M. Heaven	1993	Parametric techniques	Paper machine
6	G Lightbody et al	1994	System identification techniques	Polymerization reactor
7	G. Rigler, et al	1994	Simulation and control	Multiple stand hot strip mill
8	Zhang Qiping et al	2003	APC	Industrial application
9	H. Ren-Chu et al	2013	MVC controller design	Ammonia synthesis process
10	P. K. Juneja et al	2013	MPC strategy	Lime kiln process
11	P. K. Juneja, et al	2018	Modeling and simulation	Paper machine headbox
12	P. K. Juneja et al	2020	Modeling, control and instrumentation	Review on lime kiln process

Constantinescu et al., presented the hierarchical structure of an integrated control and management system for industrial applications. Its main levels are: distributed process control, process operations support and plant management. This structure has been specialized in the case of a cement plant. The architecture of a gracefully degrading advanced control station is given which supports the process operations in the case of medium size applications, such as cement factories. Result of the global and local repair actions, effect of the spare processing elements and influence of different types of faults have been emphasized. The main advantages of the presented advanced control station are that only the redundancy manager and the application software have been especially designed. This has led to a low cost implementation of the station [5].

Lewis discussed a practical process industry problem of considerable difficulty and defeated conventional approaches [6]. The chemical company Du Pont uses advanced process control to improve operability of plants. This enhances safety, protecting people and the environment, and defines product quality within narrow limits. It reduces costs and increases profits. It is shown that the use of advanced process control enhances safety and increases efficiency [7]. Aslin presented advanced control process for pH control in a sugar industry [8]. Jones, discussed the application of dynamic simulators in industry. The use of training simulators is quite widespread within certain process industry sectors such as refining,

petrochemical and oil and gas. Due to advances in technology and a wider awareness of the benefits these simulators are increasingly used for engineering and operations purposes. In the future the distinction between training and engineering simulators can be expected to be less clear. It is thought that soon a single tool will be available to enable a simulator to be used from conceptual design through to operations support [9]. Gough shown that the use of orthogonal filters allows transfer function identification [10]. Vachtsevanos et al., presented a new technique to optimize the slasher control parameters [11].

Heaven et al., examined some of the traditional parametric identification techniques [12]. Cameron presented the results of field applications in mercury reclaim and sulfur recovery [13]. Model and rule-based controllers can provide adaptive feature within a robust controller design framework [14].

Lightbody et al., proposed to utilize model-based approaches to improve the control performance of an industrial polymerization reactor. This involves the development of a process model using system identification techniques, the simulation of the plant within the Simulink environment to allow for the design and validation of control strategies. From these studies a Smith predictor was implemented to significantly improve the polymer viscosity control. Finally, a hardware platform is developed to facilitate the implementation of sophisticated algorithms such as recursive least squares that could not be accommodated on the present DCS [15]. Rigler et al., attempted the simulation and control of a multiple stand hot strip mill [16]. Arruda et al., proposed integrating different software resources with an application example of an oil industry [17]. Graebe Goodwin et al., described three different central strategies that were implemented and evaluated viz. a dithering controller, a linear cascade controller, and a nonlinear cascade controller [18]. Wilkinson et al., reviewed chronologically the status of fuzzy logic from the start to the present scenario [19]. Zhang Qiping et al., applied APC technique on an industrial process and shown that APC technique is capable of mastering and improving the key process targets [20]. Xiaoming Jin et al., performed advanced process control techniques [21]. Li-hong Dai et al., applied a method based on human machine intelligence [22]. Ren-Chu et al., designed six MVC controllers for the ammonia synthesis process. The final industrial application resulted in good control performance and economic benefit [23]. Juneja et al., applied MPC strategy on a lime kiln process. The lime kiln model is perturbed and the responses achieved are compared for controller designed based on MPC strategy [24]. Satisfactory system performances have been achieved by implementing MPC on real plant [25]. Juneja et al., showed that FOPDT model resembles consistency parameter [26, 27]. MOR and MIMO control system analysis techniques are depicted with the aid of flow chart [28, 29]. Modeling and control features of lime kiln process are attempted [30].

Yunhui Luo et al. [31] gives an improved version for the identification of FOPDT model when the response data is very less. So to approximate the step responses a B-spline series expansion are used which in turn provide more operative interpolation values for modeling computation. Least squares method diminishes the eccentricity of response between identified model and actual process by adjusting the error weight coefficients. Anindo Roy et al., [32] used the solidity framework of Hermite-Biehler

theorem for FOPDT process model. The resultant simulation results illustrates that this frame work can be efficiently used for the synthesis of PID controller of the FOPDT model. The proposed method gives comparably more superior results over traditional PID tuning approaches.

Qiang Bi et al. [33] projected a robust identification technique that has been derived from a step test. This method gives improved identification result then that of the prevailing method under step testing and can be easily applied to PID auto tuning. IMC-based control system displays better control action in comparison to Ziegler Nichol's based control system [34].

2 Conclusion

An attempt has been done to review the applications of advanced control strategies which are applicable in process industry. It covers many control strategies viz. programmable logic control, distributed control system, system identification, multivariable control system, modeling and simulation, model predictive control, etc. And many application areas such as, cold strip mill, sugar mill, paper machine head box, polymerization reactor, multiple stand hot strip mill, ammonia synthesis process, limekiln process.

References

1. Y. Lu, Application of modern control strategies to thermal processes in metal industry. 1987 American Control Conference (1987), pp. 1053–1058
2. S. W. Nixon, Advancements in batching control systems [glass industry]. Conference Record of the 1988 IEEE Industry Applications Society Annual Meeting, vol 2 (1988), pp. 1073–1075
3. H. T. Jaggers, A new measurement and control system for rubber calendaring. IEEE Conference Record of 1988 Fortieth Annual Conference of Electrical Engineering Problems in the Rubber and Plastics Industries (1988), pp. 28–34
4. T. Taniguchi, H. Tanaka, T. Kawabata, E. Yasui, T. Ooi, A new control system for reversing cold strip mill. Conf. Record IEEE Indus. Appl. Soc. Ann. Meet. **2**, 1472–1477 (1989)
5. C. Constantinescu, C. Sandovici, Towards the fault-tolerant advanced control of a cement plant. 1990 Second International Conference on Factory 2001-Integrating Information and Material Flow (1990), pp. 168–172
6. D. G. Lewis, Experiences in the application of advanced control engineering in the process industry. IEE Colloquium on Case Studies in Industrial Control (1990), pp. 3/1–3/3
7. J. P. McCormick, K. J. Kelly, Computer based process control applications. IEE Colloquium on Case Studies in Industrial Control (1990), pp. 4/1–4/4
8. P. P. Aslin, Connoisseur applications in the food industry. IEE Colloquium on Automation and Control in Food Processing (1992), pp. 8/1–8/4
9. D. R. Jones, Current application of simulators in the process industries and future trends. IEE Colloquium on Operator Training Simulators (1992), pp. 3/1–3/4
10. B. Gough, Advanced adaptive control applications (in industry). Conference Record on Pulp and Paper Industry Technical Conference (1992), pp. 122–132

11. G. Vachtsevanos, J. L. Dorrity, A. Kumar, S. S. Kim, Advanced application of statistical and fuzzy control to textile processes. [Proceedings] IEEE 1993 Annual Textile, Fiber and Film Industry Technical Conference (1993), pp. 6/1–6/8
12. E.M. Heaven, T.M. Kean, I.M. Jonsson, M.A. Manness, K.M. Vu, R.N. Vyse, Applications of system identification to paper machine model development and controller design. Proc. IEEE Int. Conf. Control Appl. **1**, 227–233 (1993)
13. M. M. Cameron, Use of advanced controls in environmental remediation. Industry Applications Society 40th Annual Petroleum and Chemical Industry Conference (1993), pp. 177–183
14. P. J. King, K. J. Burnham, D. J. G. James, Combined model-based and rule-based controller for process control. IEE Colloquium on Advances in Control in the Process Industries: An Exercise in Technology Transfer (Digest No. 1994/081) (1994), pp. 4/1–4/5
15. G. Lightbody, G. W. Irwin, A. Taylor, K. Kelly, J. McCormick, Advanced control of a polymerisation reactor. IEE Colloquium on Advances in Control in the Process Industries: An Exercise in Technology Transfer (Digest No. 1994/081) (1994), pp. 7/1–7/3
16. G. Rigler, H. Aberl, W. Staufer, K. Aistleitner, K. H. Weinberger, Improved rolling mill automation by means of advanced control techniques and dynamic simulation. Proceedings of 1994 IEEE Industry Applications Society Annual Meeting, vol 3 (1994), pp. 2030–2037
17. L. Arruda, Amaral, Gomide, An object-oriented environment for control systems in oil industry. Proc. IEEE Int. Conf. Control Appl. **2**, 1353–1358 (1994)
18. Graebe, Goodwin, West, Stepien, An application of advanced control to steel casting. Proc. IEEE Int. Conf. Control Appl. **3**, 1533–1538 (1994)
19. J. Wilkinson, Additional advances in fuzzy logic temperature control. IAS '95. Conference Record of the 1995 IEEE Industry Applications Conference Thirtieth IAS Annual Meeting, vol 3 (1995), pp. 2721–2725
20. Z. Qiping, G. Jinbiao, W. Xiangyu, W. Youhua, An industrial application of APC technique in fluid catalytic cracking process control. SICE 2003 Annual Conference (IEEE Cat. No.03TH8734), vol 1 (2003), pp. 530–534
21. X. Jin, G. Rong, S. Wang, Advanced process control and its application to industrial distillation chain. Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788), vol 4 (2004), pp. 3400–3404
22. D. Li-Hong, C. Xue-Bo, Y. Zheng-Jun, F. Yong-Jun, Method of intelligent object-oriented process control design and application on wire cooling system. 2009 Chinese Control and Decision Conference (2009), pp. 5775–5780
23. H. Ren-Chu, W. Hao, G. Xiao-Jing, Advanced process control technology implementation in ammonia plant. 2013 25th Chinese Control and Decision Conference (CCDC) (2013), pp. 1200–1204
24. P. K. Juneja, A. Ray, Robustness analysis using prediction based control strategy for an industrial process. 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC) (2013), pp. 1–3
25. S. M. Zanoli, C. Pepe, M. Rocchi, G. Astolfi, Application of advanced process control techniques for a cement rotary kiln," 2015 19th International Conference on System Theory, Control and Computing (ICSTCC) (2015), pp. 723–729
26. P. K. Juneja, M. Chaturvedi, S. Suman, K. Antil, Modeling of Stock Consistency in the Approach Flow System of the Headbox," 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU) (2018), pp. 1–4
27. P. K. Juneja, M. Chaturvedi, A. K. Ray, V. Joshi, N. Belwal, Control of Stock Consistency in Head Box Approach Flow System. 2019 International Conference on Innovative Sustainable Computational Technologies (CISCT) (2019), pp. 1–5
28. P. K. Juneja, A. Sharma, A. Sharma, R. R. Mishra, F. S. Gill, A Review on Model Order Reduction Techniques for Reducing Order of Industrial Process Transfer Function Model. 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM) (2020), pp. 346–350
29. P. Juneja, A. Sharma, V. Joshi, H. Pathak, S. K. Sunori, A. Sharma, Delayed complex multivariable constrained process control—a review. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (2020), pp. 569–573

30. P. K. Juneja, S. Sunori, A. Sharma, A. Sharma, V. Joshi, Modeling, control and instrumentation of lime kiln process: a review. 2020 International Conference on Advances in Computing, Communication and Materials (ICACCM) (2020), pp. 399–403
31. Y. Luo, W. Cai, H. Liu, R. Song, Identification of first-order plus dead-time model from less step response data. 9th IEEE Conference on Industrial Electronics and Applications (2014), pp. 1410–1415
32. A. Roy, K. Iqbal, PID controller tuning for the first-order-plus-dead-time process model via Hermite-Biehler theorem. ISA Trans. **44**(3), 363–378 (2005)
33. Q. Bi, W.-J. Cai, E.-L. Lee, Q.-G. Wang, C.-C. Hang, Y. Zhang, Robust identification of first-order plus dead-time model from step response. Control. Eng. Pract. **7**(1), 71–77 (1999)
34. S. K. Sunori, P. K. Juneja, M. Chaturvedi, N. Agarwal, Design and analysis of control systems for heat exchanger system of sugar mill. 8th IEEE International Conference on CICN 2016 (2016)

Analysis of Algorithms for Effective Cryptography for Enhancement of IoT Security

Valerie David, Harini Ragu, Vemu Nikhil, and P. Sasikumar

Abstract The Internet of Things has emerged as one of the most prevalent technologies of the current times, finding its place in a myriad of applications and is widely used for digitization and automation applications such as smart city development, automated monitoring systems, healthcare, energy management and much more. With more devices being connected to the Internet, one of the biggest challenges faced by the Internet of Things surfaces—privacy and security risks. The vulnerability of IoT devices and networks have been brought to light, presenting a threat to the integrity of data. Cryptography has proved itself as a method to secure communication channels and data, as a way to ensure IoT security. In this paper, we aim to compare cryptographic algorithms, namely—AES, DES, RSA and lightweight cryptographic algorithm Fernet, to determine which cryptographic algorithm is the most efficient and secure, and can thereby minimize the risk to data integrity and security in IoT applications.

Keywords IoT · Security · Lightweight · Cyberattack · Cryptography · DES · AES · Fernet · RSA

1 Introduction

The Internet of Things (IoT) is a network that connects physical entities or "things" to the Internet through a wireless network to exchange data without the need for human intervention. The Internet of Things is bringing about a transformation in our physical world by introducing a dynamic system of devices that are connected at an unparalleled scale. Technological advancements are allowing IoT to be used in many fields to enhance and automate processes, be it smart sensors that could be

V. David (✉) · H. Ragu · V. Nikhil · P. Sasikumar
School of Electronics Engineering, Vellore Institute of Technology, Vellore, India
e-mail: valeriedavid2101@gmail.com

P. Sasikumar
e-mail: sasikumar.p@vit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_10

used in home automation or even healthcare applications to help provide remote aid and medical developments such as smart inhalers.

Although the Internet of Things has helped make lives easier by connecting devices and performing tasks that would otherwise require manual labour and human intervention, it is also susceptible to security threats, as there is a deficit of built-in security. IoT devices are often vulnerable to targeting malware and hacking, which can then set off botnet attacks at a large scale and cause a threat to the stability and efficiency of networks and the associated devices. According to reports by Kaspersky, as of early 2019, 276,000 IP addresses launched around 105 million attacks on IoT endpoints. The number of attacks of IoT devices increased by around nine times in comparison to early 2018. Since IoT connects devices through networks, continued attacks could cause a threat to developed smart cities and industries as well. These cyberattacks are known to be frequent and are expected to escalate in the near future. Hence there arises a dire need to ensure that all data exchanged over the Internet of Things can be secure and their integrity remain unharmed. Security has an important role to play in order to prevent unauthorized access and misuse of data. In order to protect the data exchanged over IoT from cyberattacks, cryptography has emerged as a method to encrypt and thereby protect data. Cryptography is the process of making data unrecognizable to users that are unauthorized, hence providing confidentiality to authorized users. Encryption and data integrity are necessary requirements for authentication, secure communication and protection of firmware. Ideally, a preferred cryptographic algorithm is one that combines high performance with low cost. However, in most cases it can be seen that there are often performance-cost trade-offs when it comes to choosing an algorithm.

Lightweight cryptography, a section of classical cryptographic algorithms, is often used for IoT devices and applications. This is because lightweight algorithms consume less memory and resources for computation, and is pertinent for resource-constrained conditions whilst delivering adequate security. In cryptography, encryption is the process of converting plaintext sent by the user into an unintelligible ciphertext whereas decryption performs the reverse operation, where the encrypted ciphertext is translated back into an intelligible plaintext. Cryptographic algorithms are largely split up into types—symmetric and asymmetric key cryptography. Symmetric Key Cryptography—This kind of encryption involves a single secret key that is used to encrypt and decrypt a message. It uses a secret key made up of various characters and it is necessary for both the authorized sender and recipient to be aware of the key in order to access the messages. Asymmetric Key Cryptography—This kind of encryption uses two keys in the process of encryption. A message that has been encrypted making use of a public key can be translated back to plaintext only by using a private key. The symmetric key cryptographic algorithms compared in this paper are AES, Fernet and DES whereas the asymmetric key algorithm is RSA.

2 Literature Review

Abu-Tair, Mamun et al. recognized the security threats associated with IoT-enabled smart homes such as data being hacked or a large amount of data being erroneous, and have highlighted a few lightweight cryptographic algorithms such as TRIVIUM and CLEFIA that could be applied in order to prevent the same [1]. Similarly, Saraiva, Daniel AF, et al. analysed cryptographic algorithms such as AES, Twofish, SPECK128, RC6, and many more in order to find the most efficient algorithm for devices connected to the Internet that are constrained by resources. These algorithms were compared specifically to test execution times, power consumed and throughput for IoT devices using them [2]. Datta, Debajit, et al., propose a system in order to combat unprecedented cyberattacks by secure authenticated communication between devices via sound, as opposed to the conventional QR or pin-based authentication systems. The system involves encryption of a random signal before it is transmitted through sound. Many encryption algorithms were combined in order to determine what the most accurate and efficient algorithm would be [3].

Ismail et al., identified that the Message Queue Telemetry Transport (MQTT) protocol, popularly used in IoT environments and machine to machine interactions owing to its small footprint and efficiency, and consuming less memory, energy and time. However, the protocol is susceptible to exploitation and hence implemented the lightweight Fernet algorithm, based on AES-128-CBC to ensure security of IoT devices and the messages sent through them [4]. Umesh V. Nikam et al. also worked towards finding methods to ensure secure communication in an IoT network through the MQTT protocol through standard techniques of cryptography such as end-to-end payload encryption and digital signature implementation [5]. Effy Raja Naru, et al., have compared various lightweight cryptographic algorithms for resource-constrained IoT devices for secure transmission of data. The comparison of algorithms was based primarily on computation and required storage space in order to protect and ensure the privacy, integrity and security of the data being transmitted [6].

Pradeep Semwal et al., studied various cryptographic algorithms for data security applications in cloud computing. The algorithms were compared based on criteria including computation time for encryption and decryption, memory consumption and Avalanche effect. The algorithms compared in this study were DES, 3DES, RSA, AES, Blowfish, IDEA and CAST-128 [7]. Similarly, Priyadarshini Patil et al., studied the strengths, weaknesses, performance and cost of the same cryptographic algorithms on the basis of the time it takes for the algorithm to perform encryption and decryption operations on the data, the amount of memory consumed, entropy, and the requirement of certain bits to encode data flawlessly. The comparison was implemented using Java cryptography [8]. A survey was performed on cryptographic techniques by Padmavathi et al., on algorithms such as AES, DES and RSA with an added LSB substitution methodology on the basis the time it takes for the algorithm to perform encryption and decryption operations on the data, and buffer size given data of multiple packet sizes. The data transmitted is in the form of images [9]. Along

Table 1 Comparison of algorithms

Factors	RSA	DES	Fernet	AES
Cipher type	Asymmetric	Symmetric	Symmetric	Symmetric
Key length (bits)	>1024	56	256	128/192/256
Block size (bits)	>=521	64	128	128
Speed	Slowest	Slow	Very fast	Fast
Scalability	Not scalable	Scalable	Scalable	Not scalable
Power consumption	High	Low	Low	Low
Security	Least secure	Medium	Excellent	Excellent

similar lines, Prerna Mahajan et al., analysed encryption algorithms for security purposes. The algorithms analysed were RSA, AES and DES. The analysis was on the basis of evaluation parameters of the time it takes for the algorithm to perform encryption and decryption operations on the data packed into multiple sizes for comparison purposes [10]. Gurpreet Singh et al., evaluated encryption algorithms for ensuring the security of information being communicated via systems on the internet. Four algorithms were analysed during this study, namely RSA, AES, DES and 3DES. On comparing these algorithms in real-time, it was found that on the basis of speed, avalanche effect and throughput, AES resulted in being the most efficient algorithm for applications in information security [11].

3 Proposed System and Implementation

Through our work, we aim to compare a few lightweight cryptographic algorithms that can be used in resource-constrained IoT environments in order to determine which algorithm best suits the criteria of efficiency and security. The cryptographic algorithms explored within the scope of this paper are RSA, AES, Fernet and DES. Table 1 shows a comparative study between various parameters for the four cryptographic algorithms.

4 Results

The algorithms are evaluated based on five evaluation parameters namely: encryption time, decryption time, throughput of encryption, key lengths and average entropy per byte. The comparison is performed against multiple packet sizes for each algorithm.

On analysing Table 2 and Figs. 1, 2 and 3, it can be concluded that the Fernet algorithm takes the least amount of time for both encryption and decryption of data of all packet sizes, followed by AES and DES. The RSA algorithm takes the most time for

Table 2 Comparison of parameters

S. No	Algorithm	Packet size (KB)	Encryption time (sec)	Decryption time (Sec)	Throughput (KB/sec)
1	RSA	118	10.0	5.0	11.8
	DES		3.2	1.2	36.88
	Fernet		0.066	0.065	1787.8
	AES		1.7	1.2	69
2	RSA	153	7.3	4.9	20.9
	DES		3.0	1.0	51
	Fernet		0.066	0.066	2318.18
	AES		1.6	1.1	95.63
3	RSA	196	8.5	5.9	23.058
	DES		2.0	1.4	98
	Fernet		0.066	0.063	2969.7
	AES		1.7	1.24	115.29
4	RSA	312	7.8	5.1	40
	DES		3.0	1.6	104
	Fernet		0.076	0.070	4105.26
	AES		1.8	1.3	173.33
5	RSA	868	8.2	5.1	105.85
	DES		4.0	1.8	217
	Fernet		0.077	0.079	11,272.72
	AES		2.0	1.2	434

Fig. 1 Time for encryption of all the algorithms

Fig. 2 Time for decryption of all the algorithms

Fig. 3 Comparison of encryption throughput

both encryption and decryption. Hence, Fernet has the highest encryption throughput for all packet sizes, whereas RSA has the least.

Table 3 and Fig. 4 show that RSA has the largest key size, followed by Fernet, AES and DES. However, RSA is an asymmetric algorithm, which means that its key size of 1024-bit is analogous to an 80-bit key length for a symmetric algorithm, when algorithm strength is compared. This means that amongst the algorithms compared, Fernet is considered the strongest depending on the key size.

Table 3 Comparison of key lengths

Algorithm	RSA	DES	Fernet	AES
Key length (bits)	1024	56	256	128

Fig. 4 Comparison of key lengths

Table 4 Comparison of entropy per byte

Algorithm	RSA	DES	Fernet	AES
Avg. Entropy/byte	3.0958	2.9477	3.47248	3.84024

On analysing Table 4 and Fig. 5, it can be seen that AES has the highest entropy per byte, followed by Fernet, RSA and lastly, DES. This shows that each time the algorithm is run, AES tends to produce a more randomized key, hence making it more difficult to crack, in comparison to the other algorithms compared within the study.

5 Conclusion

Through this paper, we have aimed to find a solution to the security challenges faced by IoT applications. Cryptographic algorithms, including the lightweight Fernet algorithm, have been compared based on evaluation parameters of the time it takes for the algorithm to perform encryption and decryption operations on the data, throughput of encryption, key length and average entropy per byte in order to determine the

Fig. 5 Comparison of average entropy per byte

most secure and efficient algorithm which can alleviate the risk to data security in IoT applications.

Since Fernet is a lightweight algorithm, it emerges as the most efficient algorithm, with the least encryption and decryption time and the highest encryption throughput. AES has the highest average entropy per byte, which means that it is the most secure algorithm amongst the algorithms compared, followed by the Fernet algorithm. As IoT devices require cryptographic security provided by lightweight sources in order to reduce latency as well as the consumption of resources, yet maintain the authenticity of data, which can be determined by the difficulty in cracking the key, it can be concluded that out of the algorithms compared, the Fernet algorithm emerges as a better fit for security applications in IoT.

In the future, more lightweight cryptographic algorithms can be compared for IoT device security and authentication, in order to determine the strongest algorithm that can be used.

References

1. M. Abu-Tair et al., "Towards secure and privacy-preserving IoT enabled smart home: architecture and experimental study." Sensors **20**(21), 6131 (2020)
2. D.A.F. Saraiva et al., Prisec: comparison of symmetric key algorithms for iot devices. Sensors **19**(19), 4312 (2019)
3. D. Datta et al., An efficient sound and data steganography based secure authentication system. CMC-Comput. Mater. Continua **67**, 723–751 (2020)
4. E.L.G. Ismail, A. Chahboun, N. Raissouni. "*Fernet Symmetric Encryption Method to Gather MQTT E2E Secure Communications for IoT Devices*." (2020)

5. U. V. Nikam, H. D. Misalkar, A. W. Burange, "*Securing MQTT protocol in IoT by payload Encryption Technique and Digital Signature*" (2018)
6. E. R. Naru, H. Saini, M. Sharma, "A recent review on lightweight cryptography in IoT." *2017 international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC)* (IEEE, 2017)
7. P. Semwal, M. K. Sharma, "Comparative study of different cryptographic algorithms for data security in cloud computing." *2017 3rd International Conference on Advances in Computing, Communication and Automation (ICACCA)(Fall).* IEEE (2017)
8. P. Patil et al., "A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA and Blowfish." Procedia Comput. Sci. **78**, 617–624 (2016)
9. B. Padmavathi, S. Ranjitha Kumari, "A survey on performance analysis of DES, AES and RSA algorithm along with LSB substitution." IJSR, India (2013)
10. P. Mahajan, A. Sachdeva, "A study of encryption algorithms AES, DES and RSA for security." Global J. Comput. Sci. Technol. (2013)
11. G. Singh, "A study of encryption algorithms (RSA, DES, 3DES and AES) for information security." Int. J. Comput. Appl. **67**(19) (2013)

Relevance of Artificial Intelligence in the Hospitality and Tourism Industry

Smrutirekha, Priti Ranjan Sahoo, and Ravi Shankar Jha

Abstract The authors provide a crisp yet in-depth summary of the relevance of artificial intelligence in the hospitality and tourism industry. Focusing on artificial intelligence, the chapter draws the attention of the reader toward its usage and role in the lives of the customers and service providers in the hospitality sector. It also focuses on the costs involved in the incorporation of artificial intelligence in the hospitality industry. Some of the salient features of artificial intelligence in the hospitality sector have been discussed in the chapter by elucidating the benefits of artificial intelligence like financial benefits, technological benefits, and resource management benefits. This chapter sought to project the various factors affecting the role of artificial intelligence in the industry and how the hospitality and tourism industry adopts the technology with time, demand, and expectations of the customer. Both perspectives have been highlighted, i.e., from the customer's point of view as well as the managerial insights. The overall intention and thrust were to provide the reader with perspectives of the importance of advancing stages of artificial intelligence in the chosen industry along with a few contrasting thoughts as artificial intelligence cannot replace the essence of human touch completely as it is one of a kind of a servicescape involving humans as the most crucial and critical link.

Keywords Artificial intelligence · Hospitality · Tourism · Technology

1 Introduction

Technology is a necessary evil. It has evolved and become more sophisticated with time, but it has also disrupted many industries with its pros and cons to consider. The hospitality and tourism industry is also one sector that used to be related to humans with a human touch and experience. Enhanced guests' experiences were

Smrutirekha (✉) · P. R. Sahoo · R. S. Jha
KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India
e-mail: smrutirekha195@gmail.com

P. R. Sahoo
e-mail: prsahoo@ksom.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_11

due to a high level of human service aspect. The service providers were involved in providing services to human consumers like travelers, passengers, tourists, guests, and event attendees, to name a few. But, with technologies like artificial intelligence, the industry has been revolutionized and has gained popularity in incorporating automated services, leading to a better time, revenue, and human resource management.

Hospitality may be defined as the process involving the guests and service providers where they are provided with a set of experiences, be it for leisure or business. Food, accommodation, and experience are the primary concerns of hospitality services, which can be provided in commercial and non-commercial settings. Customer service, particularly the personal relationship between the host and the visitors, distinguishes hospitality as a service sector centered on the people who use it.

Artificial intelligence (AI) in the travel, tourism, and hospitality industries is gaining momentum as time passes by. Hospitality and tourism are now using AI to revamp their operations, thereby lowering costs, increasing production, and improving the efficiency and dependability of the services they provide [7].

In response to the guests' expectations and as a means of enhancing the customer's experience management, numerous hospitality and tourism organizations have begun to employ service automation, mobile applications, and artificial intelligence for ease and smart work. AI is projected to have a significant impact on all aspects of life and society. Hence, this chapter discusses the usage of artificial intelligence in the hospitality industry and how the management has adopted it. The chapter begins with the evolving AI and gradually highlights the relevance of AI in the chosen sector of study, thereby paving a pathway for further research in this field.

2 Overview of Artificial Intelligence

Artificial intelligence is one such innovation of the human mind that puts science in a dilemma whether to be admired or feared. John McCarthy coined the term artificial intelligence and defined it as the ability of a computer's system to interpret external data, learn from the external data obtained, and finally put them to use to achieve specific goals and targets, thereby adopting itself at its flexibility.

The ability of computers or machines to perform seemingly intelligent actions is referred to as artificial intelligence. Although artificial intelligence has been around since the 1950s, technology has only recently advanced to the point where it can be considered reliable enough to be used for critical commercial activities. When addressing artificial intelligence, it is essential to believe that scientists' programs, computers, and models don't have exact human intelligence; they perform the desired act intelligently when needed as per the instructions fed.

Artificial intelligence (AI) is the ability of a machine to understand and use human language before continuing to function on its own. Reasoning, knowledge, learning,

Fig. 1 Phases of artificial intelligence

communication, perception, planning, and other uses of modern AI are all used in society.

There are three types of artificial intelligence, which are as follows [8] (Fig. 1).

3 Adoption of Artificial Intelligence in the Hospitality Industry and Its Implications

The deployment of artificial intelligence significantly impacts corporate processes across all critical horizontals and verticals of the functional areas of hospitality and tourism organizations. AI adoption refers to how quickly hospitality businesses and consumers adopt AI. If consumers sense a financial benefit from employing AI, they will become more favorable. Consumers may be more willing to embrace AI solutions if the industry passes labor savings to them. Customers will eventually come around and embrace technologies after being exposed to them and become more normalized effective and efficient in hospitality processes. Still, to kick-start the process, it is expected that consumers will be informed about the technologies' benefits.

In recent times, AI has been applied in most business sectors, primarily those involving services like the hospitality industry. Tourism arrivals/demand/expenditure, hotel occupancy, waste generation rates in hotels, and energy demand have all been forecasted using AI (neural networks and machine learning) [12].

In the hospitality and tourism industries, AI offers substantial prospects by integrating with machine learning or big data or IoT to innovate products and services,

improve decision-making, accelerate the rate of innovation, enhance business performance, obtain sustainable advantage, and remove business impediments [11]. These technologies are diverse in type and can be used at many stages of a tourism service encounter, such as booking, vacationing at the destination, and returning home. Personalization and customization of services in a unique and time-saving manner are one of the goals of incorporating artificial intelligence in the hospitality and tourism industry.

Intelligent agents, collaborative robotics (robots), biometrics, facial and gesture recognition, intelligent automation, recommendation systems, intelligent products, personalization, text, speech, image, and video recognition, and extended are all AI-enabled business applications. All these are relevant and applicable to the hospitality industry. Almost all jobs presently performed by humans will be performed by robots, AI, and natural language systems in the final stage, when AI is wholly developed and applied in the hospitality industry, and all systems are integrated and interoperable.

Researchers have employed AI for analytical reasons such as identifying destination features, sentiment analysis of Internet reviews, assessing hotel employee satisfaction, and market segmentation. Because it directs the evolution of a hotel firm, innovation can be characterized as the most significant component of its corporate strategy.

Robots, artificial intelligence, and service automation, among other emerging technology, are causing dramatic changes in the way hotels cater to their customers [15]. AI provides a significant opportunity for hotel companies to improve operations, increase productivity, and maintain a constant level of quality [6].

As per the customer expectations of the services, some of the new facets of measuring service quality have consistently induced the necessity for service automation and mobile technologies, mainly with the rise of readily available lodging and restaurant service providers. To meet these demands, an increasing number of hospitality businesses are investing in technology to improve the efficiency and reliability of their services. In a larger sense, numerous macro-environmental forces influence the adoption of AI technology in the hospitality and tourism industry (Fig. 2).

There are various implications for improving hotel customer experience utilizing AI before, during, and after the guest's stay. By making the hotel visible to potential clients, AI search platforms and VR promote the purchase. Virtual reality can be used to test a service that would otherwise be impossible to do. AI's application in consumer research provides limitless possibilities while still cost-effective [10].

This enables hotel firms to create personalized products and services by providing services that best match the needs and expectations of their clients. AI helps guests save time by consuming services when it is convenient for them [4]. This aids in the improvement of operational and service efficiency and effectiveness.

The human aspect of the hospitality and tourism industry operations has made the sector mostly a conventional labor-intensive one. This has shown a direction to the managers for using AI for innovative work instead of hard work, which leads to improving the lives of both the customer and the employees.

Two examples of product innovations include the expanding popularity of guest room automation and artificial intelligence. The Moon Pod, for example, is a

Fig. 2 Macro-environmental forces affecting artificial intelligence

room tablet designed by CitizenM in London that provides hotel guests complete control over their room's temperature, lighting, television, and even window blinds. When a guest is not there, the Shangri-La in Abu Dhabi has incorporated a guest automation system that transforms the room's status from occupied to vacant and adjusts the temperature while sending vital information to the front desk. In addition to technological developments, demographics tend to be the most important determinant.

Customers will be more inclined to employ automation technologies for specific activities if they believe they are appropriate for automation. Customers' perceptions toward AI are influenced and changed by their actual contacts with AI technologies and their participation in service processes. They may lead to a lack of willingness to use them [17].

Customers and managers will see AI as a tool to boost productivity, service quality, and, ultimately, the company's financial outcomes. They can do a cost–benefit analysis of AI adoption and choose whether to apply AI or not, which is then put into practice through an AI adoption program. On the other hand, employees may have more unfavorable attitudes toward AI since they see AI as a tool for their replacement as employees rather than a technique to improve their performance at work, leading to fear of losing their employment and increased turnover intentions.

The hospitality business's scale, market positioning, and corporate culture will all be important factors to examine. Another critical factor to consider is the technological complexity of AI solutions. The simplicity of use of new technology is one of the essential aspects of adopting new technology [14]. Customers' and service providers' cultural attributes may significantly impact how AI technologies are viewed and used. While some civilizations (like Japan, South Korea, and the United States) appear to be open to new ideas and technologies, other countries like Japan and South Korea

are more conservative and dubious of or outright reject such advances. Some of the AI systems, particularly, those that can cause bodily harm to humans, should have safety qualities as a top priority (customers or employees). While technologies like chatbots may annoy some customers, they are unlikely to cause bodily harm to anyone.

When pricing AI-delivered services as part of the hospitality marketing mix, customers' perceptions of AI-provided services and their willingness to pay for non-human delivered tourism/hospitality services must be taken into account. AI technologies that are practical, effective, and economical to acquire will need to be developed. There is some opposition to AI's continuing growth, not only from the working class, who fear being replaced by robots and AI, but also from respected intellectuals and thinkers who see hazards in a technology that could become self-aware [16].

Innovative technologies, such as interactive social hubs, chatbots, in-room intelligent technologies, and robots are used during a stay to create a unique customer experience by providing guests with various helpful information and entertainment, saving them time, and being available to them 24 hours in the hospitality and tourism industry. Incorporating AI into the service process also eliminates the language barrier, resulting in a more efficient service [13]. For hotel visitors, innovative technology solutions are not only helpful but also amusing. Guests are frequently willing to use a service that involves a robot to observe how it works. Innovative technologies are a source of advertising, sales growth, and cost savings for the hotel industry.

Inhouse entertainment, keyless room access, and the ease with which AI-enabled virtual agents can be used are increasingly seen as the new adoptive technologies of guest satisfaction and innovation. Not only this, but also AI is empowered with recognition of the likes, dislikes, interests, and passion through the observation of behavioral patterns and gestures of the guest in a servicescape.

Current revenue management software, which is mainly used in lodging properties, will eventually spread to the inventory management systems of intermediaries and eventually cover the entire industry [5]. Intelligent systems will estimate, forecast, plan, and control operations across channels and suppliers, maximizing resource efficiency, revenue growth, and customer satisfaction. AIs will be the primary tool for improving operations, redesigning, and reengineering business processes, focusing on the customer while leveraging revenues and costs.

While present systems rely primarily on previously acquired data (Big data) to extract any implications and recommendations to tourists, predictive analytics solutions to anticipate customer behavior and project industry success will see a significant advancement in the future [9]. There are few risks involved in using AI, which directs toward loss of privacy, human bias, and the threat of technology guiding the society leading to a mechanized community rather than an ethical one.

4 Benefits and Costs Associated with Incorporation of AI

The adoption of AI technologies has enormous non-financial benefits for hospitality businesses and indirect financial consequences. AI technologies, above all, improve the quality of an employee's job [1]. AI assists staff in avoiding or reducing errors in the service process, such as incorrectly logging orders at a restaurant or incorrectly processing customer data in a hotel's property management system, lowering customer service expenses. At times, AI technologies can make the service process hilarious and enjoyable, therefore stimulating customer purchases.

Thanks to AI technologies, employees can save time, which frees them from tedious duties and focuses on more revenue-generating activities [3]. Reducing time, energy, and costs associated with human resource management is one of the most critical non-financial benefits of employing AI technologies.

The most obvious financial benefit is labor cost reductions due to AI technologies' benefits—robots, kiosks, or chatbots work around the clock and have incredibly high service capacity, meaning they can serve multiple consumers at once or for a set period [18].

The reduction of time, energy, and costs associated with human resource management is a non-financial benefit of employing AI technologies [2]. AI can assist in resolving some of the issues related to hiring and terminating personnel, particularly seasonal workers.

Through novel appealing and interactive ways of delivering services, connecting with clients, and engaging with them, AI could improve the perceived quality of service.

Figure 3 gives a bird' eye view of the benefits of artificial intelligence in the hospitality industry.

AI has gifted the hospitality industry with some very unique and innovative services like:

Fig. 3 Benefits of AI in the hospitality industry

- Dynamic pricing solutions using the market data, i.e., demand-based pricing of products and services, which depends on the current market demands
- Customized and personalized services as per the customer requirements and demands
- Intelligent and instant assistance concerning travel and tourism
- Guest rating and review analysis for generation of the accurate feedback interpretations
- Demand forecasting in terms of hotel occupancy predictions
- Designing of mobile business solutions to the management of the hospitality sector
- Modernization of the existing services and guest experiences leading to smart work
- Resource recruitment and management.

Costs associated with incorporation of artificial intelligence:

- *Acquisition costs*—these include the cost of buying a robot or a kiosk and buying a chatbot, or paying for AI software development.
- *Installation costs*—a robot or a kiosk must be physically delivered to the location where it will be used and installed.
- *Maintenance costs*—these expenditures include the robot/electricity kiosk's usage, spare components, routine maintenance, and machine repair, among other things.
- *Software update costs*—these costs include installing new and updated software for robots, kiosks, AI packages, chatbots, and other devices.
- *Costs for adapting the premises to facilitate robot's mobility.*
- *Costs for hiring specialist*—to operate and maintain the robots/kiosks/chatbots.
- *Costs for staff training*—who needs to know how to use the new technology efficiently and adequately.
- *Insurance costs*—they include insurance for robots/kiosks as assets and insurance for damages caused by a robot, while it is in use.

5 Conclusion

Advances in AI technology have allowed them to be used in various industries and societies, including manufacturing and smart factories, warehousing and supply chain management, and autonomous cars, to name a few. AI is also employed in the service industries, such as education, journalism, financial trading, and legal services.

Businesses use AI to reduce costs, eliminate waste, improve productivity, economic efficiency, and financial bottom lines, streamline operations, design service experiences, and increase revenues, resulting in significant changes to their business models and the nature of work.

It is undeniable that technological growth will continue unabated. Robots, AI algorithms, chatbots, and kiosks will enhance their technical properties and become more

affordable, allowing them to be deployed more broadly in the hospitality industry. The introduction of AI will inevitably result in the loss of jobs, but it will also result in numerous new jobs that do not currently exist. We are unlikely to feel sorry for dishwashers and cleaners if robots do their jobs or if we utilize chatbots and kiosks to deliver faster, cheaper, and more efficient service than human staff.

While we are still in the early phases of significantly integrating AI into the industry, we can only expect an increase in the adoption of such technology. Human technological skills, a lack of substantial and organized opposition to the technology's adoption into the industry, and the desire to employ human labor more efficiently in industrialized countries are all driving factors in the technology's rise in the industry. While some externalities are expected, numerous issues will arise regarding how they affect the labor market, workplace culture, and international relations.

However, we must remember that technology is a tool, not an end in itself. As a result, the economic evaluation of the costs and advantages of AI adoption must go beyond traditional financial measurements and take a more holistic approach.

Innovative technologies bring value and personalize the stay, but allowing them to take precedence may jeopardize the human ties that visitors seek and enjoy. Finally, AI will never replace the human touch, which is critical for the hospitality industry to stay hospitable.

References

1. E. Brynjolfsson, A. McAfee, Will humans go the way of horses. Foreign Aff. **94**, 8 (2015)
2. P.K. Chathoth, The impact of information technology on hotel operations, service management and transaction costs: a conceptual framework for full-service hotel firms. Int. J. Hosp. Manag. **26**(2), 395–408 (2007)
3. T.H. Davenport, *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work* (MIT Press, 2018)
4. S.J. DeCanio, Robots and humans-complements or substitutes? J. Macroecon. **49**, 280–291 (2016)
5. S. Ivanov, *Hotel Revenue Management: From Theory to Practice* (Zangador, 2014)
6. S. Ivanov, U. Gretzel, K. Berezina, M. Sigala, C. Webster, Progress on robotics in hospitality and tourism: a review of the literature. J. Hospitality Tourism Technol. (2019)
7. S.H. Ivanov, Tourism beyond humans–robots, pets and Teddy Bears, in *Paper to Be Presented at the International Scientific Conference "Tourism and Innovations"* (2018)
8. S.H. Ivanov, C. Webster, K. Berezina, Adoption of robots and service automation by tourism and hospitality companies. Rev. Turismo Desenvolvimento **27**(28), 1501–1517 (2017)
9. S. Ivanov, C. Webster, Economic fundamentals of the use of robots, artificial intelligence, and service automation in travel, tourism, and hospitality, in *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and Hospitality* (Emerald Publishing limited, 2019)
10. S. Ivanov, V. Zhechev, Hotel revenue management–a critical literature review. Tourism: An Int. Interdisc. J. **60**(2), 175–197 (2012)
11. R.S. Jha, P.R. Sahoo, Internet of Things (IOT)–enabler for connecting world, in *ICT for Competitive Strategies: Proceedings of 4th International Conference on Information and Communication Technology for Competitive Strategies (ICTCS 2019), December 13–14, 2019, Udaipur, India*, 1 (CRC Press, 2020)
12. J. Kim, S. Wei, H. Ruys, Segmenting the market of West Australian senior tourists using an artificial neural network. Tour. Manage. **24**(1), 25–34 (2003)

13. C.-H. Ko, Exploring how hotel guests choose self-service technologies over service staff. Int. J. Organ. Innov. **9**(3), 16–27 (2017)
14. V. Liljander, F. Gillberg, J. Gummerus, A. Van Riel, Technology readiness and the evaluation and adoption of self-service technologies. J. Retail. Consum. Serv. **13**(3), 177–191 (2006)
15. C. Liu, K. Hung, Understanding self-service technology in hotels in China: technology affordances and constraints, in *Information and Communication Technologies in Tourism 2019*, 225–236 (Springer, 2019)
16. A. Parasuraman, C.L. Colby, An updated and streamlined technology readiness index: TRI 2.0. J. Serv. Res. **18**(1), 59–74 (2015)
17. D.C. Ukpabi, H. Karjaluoto, Consumers' acceptance of information and communications technology in tourism: a review. Telematics Inform. **34**(5), 618–644 (2017)
18. C. Waxer, Get ready for the BOT revolution. Comput. World (2016)

Deep Learning-Based Smart Surveillance System

G. Sreenivasulu, N. Thulasi Chitra, S. Viswanadha Raju, and Venu Madhav Kuthadi

Abstract The role of CCTV cameras has been overgrown in this generation. CCTV cameras are installed all over the places for surveillance and security. Many surveillance systems still require human supervision. Recent advances in computer vision are, thus, seen as an important trend in video surveillance that could lead to dramatic efficiency gains. Various public places like shopping malls, supermarkets, ATMs, banks, and other places, where CCTV cameras are available, are the places we should concentrate on. Security can be characterized in various terms in various settings like robbery distinguishing proof, brutality recognition, odds of a blast, and so on. In jam-packed public places, the term security covers practically a wide range of strange occasions. So, it is important and challenging to build a model which detects these abnormal activities and generates some kind of alert. We used a combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) which involve the concept of deep neural networks. It extracts the spatial–temporal features of the images and calculates the Euclidean distance between the original and reconstructed batch of images. We converted the training videos into images to train the model and calculated the loss between the images to identify the abnormality. To validate the proposed algorithm, 4 datasets as HOLLYWOOD, UCF101, HMDB51, and WEIZMANN are used for action recognition. The proposed technique performs better than the existing one. We made use of Jupiter notebook and Python frameworks.

Keywords Smart surveillance · Deep learning · CNN · LSTM and UCF101

G. Sreenivasulu (✉)
Computer Science and Engineering, ACE Engineering College, Hyderabad, India
e-mail: aceaimldept@gmail.com

N. Thulasi Chitra
Department of CSE, MLR Institute of Technology, Secunderabad, Telangana, India

S. Viswanadha Raju
CSE, JNTUHCEJ, Jagityal, Karimnagar, Telangana, India

V. M. Kuthadi
Department of CS&IS, School of Sciences, BIUST, Palapye, Botswana

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_12

1 Introduction

In the recent past, there is a sizable growth in multimedia content material, and this is growing everyday. Surveillance way regulation or maintaining an eye fixed on a few get-togethers or unique events. So, it is far taken into consideration that video investigation is the satisfactory alternative for tracking and detecting. "Manual surveillance" is just too stressful and inefficient. By the use of this observation system, we will easily stumble on what is occurring at a selected area, and remotely we will screen many locations within side the meantime [1]. Also, there has been fantastic development with inside the video evaluating strategies and methods.

Many scholars have endeavored to expand correct clever observation structures that can apprehend any person thru extraordinary methods. Correctness and competence are the principle issues as 100% efficiency is not always reached [2]. There are a whole lot of "video surveillance systems" and the point of interest of every of them is to yield its area with inside the arcade. Video observation entails investigation that includes a listing of steps like "video preprocessing, object detection, movement detection, recognition, and classification" of actions [3].

Deep neural learning methods are greater appropriate for conduct and studying such datasets [4]. These methods can carry out an evaluation of the picture and video datasets that are to be had openly. These skilled fashions of deep mastering can acquire an precision of greater than percent of 95 in a few cases (Fig. 1).

"Support vector machine (SVM)" is the rendition of administered dominating. SVMs utilize carefully an immense wide assortment of capacities for dominating denied of the utilization of extra computational force. These are effective to represent nonlinear highlights and fit for utilize green strategy for dominating.

There are several foremost levels of our method gadget, e.g., attainment of image, pre-processing of image, selection feature, train our data, and test our data classification [5]. Due to brand new capabilities and picture dispensation used. This paper is comprising of six sections. Section 1 is enabling introduction, Sect. 2 is talking about literature survey, Sect. 3 enables system methods, Sect. 4 is focusing on working model, Sect. 5 is results, and Sect. 6 is conclusion and future scope.

Fig. 1 Illustration of deep neural networks

2 Literature Survey

A major work has previously been carried out on smart investigation structures; many researchers have tried to increase right clever surveillance structures that can apprehend any human hobby via specific methodologies. The below are the specific strategies recycled.

A. **Recognition on Human Action (RHA) Methods**

Majority of scholars labored with inside the area of RHA. In the place of computer vision (CV), members are the use of a spread and dynamic surroundings for the overall presentation assessment of visual dataset. Prof. Ionut [2] future a scheme for indoctrination of functions and their withdrawal to get original dispensation of body charge for motion popularity structures.

A technique is future to get the movement material inside the taken video. The descriptor is proposed. The proposed approach has progressed with good fit.

B. **Features for shape primarily-based totally methods**

Azher [1] added a singular method to apprehend social movements. A new function description is added specifically. ALMD with the aid of using thinking about movement and presence. This method is an allowance of the local ternary pattern (LTP) that is used for stationary textual content evaluation. Spark machine learning library and random forest approach are hired to apprehend human movements. Dataset of six hundred movies such as six human motion training is used for trying out functions. UCF sports motion and UCF-50 records units also are castoff for end outcome evaluation.

The studies explain the methods of temporal units of nearby functions. Aggregation of extracted functions is carried out with VLAD algorithm. Supervised method is used to deliver efficiency in consequences. MSR-Action three-D, UT-Kinect Action 3D, and the Florence 3D.

The consequences indirect that efficiency in out of doors images changed into smaller than the managed surroundings. Zhong et al. added a singular method to apprehend cross-view movements completed with the aid of using the use of digital paths. Virtual view kernel is castoff for locating similarities among multi-dimensional model functions.

C. **Point features grade points**

Gao delivered a method for the gratitude of a couple of fused physical actions. The preliminary stage and wage functions are removed after which the multimedia bag of words version is organized. The graph version is likewise carried outs eliminates the intersecting arguments of the hobby from the data.

Yang supplied a method for RHA in multimedia datasets. For the removal of forefront movement, background motion is remunerated via way of means of the usage of a movement version. The wished foreground patch is over segmented in spatiotemporal patches the usage of methods are more efficient.

Zhang supplied a coding version which can examine greater correct and pictures then different present models. The proposed methodology brings strong consequences in opposition to obstruction and multi-dimensional views. For type of functions, classifiers of supervised learning methods are used.

3 System Approach

In our strategy, first, we put into impact the preprocessing technique that is the combination of various procedures. In the essential advance, we select the channel, practice the formal hat channel, changing the profundity esteems, and difference extending through method of methods for limit esteems executed to embellish the nature of the picture. After the pre-preparing, a weight-basically based absolutely the division approach is applied for discovery to figure body qualification the utilization of combined propose and to supplant the set of experiences through method of methods for the utilization of loads and also to distinguish the closer view regions, and correspondingly, a crossover trademark extraction procedure is utilized for notoriety of human activity. The removed capacities are melded essentially dependent on sequential fundamentally-based absolutely combination, and in some time, the intertwined trademark is applied for arrangement.

A. **Proposed Method**

Our proposed method will analyzes the video feed in real-time and identify any abnormal activities like violence or theft. We are using combination of convolution and LSTM networks (ConvLSTM). We will have to extend deep neural networks to three dimensions for learning spatiotemporal features of the video feed. The abnormal events are identified by computing the **reconstruction loss using Euclidian distance between original and reconstructed batch. With minimal** manpower, abnormal activities are detected, and an alarm will be activated.

B. **CNN and LSTM Architecture**

The CNN and LSTM engineering include utilizing convolutional neural network (CNN) layers for highlight extraction on input information joined with LSTMs to help succession forecast. A CNN LSTM can be characterized by adding CNN layers toward the front followed by LSTM layers with a dense layer on the yield. It is useful to consider this design characterizing two sub-models: the CNN model for highlight extraction and the LSTM model for deciphering the highlights across time steps (Fig. 2).

Fig. 2 CNN and LSTM architecture

C. Code Implementation

This section is divided into seven sections. Each section is considered to be equal importance and having its own importance. Each section is discussed in detailed below.

C.1: Importing Packages

A segment can encompass executable declarations in addition to characteristic definitions. These statements are meant to initialize the module. They are performed handiest the first time, and the module call is encountered in an import statement.

Each module has its very own non-public image desk that is used as the worldwide image desk through all features described with inside the module. In this manner, the author of a module can utilize overall factors with inside the module without irritating around inadvertent conflicts with a client's overall factors. On the contrary hand, on the off chance that you understand what you are doing, you could contact a module's overall factors with the indistinguishable documentation used to counsel its highlights, modname, itemname.

Modules can import various modules. It is ordinary yet at this point not, at this point needed to region all import proclamations toward the beginning of a module (or content, so far as that is concerned). The imported module names are situated with inside the transferring module's worldwide image work area.

C.2: Path Creation

Training videos are placed in train directory, creating a new directory named frames for storing converted image frames

```
❖ store_image=[]
❖ train_path='./train'
❖ fps=5
❖ train_videos=os.listdir(train_path)
❖ train_images_path=train_path+'./frames'
❖ from playsound import playsound
❖ from gtts import gTTS
❖ os.makedirs(train_images_path)
❖ def store_inarray(image_path):
❖ image=load_img(image_path)
❖ image=img_to_array(image)
❖ store_image.append(gray)
❖ text=gTTS("Abnormal event detected!")
❖ text.save("alert.mp3") .
```

Fig. 3 Conversion of video to image (for every 0.5 s is the frequency of frame)

C.3: Conversion of Videos to Images

Capturing an image for every 0.5 s, creating path for every image (Fig. 3).

```
✓ for i in range(1,17):
✓ vidcap =
  cv2.VideoCapture('./train/0{}.avi'.format(i))
✓ def getFrame(sec):
✓ vidcap.set(cv2.CAP_PROP_POS_MSEC,sec*1000)
✓ hasFrames,image = vidcap.read()
✓ if hasFrames:
✓ cv2.imwrite("./train/frames/image"+str(count)+".
  jpg", image)
✓ return hasFrames
✓ sec = 0
```

C.4: Saving Array of Images into .npy file

Image frames are converted into an array. After resizing and clipping, it is stored in training .npy file in a separate file.

Allow us to comprehend the interaction of convolution utilizing a basic model. Consider that we have a picture of size 3 × 3 and a channel of size 2 × 2 (Fig. 4; Table 1).

Fig. 4 CNN working model

Table 1 Convolution matrix

1	1	2
11	1	23
2	2	2

1	7
0	1

C.5: Loading the Model

```
training_data=training_data[:::frames]
training_data=training_data.reshape(-
1,227,227,10)
training_data=np.expand_dims(training_data
, axis=4)
target_data=training_data.copy()
epochs=3
batch_size=1
callback_save
=ModelCheckpoint("saved_model.h5",monito
r="mean_squared_error",
save_best_only=True)
stae_model.fit (training_data, target_data,
batch_size=batch_size, epochs=epochs,
callbacks = [callback_save,
callback_early_stopping])
stae_model.save("saved_model.h5")
def mean_squared_loss(x1,x2):
difference=x1-x2
a, b, c, d, e=difference.shape
n_samples=a*b*c*d*e
sq_difference=difference**2
Sum=sq_difference.sum()
distance=np.sqrt(Sum)
mean_distance=distance/n_samples
return mean_distance
```

C.6: Running the model with static video

```
✓ model=load_model("saved_model.h5")
✓ cap =
  cv2.VideoCapture("test/testing_video.mp4")
✓ print(cap.isOpened())
✓ while cap.isOpened():
✓ x=0
✓ imagedump = []
✓ ret, frame = cap. read ()
```

See Fig. 5.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv3d (Conv3D)              (None, 55, 55, 10, 128)   16616
_________________________________________________________________
conv3d_1 (Conv3D)            (None, 26, 26, 10, 64)    204864
_________________________________________________________________
conv_lst_m2d (ConvLSTM2D)    (None, 26, 26, 10, 64)    295168
_________________________________________________________________
conv_lst_m2d_1 (ConvLSTM2D)  (None, 26, 26, 10, 32)    110720
_________________________________________________________________
conv_lst_m2d_2 (ConvLSTM2D)  (None, 26, 26, 10, 64)    221440
_________________________________________________________________
conv3d_transpose (Conv3DTran (None, 55, 55, 10, 128)   204928
_________________________________________________________________
conv3d_transpose_1 (Conv3DTr (None, 227, 227, 10, 1)   16480
=================================================================
Total params: 1,068,215
Trainable params: 1,068,215
Non-trainable params: 0
```

Fig. 5 Model for "sequential" to add convolution layers

```
for i in range(10):
ret, frame = cap.read()
try:
image = imutils.resize(frame, width=700,
height=600)
except AttributeError:
x=1
break
gray=(gray-gray.mean())/gray.std()
```

```
imagedump.resize(227,227,10)
image=np.expand_dims(imagedump,
axis=0)*np.ones(image.shape)
(imagedump, axis=4)
output=model.predict(imagedump)
loss=mean_squared_loss(imagedump,
output)
if frame.any() ==None:
print("none")
```

✓ cv2.putText(image, "AbnormalEvent", (100,80), cv2.FONT_HERSHEY_SIMPLEX,2, (0,0,255),4)

✓ cv2.imshow("video", image) cap.release() cv2.destroyAllWindows()

✓ stae_model.summary()

C.7: Running the Model with Dynamic Video as an Input

LSTM needs one measurement, yet in the event that you get MobileNetV2 without "top layers" (that is the classification yields, we need not bother with them) so you will have a "convolution" layer. We need to add one layer to have the one measurement shape viable with GRU. Also, MobileNet is not at first intended to identify "activities," however, just to make picture acknowledgment. In this way, on the off chance that you change the whole layers as not "teachable," you will not have precision upgrades or "misfortune" improvement. For instance, I make the last 9 layers teachable.

```
➢ screen_size = (1366, 768)
  fourcc = cv2.VideoWriter_fourcc(*"XVID")
  video = cv2.VideoWriter("captured_video.avi", fourcc, 20.0, (screen_size))
➢ def min():
            M=3Minimize = win32gui.GetForegroundWindow()

    while M >0:
            time.sleep(1)
            M -=1
    time.sleep(1)
    os.startfile("C:/Users/mpsre/test/test_video.mp4")
  time.sleep(2)
```

```
win32gui.GetForegroundWindow()

    while M >0:
            time.sleep(1)
            M -=1
    time.sleep(1)

    os.startfile("C:/Users/mpsre/test/test_vide
o.mp4")
  time.sleep(2)

    win32gui.ShowWindow(Minimize,
win32con.SW_MINIMIZE)

➢ for i in range(20):
      img = pyautogui.screenshot()
      s=str(c)
      img.save("./images/image"+s+".jpg")
      c+=1
      fram = np.array(img)
      fram = cv2.cvtColor(fram,
cv2.COLOR_BGR2RGB) video.write(fram)

➢ model=load_model("saved_model.h5")
  cap = cv2.VideoCapture("captured_video.avi")
  print(cap.isOpened())
  sum()
  while cap.isOpened():
      img = pyautogui.screenshot()
```

```
      for i in range(10)
img = pyautogui.screenshot()
      s=str(c)
      img.save("./images/image"+s+".jpg")
      c+=1
      fram = np.array(img)
fram = cv2.cvtColor(fram,
cv2.COLOR_BGR2RGB) video.write(fram)
ret,frame=cap.read()
      try:
      image =
imutils.resize(frame,width=700,height=600)
            except AttributeError:
            x=1
break
imagedump=np.array(imagedump)
imagedump.resize(227,227,10)
imagedump=np.expand_dims(imagedump,axis
=0)
imagedump=np.expand_dims(imagedump,axis
=4)
output=model.predict(imagedump)
loss=mean_squared_loss(imagedump,output)

            if frame.any()==None:
            print("none")
            if cv2.waitKey(10) & 0xFF==ord('q'):
break
            if loss>0.00065:
```

In the event that you need to utilize a non-prepared model (setting the loads boundary to none), so you will change shape. However, you should leave every one of the layers to be "teachable." In the event that you as of now have an opened scratch pad where you prepared the past model, if it is not too much trouble, slaughter the bit first.

4 Results and Screenshots

See Figs. 6, 7, 8, 9, and 10.

Fig. 6 Normal video

5 Conclusion and Future Scope

For detecting abnormal activities, we used combination of convolutional neural networks and LSTM (ConvLSTM2D) which is a concept of deep neural networks.

As we can see that the accuracy is 0.7327 which means 73.27% of total abnormal activities will be detected correctly. So, majority of the abnormal activities will be detected and appropriate action can be taken immediately. As the dataset has limited training videos, the accuracy is low. If we add more training videos to the model, the accuracy will increase, and the model will be able to differentiate between normal and abnormal events more accurately.

We have tested the model by giving input videos in both static and dynamic (live) ways. This project overcomes the drawbacks of the existing system which involves constant monitoring and decreases man power. So, any abnormal activities like theft, violence, and others at public places will be detected and produce a display and voice messages as alert.

Fig. 7 Common moment video

Fig. 8 Abnormality during theft (ATM)

Fig. 9 Abnormality detected

Fig. 10 Abnormal event detected (during street fight)

5.1 Future Scope

Our method is built to detect abnormal activities. Our model is currently detecting abnormal activity when a static video and dynamic video are passed. Whenever a dynamic video is passed, there is a time lapse of 5–10 s, and the text cannot be displayed as alert as it is live recording. So we can generate alert only through voice message. In future, we would like to work to eliminate the time lapse and try to increase the accuracy which can detect almost majority of the abnormal situations.

Rerferences

1. B. Janakiramaiah, G. Kalyani, A. Jayalakshmi, Automatic alert generation in a surveillance systems for smart city environment using deep learning algorithm. Evol. Intel. (2020)
2. R. Slama, H. Wannous, M. Daoudi, A. Srivastava, Accurate 3D action recognition using learning on the Grassmann manifold, Pattern Recognit. **48**, 556–567 (2015)
3. G. Sreenu, M.A. Saleem Durai, Intelligent video surveillance: a review through deep learning techniques for crowd analysis. J Big Data **6**, 48 (2019)
4. M. Baba, V. Gui, C. Cernazanu, D. Pescaru, A sensor network approach for violence detection in smart cities using deep learning. Sensors (2019)
5. G. Sreenivasulu, S. Viswanadha Raju, et al., A Review of clustering techniques, in *International Conference on Data Engineering and Communication Technology (ICDECT)*, Springer, March-2016.J ISSN: 2250-3439)

Pronunciation-Based Language Processing Tool for Tai Ahom

Dhrubajyoti Baruah, Joydip Sarmah, Kamal Gayan, and Satya Ranjan Phukon

Abstract Words with similar pronunciation may not be of same in its written form. Such similarities can be measured by means of computer algorithms. For English language, an algorithm is available, which may be utilized as a database function "Soundex". This algorithm returns codes on the base of pronunciation of words. Same code signifies similar pronunciation, and dissimilar code indicates different pronunciation. This paper focuses on development of Soundex for Tai Ahom language. In some cases, a single word carries several different meanings. For such matters, exact meaning may be guessed on the basis of other words in the sentence. For this purpose, we have focused on developing another algorithm backed by sentiment analysis.

Keywords Text mining · Soundex · Natural language processing · Sentiment analysis · Tai Ahom

1 Introduction

Soundex is an algorithm used to index words of English language as pronounced. RC Russell and MK Odell had developed the English Soundex algorithm and was patented in the year 1918. Soundex is a phonetic algorithm that returns a specific code against words that sounds similar. The code is a combination of the starting alphabet of the word along with three digits fetched from a mapping table. Words having similar pronunciation get same code from Soundex. Initially, it was used in the US. census from 1890 to 1920. Soundex was used by US census to index people and also for other official purposes. Soundex system helps us to avoid problems like misspellings or alternate spellings of any word. Let us take two naming words "Rima" and "Reema". Though these words are written in different forms, the Soundex code will be the same. The words "Rima" and "Reema" both will produce the same Soundex code "R500". The Soundex system is also useful for searching the ancestor's data. It is common that family names are misspelled in official records. Also it is

D. Baruah (✉) · J. Sarmah · K. Gayan · S. R. Phukon
Department of Computer Application, Jorhat Engineering College, Jorhat, Assam, India
e-mail: dhrubaghy@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_13

quite obvious that written form of family names could change with time. National Archives and Records Administration (NARA) of USA applies Soundex system to search citizen data for different administration purposes. Soundex is also used to check the spelling of words in large databases. For example, in a dictionary, we can use Soundex to find out words on the basis of the pronunciation. Soundex algorithm for English language has some limitations, like the algorithm depends on the first letter of the word and, hence, cannot detect similarity in pronunciation of words with different starting letters. For example, "college" and "kollege", although similar in pronunciation, Soundex yields different codes that incorrectly indicates dissimilarity. We are focusing on design and development of Tai Ahom language Soundex. The other focus of this paper is based on sentiment analysis. In Tai Ahom language, a single word may carry several different meanings, which meaning fits best in a sentence depends on the other present words of the sentence. A word is likely to carry the positively sensed meaning within a sentence with optimistic, positive flavour. The same word may carry the negative meaning (if any) within a pessimistic, negative flavoured sentence.

2 Related Work

A lot of projects are going on Soundex for different languages. "The Soundex Phonetic Algorithm Revisited for SMS Text Representation" [1] work was carried out by David Pinto and his team at BUAP, University of Mexico. Gunma University, Japan, and RMIT University, Australia, are developing a set of phonetic matching functions for the Japanese language [2]. University of Science and Technology, Algeria, researchers have designed Soundex for Arabic language [3]. India, having several different languages, is also focusing deeply on phonetic research. Odia language Soundex is designed by Indian Institute of Information Technology, Bhubaneswar, Orissa [4]. One project work was carried out for the Assamese language by Jorhat Engineering College, Assam [5]. This project is a web dictionary, where similar pronunciation words are displayed from database according to the Soundex code. Soundex algorithm improvement for Indian language on the basis of phonetic matching [6] work was carried out by Rima Shah and Dheeraj Kumar Singh of Gujarat Technological University. For Sindhi language, a phonetic algorithm is developed, to be implemented for Sindhi Spell Checker [7].

3 Tai Ahom Language

The Tai Ahom language is an obsolete, extinct language which was spoken in the North East part of India, by the Ahom people [8]. The Tai Ahom are descendants of Tai-speaking people, who came to North East India in 1228AD [9], led by King Siukapha, and they settled on the Brahmaputra Valley. The Ahoms ruled in Assam

from 1228 to 1826AD. The Ahom people came from Yunnan province of South East Asia [9]. Ahom was used as the prime language of the Ahom kingdom but was gradually replaced by Assamese language. Although Ahom language had lost its use as first language, it continued to be used in marriage and other religious ceremonies, and recently many efforts are made to re-establish the Ahom language [10]. Therefore, although this language is usually regarded as a dead language, it survives in the form of large number of documents written in the heritage Ahom script and exists as a ritual language in Ahom religious ceremonies. The old Ahom manuscripts have somehow helped to understand this dead language and also all the other ways in which it survives today. The derivation of Ahom script was probably from Brahmi origin [10]. The Tai Ahom language is tonal, but tones were not written; so it is not known properly as how tones were arranged and exact number of tones used in the language [11]. The Ahom language now has a place on Unicode [12].

4 Our Approach

Following list is prepared according to their pronunciation. The letters having same pronunciation are grouped together and assigned with a mapping code. The list is prepared based on pronunciation similarity, instead of human articulation base:

Now lets us take “ 𑜈” and “ 𑜉”. Though they are not same letter, both are pronounced as “SA”. Therefore, we group them and assign to mapping code “SA”. Below is a graph showing the intensity while pronouncing “ 𑜈” and “ 𑜉” (Fig. 1).

Now let us take “𑜈𑜨𑜫” and “𑜉𑜨𑜫” Ahom words as example which are pronounced as “sang” and “shang”. Now we will generate Soundex code for these two Ahom words according to the mapping codes mentioned in Table 1.

Mapping code for “𑜈𑜨𑜫”:

𑜈 -> SA
𑜨 -> NA
𑜫 -> O

Therefore, the Soundex code will be SANAO.

Likewise mapping code for “𑜉𑜨𑜫”:

𑜉 -> SA
𑜨 -> NA
𑜫 -> O

Therefore, the Soundex code will be SANAO.

Fig. 1 Graph generated for "𑜀𑜫" and "𑜀"

Table 1 Soundex mapping code of Tai Ahom language

Ahom letter	Code	Ahom letter	Code
𑜀, 𑜁	SA	𑜬	PHA
𑜂, 𑜃	JA	𑜭	BA
𑜄, 𑜅, 𑜆	NA	𑜮	MA
[illegible]	RA	𑜯	THA
𑜇	LA	𑜰	HA
𑜈, 𑜉	AA	𑜱	DA
[illegible]	O	𑜲	DHA
[illegible]	U	𑜳	GA
[illegible]	I	𑜴	GHA
𑜊	KA	𑜵	BHA
𑜋	KHA	𑜶	E
𑜌	TA	[illegible]	AW
𑜍	PA	[illegible]	AI
		[illegible]	AM

Now we observe that the words with similar pronunciation will have same Soundex code. On the basis of the grouping, an algorithm is developed to generate Soundex codes. Algorithm to generate Soundex for Ahom words is given below:

1. Take a variable name as word;
2. Take another variable name as soundex_code to store the generated code.
3. For i = 0;i < word.length;i++

If word[i] == 𑜎 or 𑜏 Then soundex _code = soundex _code + SA Else if word[i] == 𑜐 or 𑜑 Then soundex _code = soundex _code + JA Else if word[i] == 𑜒 or 𑜓 or 𑜔 Then soundex _code = soundex _code + NA Else if word[i] == 𑜕 or or ◌𑜵 or ◌𑜶 Then soundex _code = soundex _code + RA Else if word[i] == 𑜖 Then soundex _code = soundex _code + LA Else if word[i] == 𑜗 or 𑜘 or 𑜙 Then soundex _code = soundex _code + AA Else if word[i] == ◌𑜠 or ◌𑜡	Then soundex _code = soundex _code + PHA Else if word[i] == 𑜚 Then soundex _code = soundex _code + BA Else if word[i] == 𑜛 Then soundex _code = soundex _code + MA Else if word[i] == 𑜜 Then soundex _code = soundex _code + THA Else if word[i] == 𑜝 Then soundex _code = soundex _code + HA Else if word[i] == 𑜞 Then soundex _code = soundex _code + DA Else if word[i] == 𑜟 Then soundex _code = soundex _code + DHA Else if word[i] == 𑜠

Then soundex _code = soundex _code + O Else if word[i] == 𑜍 or 𑜎	Then soundex _code = soundex _code + GA Else if word[i] == 𑜌
Then soundex _code = soundex _code + U Else if word[i] == 𑜏 or 𑜐	Then soundex _code = soundex _code + GHA Else if word[i] == 𑜍
Then soundex _code = soundex _code + I Else if word[i] == 𑜒	Then soundex _code = soundex _code + BHA Else if word[i] == 𑜑
Then soundex _code = soundex _code + KA Else if word[i] == 𑜓	Then soundex _code = soundex _code + E Else if word[i] == 𑜕
Then soundex _code = soundex _code + KHA Else if word[i] == 𑜔	Then soundex _code = soundex _code + AW Else if word[i] == 𑜖
Then soundex _code = soundex _code + TA Else if word[i] == 𑜗	Then soundex _code = soundex _code + AI Else if word[i] == 𑜘
Then soundex _code = soundex _code + PA	Then soundex _code = soundex _code + AM

4. Soundex code generated and stored in variable Soundex_code.

Algorithm 1: Tai Ahom Soundex
Experimented Results for algorithm 1:

See Table 2.

This way, the solution path has been framed for getting correct word among different probable alternative words with variant spelling. In another scenario, a single word may have several meanings. Soundex algorithm is of no use for such cases. In this case, we can consider "sentiment analysis and subjectivity" [13] concept

Table 2 Tai Ahom Soundex output

Word	Soundex Code
𑜒𑜍𑜎, 𑜒𑜍𑜎, 𑜒𑜍𑜎	NAUNAO, NAUNAO, NAUNAO
𑜌𑜎, 𑜌𑜎	JANAO, JANAO
𑜑𑜏𑜐, 𑜑𑜏𑜐	SAUPAO, SAUPAO
𑜑𑜍𑜎, 𑜑𑜍𑜎	SANAO, SANAO

to get the proper meaning of word. Sentiment analysis, also termed as opinion mining, is the computational research of sentiments, opinions and emotions expressed in text. Using this concept, we can mine or extract emotions and opinions in a given sentence.

In our approach, categorization of words, according to positive and negative sense, is done. The word having positive sense is assigned a key value "positive", and the word having negative sense is assigned a key value "negative". On the basis of the key value, computation for meaning fetch of targeted word is done. A complete sentence is needed to work for detection of meaning of targeted word within it. The targeted word carries more than one meaning, and to fetch the actual meaning, the sentence should have at least one negative or positive word in it. If majority of the non-targeted words in a sentence is positive, then the positively sensed meaning of the targeted word is returned by an algorithm. Likewise, if the counter for negative word holds a bigger value, it indicates the negatively sensed meaning of the targeted word in the sentence.

As an illustration, some targeted words are as follows:

1. 𑜈𑜫 (Ka , meaning : to dance, suffering)
2. 𑜉𑜫 (Ra, meaning : epidemic, rain)
3. 𑜊𑜫 (Ja, meaning: medicine, abandoned)
4. 𑜇𑜫 (Ma, meaning: to come, to not eat)

Considered sentences are as follows:

1. 𑜇𑜨 𑜈𑜫 𑜄𑜧𑜊𑜩	5. 𑜇𑜨 𑜈𑜨 𑜊𑜫𑜊𑜧𑜇𑜨 𑜧
2. 𑜇𑜨 𑜄𑜧𑜈𑜫 𑜉𑜫𑜈𑜫 𑜨	6. 𑜇𑜨 𑜈𑜨 𑜊𑜫𑜊𑜧 𑜉𑜫𑜊𑜧 𑜊𑜩
3. 𑜈𑜨 𑜊𑜧𑜉𑜫𑜊𑜧	7. 𑜇𑜨 𑜈𑜫𑜊𑜧 𑜇𑜫
4. 𑜉𑜫𑜊𑜧𑜉𑜫𑜊𑜧𑜄𑜧𑜄𑜧𑜊𑜩	8. 𑜇𑜫 𑜩 𑜉𑜧𑜇𑜨 𑜇𑜫𑜈𑜨 𑜊𑜩

The first sentence listed above is as follows:

𑜇𑜨 𑜈𑜫 𑜄𑜧𑜊𑜩 (She dances well)

Now we will try to find the meaning of the word " 𑜈𑜫 " which has both negative and positive sense, from the above sentence. Leaving our targeted Ahom word " 𑜈𑜫 " aside, we categorize the other words as follows:

𑜉𑜫 -> Neutral word (Neither positive nor negative)

𑜆𑜧 -> Positive word

𑜇𑜩 -> Neutral word

Now we can see that, among the three words, there is a positive word " 𑜆𑜧". Therefore, count of positive is more than negative, and we will take " 𑜊𑜨" as a positive word and will return the positive meaning of it. So, " 𑜊𑜨" will mean "to dance" (positive) here instead of "suffering" (negative).

The second Ahom sentence is as follows:

𑜉𑜫 𑜆𑜀𑜊𑜨 𑜁𑜧𑜊𑜨𑜊 (She got hurt in quarrel).

The Ahom word " 𑜊𑜨" has both negative and positive senses. Keeping the targeted Ahom word " 𑜊𑜨" aside, we categorize the other words as follows:

𑜉𑜫 -> Neutral word

𑜆𑜀 -> Neutral word

𑜁𑜧 -> Neutral word

𑜊𑜨 -> Negative word

𑜊𑜫 -> Neutral word

Now we can see that among the five words, there is a negative word " 𑜊𑜨". Therefore, count of negative is more than positive, and we will take " 𑜊𑜨" as a negative word and will return the negative meaning of it. So, " 𑜊𑜨" here will mean "suffering" or "hurt" instead of "dance". On the basis of this concept, an algorithm is developed to return the probable meaning of a word within a sentence. The algorithm is as follows:

1. Take a string variable containing the sentence as ‘sentence_str’.
2. Take an integer variable containing the position of the targeted word in the sentence as ‘pos’.
3. Take a variable name as ‘Counter’ to count the occurrence of positive and negative word and initialize it to zero
4. Now split the sentence into an array
5. For(i=0;i<sentence_str.length;i++)
6. If sentence_str[i] != sentence_str[pos]
   If sentence_str[i] is positive
   Counter++
   Else If sentence_str[i] is negative
   Counter - -
7. If(counter > 0)
   Then return positive meaning of sentence_str[pos]
   Else If(counter < 0)
   Then return negative meaning of sentence_str[pos]
   Else
   Return “No result found”

Algorithm 2: Meaning Mining

The algorithm designed was tested, and satisfactory results are found. Some of the test results are given in Table 3.

Table 3 Experimented results of algorithm 2 (Meaning Mining)

Sentence	Targeted Word	Meaning
𑜒𑜫𑜀𑜉𑜀𑜉𑜁	𑜉𑜁	Epidemic
𑜀𑜧𑜀𑜉𑜁𑜀𑜇𑜇𑜀	𑜉𑜁	Rain
𑜋𑜫𑜒𑜫𑜁𑜀𑜋𑜫𑜌	𑜁𑜁	Medicine
𑜋𑜫𑜁𑜉𑜁𑜧𑜁𑜒𑜀	𑜁𑜁	Abandoned
𑜋𑜫𑜁𑜀𑜋𑜁	𑜋𑜁	To come
𑜋𑜁 𑜋𑜫𑜀𑜋𑜫𑜋𑜁𑜒𑜫𑜋	𑜋𑜁	To not eat

5 Conclusıon and Future Scope

Algorithm of Soundex for Tai Ahom language can be implemented using any programming language. This algorithm works flawlessly and still is open for any needed modification to improve its performance. The programme developed using this algorithm can be used in many fields, for example, in linguistic and computational task ranging from word suggestion to plagiarism [14]. In future, this algorithm can be used in further development of technology for Tai Ahom language which is an extinct language struggling for revival. This algorithm will be a boon to research and development of Tai Ahom language. The other algorithm based on sentiment analysis for meaning mining may be integrated as future scope for heritage Tai Ahom language research in web computing platform [15].

Acknowledgements This work is carried out under the project "Web Technology Development For Heritage Tai Ahom Manuscripts" which is funded by Dept. Of Science and Technology (DST), Govt. of India, under Science and Heritage Research Initiative (SHRI) Scheme (sanction no DST/TDT/SHRI-11/2018).

References

1. D Pinto, D Vilariño, Y Alemán, H Gómez N Loya, and HJ Salazar, "The Soundex Phonetic Algorithm Revisited for SMS Text Representation"
2. Michiko Yasukawa, J. Shane Culpeppery, Falk Scholery, "Phonetic Matching in Japanese", SIGIR 2012 Workshop on Open Source Information Retrieval. August 16, 2012, Portland, Oregon, USA
3. ND Ousidhoum, A Bensalah, and N Bensaou, "A New Classical Arabic Soundex Algorithm", Proc. of Int. Conf. on Advances in Communication and Information Technology 2012, ACEEE
4. Rakesh Chandra Balabantaray, Bibhuprasad Sahoo, Sanjaya Kumar Lenka, Deepak Kumar Sahoo, Monalisa Swain, IIIT Bhubaneswar, "An Automatic Approximate Matching Technique Based on Phonetic Encoding for Odia Query", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012
5. D Baruah, Jorhat Engineering College, AK Mahanta, Gauhati University, "Design and Development of Soundex for Assamese Language"
6. R Shah, Department of Computer Science And Engineering, Parul Institute of Engineering and Technology, Gujarat Technological University,Gujarat, Vadodara, DKSingh, Department of Information and Technology, Parul Institute of Engineering and Technology, Gujarat Technological University,Gujarat, Vadodara, "Improvement of Soundex algorithm for Indian language based on phonetic matching
7. Z Bhatti, A Waqas, IA Ismaili, DN Hakro, WJ Soomro, "Phonetic based Sou x & ShapeEx algorithm for Sindhi Spell Checker System"
8. https://en.wikipedia.org/wiki/Ahom_language
9. Stephen Morey, La Trobe Univeristy, "Ahom and Tangsa: Case studies of language maintenance and loss in North East India"
10. https://omniglot.com/writing/ahom.htm
11. https://scriptsource.org/cms/scripts/page.php?item_id=script_detail&key=Ahom
12. http://unicode.org/L2/L2012/12309-ahom-rev.pdf
13. Bing Liu, University of Illinois at Chicago, "Sentiment Analysis and Subjectivity"

14. D Baruah, AK Mahanta, "A New Similarity Measure with Length Factor for Plagiarism Detection", International Journal of Computer Applications (0975 – 8887) Volume 72– No.14, May 2013
15. www.ahomweb.in

Improving Ecommerce Performance by Dynamically Predicting the Purchased Items Using FUP Incremental Algorithm

K. Kalaiselvi, K. Deepa Thilak, S. Saranya, T. Rajeshkumar, M. Malathi, M. Vijay Anand, and K. Kumaresan

Abstract Data mining or knowledge discovery is the way toward examining data according to substitute perspectives and summarizing it into accommodating information. This information can be then used to fabricate a pay, decreases costs, or both. Programming made with web mining as its key subject ought to permit clients to isolate information from a wide extent of assessments or centers, demand it, and sum up the affiliations perceived. Taking everything into account, information mining is the way toward discovering affiliations or models among many fields in colossal social instructive assortments. This paper effectively tracks down the rehashed bought things by clients. This proposed algorithm is having a higher running time than the existing FUP incremental algorithm. This algorithm efficiently finds the frequent items, and dynamically the items can be added. The entire history of the frequent item database was added and put into separate clusters. At last, we compare and choose the best-purchased items of the customer and also predict the past purchased items in the history. Based on the output, we can easily find the current status of the customer purchase.

K. Kalaiselvi (✉) · K. Deepa Thilak
Department of Networking and Communications, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India
e-mail: kalaisek2@srmist.edu.in

S. Saranya
Department of ECE, SRM Easwari Engineering College, Chennai, India

T. Rajeshkumar
Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India

M. Malathi
Department of ECE, Vivekananda College of Engineering for Women (Autonomous), Namakkal, Tamil Nadu, India

M. Vijay Anand
Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India

K. Kumaresan
Department of CSE, K.S.R. College of Engineering, Tiruchengode, Tamil Nadu, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_14

Keywords Data mining · FUP · Intra-cluster · Cost

1 Introduction

This undertaking is an expansion of one of the eminent sub-classes of Data Mining:—"market basket analysis (MBA)," which is an appearance framework giving data into the client buying plans [1, 2]. A market compartment is made out of the thing sets which are bought in a solitary outing to the store. MBA from an overall perspective desires to discover the relationship between the things bought in this compartment [3–5]. As a showing mechanical get-together, it is utilized to mine out the ceaseless thing sets in an epic number of exchanges. Hence, it is likewise called "steady item set mining."

2 Proposed System

The proposed paper intends to achieve an improved foreseeing calculation to view the incessant things liable to be bought by the client. This calculation has preferable running time over FUP gradual calculation. It assists with finding continuous things in a powerfully added exchange. Disintegrate the exchange history data set into intentional example isolated bunches. It maps the current client to the most appropriate bunch and performs sequencing of past acquisition of the clients. It additionally predicts the buy succession of the current client and separates the regular thing from the exchanges [6].

This paper intends to achieve an anticipating calculation to discover the incessant things liable to be bought by the client. This calculation has preferred running time over FUP steady calculation. It assists with finding regular things in a progressively added exchange. Disintegrate the exchange history data set into intentional example isolated groups. It maps the current client to the most appropriate bunch and performs sequencing of past acquisition of the clients. It additionally predicts the buy arrangement of the current client, extricate the successive thing from the exchanges [7, 8].

3 Implementation

3.1 Allocation Phase

Allocation phase the total cost can be minimized for every transaction constant variable T assigned to newly created cluster. A similar will be written in the information base. The choice of whether to remember the exchange for one of the current

bunches or to make another one is made by working out the expense of grouping. The expense comprises of intra-bunch disparity and between group closeness which is determined as follows. The decision of whether to recall the trade for one of the current gatherings or to make another is made by determining the cost of batching. The cost contains intra-bunch dissimilarity and between pack likenesses still up in the air as follows. Dispersion stage the outright cost can be restricted for each trade predictable variable T given out as of late made gathering. a comparable will be written in the database allocation phase the total cost can be minimized for every transaction constant variable T assigned to newly created cluster. The same will be written in the database.

Dissemination stage the supreme cost can be restricted for each trade reliable variable T given out as of late made gathering. A comparative will be written in the database Allotment stage the complete expense can be limited for each exchange steady factor T appointed recently made group. A similar will be written in the information base

3.1.1 (a) Intra-Cluster Dissimilarity

$$\text{Intra(U)} = |\text{Ukj} = 1\ \text{Sm(Cj, E)}|$$

S stands for little items. E—Maximum ceiling 5 Cj—jth cluster. To be known as a little thing, the most extreme roof is the greatest number of exchanges that can contain it. Intra-group uniqueness is characterized as the mix of one of a kind minuscule articles found in all bunches. The most extreme amount of exchanges that a roof can contain is called the greatest roof. It is a small thing, but it is a big deal. The association of distinct small elements existing throughout the bunches is the intra-group variation in this way.

3.1.2 (B) Inter-Cluster Dissimilarity

$$|\text{Ukj} = 1|\text{Inter(U)} = \text{kj} = 1|\text{La(Cj, S)}| - \text{Inter(U)}$$
$$- \text{Inter cluster dissimilarity La(Cj, S)}|\text{Large objects(la)}$$

S—Minimum support Cj—jth cluster.

The minimum support indicates the number of transactions in which an item must appear in order to be classified as a major item. The entire cost is computed using the formula below.

The least help indicates the minimum number of exchanges that a thing should be provided to ensure that it is a significant object. The equation that follows determines the total cost.

$$\text{COST} = \text{w} * \text{Intra(U)} + \text{Internal(U)}$$

where w stands for weight.

Intra(U)—Dissimilarity within a cluster. Inter(U)—Cluster similarity between clusters

Another exchange is first placed in every one of the current bunches, and the expense is determined for each group. Then, at that point, another bunch is made to oblige the exchange and the expense is determined. The exchange is then at last appointed to the bunch with the most reduced expense esteem as follows.

- For this reason (each new non-bunched exchange)
- Assign the transaction to the group c
- Calculate the cost for (each group, c)
- Compare the cost to the most cost-effective cost
- Assign existing expense to best cost if (new expense is superior)
- Assign the present bunch to the most suitable group
- Calculate the cost of a new bunch for the current exchange
- If the new expense is better than the best expense so far, assign the current expense to the best cost
- Assign the present bunch to the most suitable group
- End.

One more trade is first positioned in all of the current bundles and the still up in the air for each gathering. By then one more bundle is made to oblige the trade what's more, the not really settled. The trade is then finally given out to the bundle with the most minimal cost an impetus as follows. For this reason (each new non-bundled trade)

- In order to (each bundle, c)
- Assign the bundle c to the trade.
- Estimate the cost
- Compare the price to the best price
- In the event that (new cost is better)
- Assign the best cost to the existing cost
- Assign the present bundle to the most appropriate gathering
- Add another bundle to the current trade
- Estimate the cost
- Compare the price to the best price you've found thus far
- In the event that (new cost is better)
- Assign the best cost to the existing cost
- Assign the present bundle to the most appropriate gathering
- End.

3.2 Refinement Phase

In the refinement stage, the little enormous proportion (SL proportion) of the multitude of exchanges is determined as follows.

$$\text{SLR} = (\text{no. of little things})/(\text{no. of huge things})$$

The SL proportion of every exchange accordingly determined is then contrasted and the SLR limit. Assuming the SLR of the exchange surpasses the edge, the exchanges are moved to the overabundance pool. An endeavor is then made to oblige these exchanges are an alternate bunch, if the SLR of these exchanges in the new group does not surpass the edge. If not these exchanges are considered anomalies and are dispensed with from thought.

The interaction is clarified as follows.

- Calculate S-L proportion of each exchange.
- Move every one of the exchanges whose S-L proportion surpasses the edge to the overabundance pool.
- Shuffle the exchanges in the overabundance pool to various groups with the end goal that the S-L proportion esteem stays beneath the limit.
- Delete the leftover exchanges from the abundance pool. An illustrative illustration of the interaction is as per the following.

In the refinement stage, the little immense extent (SL extent) of all the not really set in stone as follows.

$$\text{SLR} = (\text{no. of easily overlooked details})/(\text{no. of colossal things})$$

The SL extent of each not really set in stone is then differentiated and the SLR edge. In case the SLR of the trade outperforms the breaking point, the trades are moved to the excess pool. An undertaking is then made to oblige these trades is a substitute pack, if the SLR of these trades in the new gathering does not outperform the edge. If not these trades are viewed as exemptions and are cleared out from thought.

The cycle is explained as follows.

- Calculate S-L extent of each trade.
- Move all of the trades whose S-L extent outperforms the edge to the excess pool.
- Shuffle the trades in the excess pool to different bundles with the ultimate objective that the S-L extent regard stays under the edge.
- Delete the extra trades from the bounty pool.

Initial Clustering (Allocation Phase) (Tables 1, 2, 3, and 4; Fig. 1):

The bunching system is consequently finished, consolidating both the portion and refinement stages. The gathering collaboration is thusly all out, intertwining both the task and refinement stages.

Table 1 Clustered transaction

Cluster 1		Cluster 2		Cluster 3	
TID	Item Set	TID	Item Set	TID	Item Set
110	B,D	210	B,I	310	B,I
120	A,B,D	220	A,B,I	320	A,B,I
130	B,C,D	230	B,E,I	330	B,E,I
140	D,F,H	240	B,C,E,I	340	B,C,E,I
150	B,G,I	250	C,I	350	C,I

Table 2 Excess transaction and elimination

Minimum Support = 60% Maximum Ceiling = 30%			
Cluster	Large	Middle	Small
1	B,D	A,C	
2	B,I	C,E	A
3	D,H		F,G
	Intra-cluster dissimilarity = 7 Inter-cluster similarity = 2 Total cost = 9		

Table 3 Excess pool

TID	Item Set
140	D,F,H
150	B,G,I
330	B,C,D,F

Table 4 Final cluster

Minimum support = 60% Maximum Ceiling = 30%			
Cluster	Large	Middle	Small
C1	B,D	C	A,F
C2	B,I	C,E	A,G
C3	D,H	F	G
	Intra(U1) = 3 Inter(U1) = 2 Cost(U1) = 5		

3.3 Incremental Association Rule Mining

Info: Transaction history information base

Yield: Successive thing sets and guaranteed incessant thing sets.

The exchange history information base contains the past exchanges made by the clients. The subtleties incorporate client id, the arrangement of things purchased

Fig. 1 Refinement

alongside the exchange id. This stage has two sub stages, viz. original information base discovery updating successive and promising incessant thing sets. Data: Transaction history database yield: perpetual thing sets and ensured customary thing sets. The trade history database contains the past trades made by the customers. The nuances consolidate customer id, the plan of things bought close by the trade id. This stage has two sub-stages, viz. unique informational collection discovery refreshing nonstop and promising unremitting thing sets.

3.4 Original Database Discovery

A unique database may make it possible to embed new exchanges. Existing affiliation controls may be invalidated, and new affiliation rules may be enacted as a result. Maintaining affiliation rules for a big database is a serious challenge. As a result, this paper provides a new computation to deal with such a refreshing situation. The new approach is based on the assumption that the measurements of new exchanges eventually diverge from those of unique exchanges. As the presumption suggests, measurements of old exchanges obtained through previous mining can be used to approximate measurements of new exchanges. As a result, the support count of object sets obtained from previous mining may differ from the help count. After embedding new exchanges into a unique data set that comprises prior exchanges, a set of item sets is created. When new exchanges are embedded into the first data set, the new calculation uses the greatest help count of 1-itemsets acquired from previous mining to assess inconsistent thing sets of a unique information base that will be equipped

to become regular thing sets when new exchanges are embedded in the first data set. Support count for rare thing sets that will be equipped for continuous thing sets, such as min pL, is displayed in Eq. 1: With the highest help count and the largest size of new exchanges that permit embed into a unique data set, support count for rare thing sets that will be equipped for continuous thing sets, such as min pL, is displayed in Eq. 1:

$$\text{min_sup}_{\text{DB}} - \left(\frac{\text{max supp}}{\text{total size}}\text{Xinc_size}\right) \leq \text{min_PL} < \text{min_sup}_{\text{DB}}$$

where min_sup(DB) is least help count for a unique information base, max supp is the greatest help count of thing sets, current size is various exchange of a unique data set, and inc_size is the most extreme number of new transactions. Here, a promising successive thing set is characterized as following definition:

A rare thing set that meets requirement 1 is a promising succeeding thing set. Apriori calculation is used in this research to find all possible sequential k-thing sets as well as promising ordinary k-thing sets. For each emphasis, Apriori filters all exchanges of a unique information base using two phases of measures: join and prune. Things in both incessant k-thing sets and promising successive k-thing sets can be merged in the join step, unlike in usual Apriori calculations. Its help count should be higher than a client indicated least help count edge for a persistent item, and it should be higher than min PL but not higher than min PL for a promising successive thing.

A remarkable database might allow install new trades. This might invalidate existing association leads just as start new alliance rules. Keeping up alliance rules for an incredible informational collection is a huge issue. Thusly, this paper proposes one more computation to oversee such reviving situation. Our assumption for the new computation is that the experiences of new trades continuously change from special trades. As demonstrated by the assumption, the estimations of old trades, gained from past mining, can be utilized for approximating that of new trades. Thus, support check of thing sets got from past mining may imperceptibly not equivalent to help count of thing sets ensuing to embedding new trades into an exceptional database that contains old trades. The new estimation uses the most outrageous assistance count of 1-itemsets obtained from past mining to evaluate uncommon thing sets of a remarkable informational collection that will be prepared for being progressive thing sets when new trades are inserted into the principal informational collection. With the most outrageous assistance count and most prominent size of new trades that grant implant into a one of a kind informational index, maintain mean uncommon thing sets that will be prepared for ceaseless thing sets, for instance min_pL , is shown in Eq. 1, where min_sup(DB) is least assist with importance a novel database, max supp is most noteworthy assistance check of thing sets, current size is different trade of a remarkable informational collection and inc_size is a most outrageous number of new trades. Here, a promising standard thing set is described as keeping definition: A promising progressive thing set is something uncommon set that satisfies the condition 1. In this paper, Apriori estimation is applied to find all possible unending

k-thing sets and promising progressive k-thing sets. Apriori channels all trades of a one of a kind informational index for each accentuation with two phases measures are joined and prune step. Unlike normal Apriori estimation, things in both nonstop k-thing sets and promising ceaseless k-thing sets can be solidified in the join step. For something standard, its assist count with being higher than not really settled least assist count with edging and for a promising persistent thing, its assist check with being higher than min_PL yet not actually the customer demonstrated least assistance count.

3.5 Updating Frequent and Promising Frequent Items

An old incessant k-thing may transform into a rare k-thing, and an old promising continuous k-thing could turn into a consecutive k-thing, when new exchanges are added to a unique knowledge base. As a result, new affiliation rules are introduced, and certain existing affiliation rules are rendered invalid. When new exchanges are introduced to a unique data set, all k-things should be refreshed to address this issue. In this segment, we'll show you how to spruce up any old item. When new exchanges are embedded into a unique data set, the size of the refreshed data set grows. Min PL should be updated in this fashion to correspond to the new size of a refreshed data collection. min PL (update) is processed in the same way as the following Eq. 2:

Then, at that point, if any k-thing has support count more prominent than or equivalent to min_sup(DBUdb), this thing set is moved to a regular k-thing of a refreshed information base. In the other case, if any k-thing has support count not exactly min_sup(DBUdb); however, it is more noteworthy or equivalent to min_PL(update), and this k-thing is moved to a guarantee continuous thing set of a refreshed data set. The accompanying calculations are created to refresh continuous and promising regular k-items of a refreshed data set. Right when new trades are added to a one of a kind database, an old relentless k-thing could transform into an uncommon k-thing and an old promising consistent k-thing could transform into a normal k-thing. This presents new alliance rules, and some current connection rules would get invalid. To deal with this issue, all k-things ought to be invigorated when new trades are added to a remarkable informational collection. In this part, we reveal how to revive each and every old thing. The size of a revived database additions when new trades are installed into a one of a kind informational index. Thusly, min_PL ought to be recalculated to associate with the new size of an invigorated database. min_PL (update) is handled as the follows:

Condition 2:

By then, if any k-thing has maintained check more conspicuous than or comparable to min_sup(DBUdb), this thing set is moved to a persistent k-thing of a revived informational collection. In the other case, if any k-thing has maintained check not actually min_sup(DBUdb) yet it is more unmistakable or comparable to min_PL(update), and this k-thing is moved to an assurance persistent thing set of an invigorated database.

Table 5 Transaction data and candidate 1-item sets

TID	Item Set	Item Set	Support
1	A,B,E	A	7
2	B,D	B	7
3	B,C	C	6
4	A,B,D	D	2
5	A,C	E	3
6	B,C		
7	A,C		
8	A,B,C,E		
9	A,B,E		
10	A,C		

Table 6 Candidate, frequent, and promise frequent 2-item sets

C2	Support	L2	Support
AB	4	AB	4
AC	4	AC	4
AD	1	PL2	Support
AE	3	AE	3
BC	3	BC	3
BD	2	BD	2
BE	3	BE	3
CD	0		
CE	1		
DE	0		

The going with estimations are made to invigorate relentless and promising persistent k-items of a revived informational collection (Tables 5, 6, and 7).

Table 7 Candidate, frequent, and promise frequent 3-itemsets

C3	Support	PL3	Support
ABC	1	ABE	3
ABE	3		
ACE	1		
BCD	0		
BCE	0		
BDE	0		

4 Conclusion

This proposed calculation has higher running time than existing FUP steady algorithm. This calculation productively tracks down the incessant things and powerfully the things can be added. The whole history of continuous thing information base added and put it into independent bunches. Finally, we look at and pick the best-bought things of the client and furthermore predict the past bought things in the set of experiences. Based on the yield, we can undoubtedly track down the current status of the client bought.

References

1. Frequent Itemset Mining Dataset Repository, http://fimi.ua.ac.be/data (2004)
2. Apache Hadoop, http://hadoop.apache.org/ (2013)
3. Apache Mahout, http://mahout.apache.org/ (2013)
4. R. Agrawal, J. Shafer, Parallel mining of association rules. IEEE Trans. Knowl. Data Eng., pp. 962–969 (1996)
5. R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in *Proceedings of VLDB*, pp. 487–499 (1994)
6. G.A. Andrews, *Foundations of Multithreaded, Parallel, and Distributed Programming* (Addison-Wesley, 2000)
7. R.J. Bayardo, Jr., Efficiently mining long patterns from databases. SIGMOD Rec., pp. 85–93 (1998)
8. M. Boley, H. Grosskreutz, Approximating the number of frequent sets in dense data. Knowl. Inf. Syst., pp. 65–89 (2009)

A Novel Approach to Privacy Preservation on E-Healthcare Data in a Cloud Environment

Kirtirajsinh Zala and Madhu Shukla

Abstract Cloud environment enables healthcare professionals to work together. Despite its many benefits, it faces several difficulties: technical, legal, and managerial. Cloud computing can enhance healthcare facilities by utilizing its skills to assist information transfer within the health monitoring system, and without any geographical limitations, users such as patients, physicians, pharmacists, and health insurance agents can obtain health-associated information at any moment. With the widespread use of healthcare information and communication technology (ICT), creating a stable and sustainable data sharing scenario has attracted increasing interest in both academic research and the healthcare sector. This paper evaluates and compare the present situation with the security criteria for cloud-based medical e-health record. In addition, this paper outlines proposed cloud model and comparison of our proposed cloud model with different frameworks listed in different research papers are discussed through which how we can safeguard patient e-healthcare record over cloud environment.

Keywords EHR · Cloud computing · Privacy preserving · Health · Security

1 Introduction

Cloud computing provides cost-effective alternatives through various services such as storage, computing resources, and much more. Since big quantities of data stored in the cloud contain sensitive information, storing this information with third-party service providers presents a major danger to information privacy. The information location in the cloud is vibrant and depends on different variables including network velocity and storage accessibility. In this situation, traditional mechanisms for data protection to safeguard the user information at a known place fail owing to uncertainty

K. Zala (✉) · M. Shukla
Department of Computer Engineering, Marwadi University, Rajkot, India
e-mail: Kirtirajsinh.zala@marwadieducation.edu.in

M. Shukla
e-mail: Madhu.shukla@marwadieducation.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_15

with regard to user data place. In today's healthcare settings, there is a powerful need to build an infrastructure that minimizes time-consuming and expensive operations to obtain a patient's full medical record and uniformly integrates this diverse collection of medical data to deliver it to healthcare professionals. EHRs are becoming more widely used, allowing healthcare professionals, insurance companies, and patients to generate, handle, and access patient healthcare information from any location and at any time. Threat to individual privacy, inference of sensitive information including personal information, patient health records splits among different stakeholders like insurance company pharmacy doctor or even patterns from non-sensitive information need to be protected [1–6]. Privacy preservation in data mining (PPDM) is a relatively recent development. Many researchers are looking into how this technology can be used in the field of cloud computing to secure data in the cloud. Furthermore, plenty of techniques have been proposed. The two factors of PPDM are data hiding and rule hiding. Data hiding methods aim to keep individuals' sensitive data private and to modify data mining algorithms so that sensitive data cannot be presumed from the results of data mining algorithms. Rules are hidden to safeguard output privacy [7].

1.1 What Is Cloud Computing?

Cloud computing relates to the remote control, configuration, and access of resources to hardware and software. It provides Internet data storage, infrastructure facilities, and implementation.

1.2 Deployment Model

Cloud can have any of the four types of access: public, private, hybrid, and community [2, 6, 8].

1.3 Service Models

Cloud computing is based on service models. These are categorized into three basic service models which are.

Infrastructure-as-a-Service (IaaS): IaaS provides access to basic resources, including physical, virtual, virtual, and storage. IaaS quickly increases demand and allows you to only pay for what you use. It enables you prevent the cost of purchasing and handling your own physical servers and other information centers [2, 6, 8, 9].

Platform-as-a-Service (PaaS): PAAS offers the apps, development and deployment instruments, etc., at runtime environment. PaaS enables you to prevent software

license purchasing and management costs and complexities, fundamental application, and middleware infrastructure or development instruments and other resources [2, 6, 8, 9].

Software-as-a-Service (SaaS): The SaaS framework enables end-user services to use software applications. Users can connect and use cloud-base applications over the Internet using the SAAS service [2, 6, 8, 9].

1.4 Security and Privacy Concerns in the Healthcare Cloud Computing Environment

There are several challenges to ensuring the level of data security and privacy over cloud in healthcare environment. Despite its vast possibility and dynamic process, cloud privacy, security, and trust continue to be areas of concern and uncertainty. The following issues have been identified in the cloud computing environment for health care.

Virtualization: This way many users can share the same physical resources by developing virtual ones. Multiple virtual machines are modeled on the same hardware in a multi-tenant setting to allow resource pooling [1, 2, 6, 8].

Information and Storage Issues: In order to meet client requirements, this fresh paradigm depends on distributed systems. To that purpose, physical servers are distributed in various geographical places across various data centers, so there are different threats to the privacy and accessibility of the information in the cloud environment like vulnerabilities for data recovery, unsanitary media, and data backup [1, 2, 6, 8].

Web Services on Internet: It provides enormously powerful access to remote data service. Despite the many advantages of the key component, Web technology use can compromise client data privacy and put them at risk of various security issues like session management, SQL injection, cross-site scripting, abbreviations, and acronyms [1, 2, 4, 6, 8].

Data Confidentiality and Integrity: The confidentiality of cloud computing medical image processing is the way medical content and patient data are being kept confidential. It therefore protects unauthorized users from disclosing medical information. Cloud computing medical image processing protects the confidentiality of medical contents and patient information. As a result, it prevents unauthorized users from accessing medical information. The medical image content in the health system must be preserved. The primary challenge for providers who offer cloud-based medical image processing as a service is thus ensuring the integrity of medical images. [1, 2, 4, 6, 8, 10].

Data Availability and Ownership: The diagnosis and clinical decisions of medical image processing tools are of major relevance. In reality, medical images provide instant access to critical data that assists physicians in diagnosing and treating

patients. As a result, this software should always be available and accessible to authorized users. Load balancing methods are also used to improve accessibility and efficiency. The process of establishing a link between medical records and their owners is commonly referred to as data ownership. In this case, the owners should have legal rights, complete control, and unrestricted access to medical data [1, 2, 4, 6, 8, 10].

2 Related Work

In this section, we briefly review the existing work. Author proposed secure multiparty computing protocol (SMC). Pailler scheme based on homographic encryption techniques is used to avoid possible disclosure of information in a collaborative environment [1]. The other proposed technique is based on the fully homographic encryption (FHE) algorithm, the attribute-based encryption (ABE), and the indexed searched algorithm, and it allows healthcare professionals to handle encryption keys in the cloud while also splitting the encryption key, so that patient information are never made available to a cloud service provider [2]. The author uses Depsky cloud storage to store health data, and the DepSpace protocol is responsible for access control. The author proposes a solution to the problem of maintaining user access control and preserving trust while storing medical information in groups. Access control (sharing) and encrypted (secure) data-based work is used [3]. Here, author suggested technique is based on the residue number system (RNS) homomorphic encryption system. Technique presented enables direct operation over the encoded data and hence facilitates the complete privacy protection [11]. With diverse users of EHR with different access privileges and permissions, the study presents distributed clinical data exchange via dynamic access control privacy employing hybrid clouds, as well as access control policy transformation and cryptographic building blocks [4]. Medcloud platform is proposed for developers to use in application development, and Restlet, a Web portal, is presented to users, to access the Medcloud system. People from different places can easily access and use, as well as help the developers in creating their own healthcare applications sharing the EHR [12]. In this paper author evaluates performance of cryptographic based identity-based encryption (IBE), and newly identity based proxy re-encryption scheme (IBPRE) which is designed for securing privacy of EHR [13]. On privacy access control, author proposed a prototype which is a combination of RBAC, PBAC, MAC, and DAC access control models. The current prototype is capable of demonstrating the process of setting the standard of access for patients and the health authority and handling access requests by health professionals [5]. The author has planned and built a simulative method for patient doctor apps to link the patient with the doctor in order to secure the privacy of patients [14]. The paper using image steganography for providing improved medical data security proposed an improved image steganography for securing medical data that anonymized and secures compressed data using swapped Huffman tree coding [15]. In order to exchange health information in the cloud, the author explores essential resources related to health information sharing and incorporation in health care

and examines emerging problems and issues. Discussion on a range of possible options and opportunities to incorporate EHR in the healthcare cloud [16]. In cloud patient data privacy and security in sensor-cloud infrastructure, author proposes generic framework steps for patient psychological parameters (PPPs) to achieve patient privacy and security in S-CI extracted from available literature [17]. For secure management of medical e-health record in cloud, author sets security functions like an encryption and authentication in the system; author suggested security steps and security process with digital rights management (DRM) to protect patient record in cloud [18]. The author proposes a cloud computing framework for EMR system collaborations to securely share health and medical records among federations of healthcare information systems, enabling for the information exchange of medical records among healthcare professionals with no preexisting relationship on a global scale. The patient-provided selective access grants govern the sharing [19]. In this paper, paper shows that using the attributes-based encryption method is a safe way to access health information in cloud computing [20]. Many authors have used cloud simulator in the future. Cloud simulator is a virtual environment framework that allows for the modeling, simulation, and experimentation of evolving cloud computing infrastructure and application services [21]. There are some research challenges in the cloud in healthcare environment that have been briefly described by certain authors. How do users secure and protect the safety of data stored in the cloud, and how do they use data storage protection to ensure the safety of medical records? Which authentication scheme will be more efficient for secure EHR transfer? What encryption scheme can be used to ensure healthcare data security? These issues must be addressed in the future work [1, 2, 4, 9, 12].

3 Proposed Model

The patient's medical history is essential for making the right decision and assisting healthcare staff in responding as quickly as possible. This information is extremely private and must be kept private for the patient's sake. Medical records should be kept accessible at the same time. To ensure that a patient's data is accessible to all healthcare institutions, it can be accessed anywhere. This is a problem because unauthorized users can access patient data because of the large number of users, associations, and patients and the frequent changes in privileges that must be made to ensure anonymity. This paper proposes a cloud framework for managing user access control in a complex universe for user data and ensuring confidentiality by storing user's healthcare data in the cloud (Fig. 1).

Step-1: Every user has to register with the system

Step-2: User information will be checked using the user management module, and the user registration details will be sent to the crypto module, which will store all data using the DES technique.

Fig. 1 Proposed model

Step-3: If registered user, e.g., patient needs to store his/her data on the cloud, patient can transfer all data to the cloud after login via steganography module, which will hide all the EHR of the patient in one image and provide the EHR of the patient with protection.

Step-4: Patient will provide access to his/her EHR data to doctor/hospital/user through access control module. At the time of defining access rights to its EHR, patient will define some sensitive EHR will not be seen to anybody which will be protected by anonymization module.

Step-5: Now, registered doctor/hospitals/user wants to access patient EHR; then, they have to login with their credentials and will request EHR of patient through access control module.

Step-6: Access control module checks Dr./hospital/user access rights/privileges and contact steganography module to provide the requested EHR of the patient stored in image.

Step-7: Steganography module will reverse the process and will give EHR of patient to access control module.

Step-8: Access control module will verify EHR with anonymization module.

Step-9: Anonymization module will hide the specified critical EHR of the patient that cannot be exposed to anyone, and the rest of the EHR will be provided to the access control module.

Step-10: Lastly, access control module will give requested EHR of patient to doctor/hospital/user.

So, here, author will be implementing below module in proposed model.

1. Cryptography encryption techniques to protect patient credentials in cloud.
2. Steganography to hide patient's EHR in image.

Table 1 Comparison with different frameworks

Paper Reference No.	Any approach?	Information hiding?	Data mining technique used?	Patient privacy maintained	Storing medical data using encryption technique?	Framework available?	Any techniques for sharing medical data over cloud available?
[12]	Yes	No	No	Yes	No	Yes	No
[1]	Yes	No	No	Yes	Yes	Yes	No
[3]	Yes	No	No	Yes	Yes	Yes	Yes
[2]	Yes	No	Yes	Yes	Yes	Yes	No
[4]	Yes	No	No	Yes	Yes	Yes	Yes
[20]	Yes	No	No	No	Yes	Yes	Yes
[19]	Yes	No	No	Yes	No	Yes	Yes
Our proposed cloud model	Yes	Yes	Yes	Yes	Yes	Yes	Yes

3. Access control module to set privileges and rights for patient's EHR to hospitals/doctor/user by patient itself.
4. Anonymized private field based on user choice provide data hiding mechanism.
5. Effectively mine the data to generate the required result.

4 Proposed Cloud Model Comparison with Different Frameworks

See Table 1.

5 Conclusion and Future Work

In this paper, we have discussed security techniques used in cloud environment to safeguard the information and comparison of different frameworks describe in different papers with the proposed cloud model presented in depth in this paper. Managing user access control, guarantee confidentiality of patient's data are complex task. So, our main aim in future publication is to mine the data efficiently, preserve privacy, data hiding, secure storing of sensitive data, and adequate access control for users. For future work, we intend to further investigate and implement the proposed

model in the context. As the proposed cloud model is in stage of development, hence, actual results will be shared in future publication.

References

1. M. Marwan, A. Kartit, and H. Ouahmane, "A Cloud Based Solution for Collaborative and Secure Sharing of Medical Data," Cloud Security, pp. 1528– 1547, 2019.
2. H. Elmogazy and O. Bamasak, "Towards healthcare data security in cloud computing," 2013 IEEE Third International Conference on Information Scienceand Technology (ICIST), 2013.
3. D. R. Matos, M. L. Pardal, P. Adão, A. R. Silva, and M. Correia, "Securing Electronic Health Records in the Cloud,"Proceedings of the 1st Workshop onPrivacy by Design in Distributed Systems - W-P2DS18, Apr. 2018.
4. F. Rezaeibagha, Y. Mu, Distributed clinical data sharing via dynamic access-control policy transformation. Int. J. Med. Informatics **89**, 25–31 (2016)
5. Randike Gajanayake,Renato lannella,Tony Sahema," Privacy Oriented Access Control for Electronic Health Records," InData Usage Management on the WebWorkshop at the Worldwide Web Conference, ACM, Lyon Convention Centre,Lyon, France2012.
6. N.A. Azeez, C.V.D. Vyver, Security and privacy issues in e-health cloud-based system: A comprehensive content analysis. Egyptian Informatics Journal **20**(2), 97–108 (2019)
7. H. Vaghashia and A. Ganatra, "Applications, vol. 119, no. 4, pp. 20–26, .A Survey: Privacy Preservation Techniques in Data Mining,"International Journal of Computer-2015.
8. M. Marwan, A. Kartit, and H. Ouahmane, "Using cloud solution for medical image processing: Issues and implementation efforts," 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), 2017.
9. Chenthara, Shekha, Khandakar Ahmed, Hua Wang et al. "Security and Privacy-Preserving Challenges of e-Health Solutions in Cloud Computing." IEEE Access, vol. 7, 30, pp. 74361–74382., doi:https://doi.org/10.1109/access.2019.2919982, May 2019.
10. Monica Adriana Dag¸ Emil Ioan ¸Razvan Dobre," Data Hiding Using Steganography"IEEE 12th International Symposium on Parallel and DistributedComputing, 2013.
11. M. Gomathisankaran, X. Yuan, and P. Kamongi, "Ensure privacy and security in the process of medical image analysis,"IEEE International Conference on GranularComputing (GrC), 2013.
12. Dalia Sobhy, Yasser El-Sonbaty and Mohamad Abou Elnasr," MedCloud: Healthcare Cloud Computing System," The 7th International Conference forInternet Technology and Secured Transactions (ICITST-2012).
13. X.A. Wang, J. Ma, F. Xhafa, M. Zhang, X. Luo, Cost-effective secure E-health cloud system using identity based cryptographic techniques. FutureGeneration Computer Systems **67**, 242–254 (2017)
14. Farheen Pathan, Prof. Jadhav H. B.," Patient Privacy Control for Health Care in Cloud Computing System," International Research Journal of Engineering andTechnology (IRJET),vol.04,issue 07,july-2017.
15. M. A. Usman and M. R. Usman, "Using image steganography for providing enhanced medical data security," 2018 15th IEEE Annual ConsumerCommunications & Networking Conference (CCNC), 2018.
16. Y. Guo, M.-H. Kuo, and T. Sahama, "Cloud computing for healthcare research information sharing," 4th IEEE International Conference on Cloud ComputingTechnology and Science Proceedings, 2012.
17. I. Masood, Y. Wang, A. Daud, N.R. Aljohani, H. Dawood, Towards Smart Healthcare: Patient Data Privacy and Security in Sensor-Cloud Infrastructure. Wirel. Commun. Mob. Comput. **2018**, 1–23 (2018)
18. H. Ko, L. Mesicek, J. Choi, J. Choi, S. Hwang, A Study on Secure Contents Strategies for Applications With DRM on Cloud Computing. InternationalJournal of Cloud Applications and Computing **8**(1), 143–153 (2018)

19. S. Shenai and M. Aramudhan, "Cloud computing framework to securely share health & medical records among federations of healthcare information systems," Biomedical Research, Jun.2017.
20. K. Geetha, Impact of cloud database in medical healthcare records based on secure access. Int. J. Eng. Adv. Technol. Regular Issue **8**(6), 652–654 (2019). https://doi.org/10.35940/ijeat.f8097.088619
21. P. Humane, J. Varshapriya, Simulation of cloud infrastructure using CloudSim simulator: a practical approach for researchers, in *International Conference on SmartTechnologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, pp. 207–211, May 2015

Temporal Reasoning of English Text Documents

Parul Patel

Abstract Temporal expression recognition and normalization (TERN) is promising research area in the field of natural language processing. Automatic temporal information extraction system is in great demand due to rapid growth of digitized information on the Internet. Lot of research has been done in temporal information extraction. But due to variety of temporal expression present in the textual data, interpretation needs more efforts. It is important to understand nature of expression, and then, its interpretation ways can be defined. For example, "on the same day," it requires information about previous sentences in the text to define reference time rather than considering document creation time as reference time statically. In this paper, a normalization system is proposed that dynamically chose the reference time from text rather than relying on traditional approach of selecting a document creation date as reference time. The system is evaluated on gold standard dataset like WikiWar, AQUAINT, and TempEval-2 datasets.

Keywords Temporal expression · Temporal tagger · TIMEX · TERN

1 Introduction

Due to rapid increase in the digitization, amount of information has increased drastically on Internet. Time plays fundamental role in retrieving meaningful information from it. Recognizing time based information from text document becomes an emerging research area in natural language processing. For understating a text from time perspective, it is important to interpret temporal expressions that are available in various forms. Understanding temporal expressions require contextual information around text. For example, "On the same day, decision of making of power plant has been taken." In this temporal expression, "the same day" requires contextual information around the text for proper interpretation. Lot of work has been done in the temporal information extraction domain, but interpretation of implicit and relative

P. Patel (✉)
Department of ICT, VNSGU, Surat, India
e-mail: pjpatel@vnsgu.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_16

temporal expression is done on the basis of fixed reference time like document publication time, document creation time, news date etc. In this paper, a novel approach is proposed to determine reference time dynamically based on the type of temporal expression rather than earlier static method. Rest of the paper is organized as follows: Sect. 2 is review of literature in the field of temporal reasoning. Section 3 describes research methodology adopted for novel approach. Section 4 shows evaluation and results of the implemented algorithm. Section 5 concludes the work and gives path to future work.

2 Literature Review

Temporal information can be expressed in variety of ways in natural language. Approaches to extract temporal information includes, machine learning-based, rule-based, data-driven-based etc. HeidelTime [1] is a rule-based temporal tagger based on UIMA (unstructured information management architecture). SUTime [2] is based on CFA (cascade of finite automata) with heuristics. Trips and Trios [3] is a system developed by UzZaman et al. by using conditional random field and hidden Markov models. Deep learning is also becoming very popular in the natural language processing. Kim et al. [4] have used recurrent neural network for recognizing temporal expressions. T-YAGO [5] uses knowledge base (KB) to extract temporal information. It extracts temporal information from Wikipedia, infoboxes, categories, and lists. ParsTime [6] is rule-based temporal tagger developed for Persian language. In [7], CRF-based approach is used to extract temporal expressions from English text document.

3 Research Methodology

Temporal information processing system also called temporal taggers involves two sub tasks:

(1) Temporal Expression Extraction
(2) Interpreting each temporal expression and converting into a standard format.

In this paper, main focus is on interpreting temporal expression by looking text around it in a sentence rather than considering document creation date, time as a static reference time. In this section, we introduce the normalization task that is used to convert the temporal expressions recognized by recognizer module into standard format. At first glance, it looks very easy to interpret temporal expressions by calculating the offset value from specific reference time, but it is not always true. There are some problematic cases that need very complex solutions. Normalization module depends upon external knowledge for interpreting implicit expression. We describe the fixed language specific knowledge. It is a rule format for temporal

reasoning. Rules are operated on the basis of the matching pattern of the temporal expression. It is applicable only when time pattern matches. Rules are divided into two parts: constant and functions. Constant part is defined as follows:

%DCT, REFTIME, OFFSET, DCT_YEAR, DCT_MONTH, DCT_DAY

DCT is a document creation time, REFTIME is a reference time to for normalization, OFFSET defines the direction to be chosen for relative temporal expression, DCT_YEAR, DCT_MONTH, DCT_DAY defines date, month, and year of the document creation date. There are set of functions applied to calculate different normalization values.

For example:

ADD (date, granularity, number)

It calculates the date after adding the number in the corresponding granularity. For example Add (2017-04-01,day,-2) will output 2017-03-30. We have used SUTime's basic rules for interpreting explicit temporal expression. For interpreting implicit and relative temporal expression, we need to identify reference time correctly. Moreover, external knowledge need to be fed into the system to handle named temporal expressions. To impart such knowledge related to a festival or holidays, a special file is developed with the knowledge of it as an external information to be embedded in a system. Moreover, in a country like India, festival do not occur on a same date every year as it follows specific religion calendar. To deal with such situation, we have collected dates of such Indian festival through year with their synonyms in different regions of India along with their values. To generate a knowledge base for the holidays, first we have extracted name of festivals occurring on fixed day or on a variable day from Wikipedia. Data about festivals and special events are gathered from various sources like Wikipedia, different calendars available on Internet, and other resources available on Web that focuses on Indian festivals. Based on available data, rules are designed to interpret such expressions. Once the knowledge base is created, it can be extended for adding more information about other named temporal expressions and event. After generating knowledge base, we have generated algorithm to select the reference time dynamically. Generally, publication time of the document is referred as a reference time for all temporal expressions which is not always true. We control discourse-level information in order to find out reference time by using previously mentioned temporal expression and handling TIMEX that are affected by tenses of the verb. Consider following example

"***News: 2020-04-18***
"*A CEO of XYZ company has declared that they are going to purchase laptop, desktop and other required equipment for their organization for giving improved performance on* ***15th March 2020****,* ***After two days,*** *tender was published in local newspaper.* ***After one month****, a contract is given to ABC 'company for purchase and maintenance. Final delivery is on* ***this Thursday***".

In above example, publication time of the document is "2020-04-18." But the expression after two days relies on the expression "***15th March 2020*** of earlier sentence. However, another expression "this Thursday" depends on the publication time. In short, there cannot be only one logic to normalize temporal expressions. In the above example, following temporal expressions are extracted by recognition module:

(a) 2010-04-18,
(b) March 8, 2010,
(c) Week, (followed by modifier ***after*** and number = 1),
(d) Thursday (followed by modifier ***this***).

First two temporal expressions are explicit temporal expressions. Next two are context specific expressions which need a specific reference time to resolve it. First relative temporal expression depends upon the temporal expression of earlier sentence whereas last temporal expression depends on publication date of the news fragment. So, here, document publication date is treated as a publication time and temporal expressions occur in the sentences around are known as local temporal expressions. We have analyzed corpora like WikiWar, aquaint, and TempEval to get clear difference between two different type of relative temporal expression (1) expressions that depends upon publication time (2) expressions that depends upon the heuristic around sentence, and finally concluded various expressions in both the categories as shown in following table: We refer temporal expressions that depends upon the publication time as global temporal expressions, whereas temporal expressions that depends on its previous sentences are termed as local temporal expressions. Expressions like next month, this Friday, tonight, and this year depend on the publication time or a global reference time like document creation date, document publication date etc., and expressions like that year, on the same day, before one day, after one day etc., may depend on the TIMEX that present in the previous few sentences. Finally, all extracted temporal expressions have assigned their respective class for interpretation as: (1) Local and (2) Global. Normalization algorithm works as follows:

(1) All explicit expression that do not require any other information for interpretation are converted into a TIMEX2 format yyyy-mm-dd. Example of such expression is shown in Table 1.

(2) In this step, all relative temporal expressions are assigned their offset value based on their temporal modifier.

Table 1 Normalizing explicit temporal expressions

No.	Expression	After normalization
1	3rd January 2010	2010-01-03
2	Sunday, 14 November 2010	2010-11-14
3	15/03/1982	1982-03-15
4	November 2016	2016-11

Table 2 Offset direction from reference time

No.	Temporal expression	From reference time
1	Last Monday	< D1
2	Next Thursday	> D4
3	This coming Sunday	> D7
4	Next December	> M12
5	Last January	< M1

– I will see you next Friday
– They had a meeting last Wednesday.

In this example, temporal modifier like next, last are separated to decide the direction of offset in normalization (Table 2).

(3) Once offset is decided, then reference time for temporal expression is chosen. As shown in above example, relative temporal expression can be considered deictic whose reference time depends upon the publication time where anaphoric expression relies on the context of earlier sentence. Some of the expressions are event-based where event is an external knowledge to be fed into a system for its interpretation. Consider following examples:

(a) *"After meeting, they have decided to accept proposal."*
(b) *"He will visit china in next Christmas."*

In sentence (a), to resolve a temporal expression "after meeting," first the time of meeting needs to be resolved which is out of the scope of this work. In sentence (b), to resolve temporal expression "next Christmas," system has to externally add a knowledge about the date when Christmas occurs. For resolving such expressions, after detecting their offset direction, their year granule is identified by taking publication time as a reference time. For example, if the publication date of document is "20-02-2017" then, 2017 will be taken as year to decide final year of the expression. For example, if the expression is next Christmas, then it is incremented by 1 and so on. Once the granule is decided, then finally granule + festival name is searched in the data file to get final value.

Following is an algorithm to find out the reference time dynamically.

```
1: Load all temporal expression of Document d into a list Tlist
2: Assign label to each expression as local or global for choosing reference time.
3: Initialize GRT and LRT with either RT(reference time) or PT(publication
 Time)
//(Divide TE  into two part: temporal modifier and temporal noun  (it means after
 one week will be divided as temporal modifier: after,  temporal noun =one week
 another e.g. previous month  then temporal noun= month and //temporal
 modifier= previous)//
4:  For each TE in a TList
        TE'=Segment(TE)
        If(TE'= explicit) then
            LRT= TE'
            Insert(TE',TList1)
        Else if(TE'=Local)
            TE'= Normalize(TE',LRT,offset)
        Else
            TE'=Normalize(TE',GRT,offset)
        Endif
            LRT=TE'
            Insert(TE',TList1)
        Endif
```

To deal with weekdays, straightforward rules are not always applicable. Consider following sentence.

(c) Meera did not go to work on Tuesday.
(d) Shlok will go to school on Monday.

In above two sentences, sentence (a) contains temporal expression “On Tuesday” and sentence (d) contains temporal expression “on Monday.” During interpretation of such expression, selecting offset direction is quite complex as “On Tuesday” in sentence (c) focus on previous Tuesday, whereas “On Monday” in sentence (d) focus on next Monday. To deal with such weekdays expression with modifier that do not specify the direction, we have applied following algorithm (Fig. 1):

```
1. Insert TE (with Weekdays)
2. Move backwards in the sentence till the beginning of the sentence to find
   verb
3. If found(verb) =true then
        Find Tense of the Verb
   Else
        Move forward in the sentence till the end of the sentence to find verb
            If found(verb)=true then
                Find Tense of the Verb
            Else
                Jump towards previous sentence.
                Goto step 2.
            End If
   End if
4. If Tense(Verb)= "PS" || " PP" || "PC"||"BPS  then
        Offset.direction="Past"
   Else
        Offset.direction= "Future"
   End
```

Fig. 1 Snapshot showing output of the system for choosing reference time dynamically

Table 3 Results on different dataset

Dataset	Precision	Recall	F-measure
TempEval-2	0.94	0.91	0.92
Aquaint	0.88	0.9	0.88
IndiaTimes	0.89	0.85	0.86
TempEval-3	0.91	0.9	0.9
WikiWar	0.92	0.91	0.91

4 Evaluation and Results

The proposed algorithms are tested on the gold standard dataset TempEval-2, Aquaint, TempEval-3 shared task, WikiWar and IndiaTimes. WikiWar contains 22 historical documents, Aquaint contains 73 documents, TempEval-2 and TempEval-3 contains 325 and 21 documents, respectively. The result is calculated in terms of precision, recall, and F-measure. Precision, F-measure, and Recall are calculated as follows (Table 3).

$$\text{Precision} = \text{TRUE POSITIVE}/(\text{TRUE POSITIVE} + \text{FALSE POSITIVE})$$
$$\text{Recall} = \text{TRUE POSITIVE}/(\text{TRUE POSITIVE} + \text{FALSE NEGATIVE})$$
$$\text{F} - \text{Measure} = 2 * (\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall})$$

5 Conclusion

In this paper, automatic temporal information recognition and normalization system is proposed that focuses on all types of temporal expressions, explicit, implicit, and relative. Two algorithm for choosing reference time for interpretation of implicit temporal expressions are proposed and implemented. Major focus of this research is to understand temporal semantics to avoid ambiguity that arise in natural language processing. Finally, the proposed algorithms are evaluated on standard dataset.

References

1. J. Strötgen, M. Gertz, HeidelTime: high quality rule-based extraction and normalization of temporal expressions, Proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010, pages 321–324, Uppsala, Sweden, 15-16 July 2010.
2. A. Chang, C. Manning, SUTIME: a library for recognizing and normalizing time expressions (2012)
3. N. UzZaman, J. Allen, TRIPS and TRIOS system for TempEval-2: extracting temporal information from text, in *Proceedings of the 5th International Workshop on Semantic Evaluation* (2010)

4. Z.M. Kim, Y.-S. Jeong, TIMEX3 and event extraction using recurrent neural networks, pp. 450–453 (2016). https://doi.org/10.1109/BIGCOMP.2016.7425968
5. Y. Wang, M. Zhu, L. Qu, M. Spaniol, G. Weikum,. Timely YAGO: harvesting, querying, and visualizing temporal knowledge from Wikipedia. 697–700 (2010). https://doi.org/10.1145/1739041.1739130
6. B. Mansouri, M. Zahedi, R. Campos, M. Farhoodi, M. Rahgozar, ParsTime: Rule-based extraction and normalization of Persian temporal expressions (2018). https://doi.org/10.1007/978-3-319-76941-7_67
7. P. Patel S.V .Patel, Article: CRF based approach for temporal information recognition from English text documents, in *IJCA Proceedings on International Conference and Workshop on Emerging Trends in Technology* ICWET 2015(1):1–4, May 2015

Design, Development, and Integration of DMA Controller for Open-Power-Based Processor SOC

Gannera Mamatha, M. N. Giriprasad, and Sharan Kumar

Abstract While the processor is momentarily disabled or busy performing other orders in parallel, the Direct Memory Access (DMA) technology allows direct access to peripherals and memory. DMA gets control of the buses to transfer the data directly to the I/O devices. DMA completes the data transfer to all peripherals without the interference of the processor. The DMA controller supports eight channels with 32-bit data transfer, and it has an interface toward user logic for data read and write. The channel assignment is done based on priority. This project proposes to design, develop, and integrate DMA controller for open-power processor A2O core-based fabless SoC through AXI4 interface. The methodology used for designing is as follows: design state machines, develop Verilog HDL code, simulate using ModelSim Questa®, and synthesis using Vivado Design Suite-Xilinx®.

Keywords A2O core · AXI4 · DMA · Verilog HDL · Questasim

1 Introduction

Direct Memory Access (DMA) is a technique that allows an information/yield (I/O) device to send or receive data directly to or from the main memory, bypassing the CPU to speed up memory operations. A device called as a DMA regulator is in charge of the cycle (DMAC).

A DMA channel allows a peripheral to transfer data without putting undue strain on the CPU. Without the DMA channels, the CPU uses a fringe transport from the I/O peripheral to duplicate each piece of data. Using a fringe transport uses, the CPU during the read/compose step and prevents other work from being done until the activity is completed.

DMA allows the CPU to do several tasks at the same time, while data are being transferred. The CPU initiates the information exchange process. The controller can transfer data in one of three ways, which are listed below. Only, the new controller

G. Mamatha (✉) · M. N. Giriprasad · S. Kumar
Department of ECE, JNTUA College of Engineering Anantapur, Anantapur, Andhra Pradesh, India
e-mail: gmamatha.ece@jntua.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_17

has operating modes based on these modes, which are essential for improving performance and data transfer time depending on the user's application.

In *burst mode*, the framework transport is delivered solely after the information move is finished.

The framework transport is relinquished for a few clock cycles during the exchange of information between the DMA channel and I/O peripheral in cycle stealing mode, so the CPU can conduct other activities. When the transfers are complete, the CPU receives an interrupt on demand from the DMA controller signaling that the data transfer is complete, and that the CPU is ready for the next transfer.

The DMAC can take over responsibility for framework transfer only when the CPU requires it in transparent mode.

If the CPU is accessing any external memories, data stored in internal memory, such as RAM, may not be sampled correctly when accessed by DMA controller. We can conclude that if sufficient care is not taken for the memories from which the DMA controller actually gets information, there may be a cache coherence problem with the controller as well.

2 DMA Features

In general, there are many design methods and techniques for the DMA controllers design based on the application. This DMA controller reference is taken from http://www.opencores.org.

This DMA consists of eight channels and supports the both read and write transactions for the address given by the user. The features of DMA controller are as follows:

1. Eight DMA channels
2. Three priority levels
3. Support different modes
4. Clock divider for slow channels
5. Block transfer in a frame context
6. Support FIFO buffer
7. Windowed arbitration mode.

The AXI4 interfaces the master device and slave device interface. The slave device interface is used to complete DMA with the configured register operations. In some applications, the bus may not include master device, and the slave device interface of AXI is used.

Fig. 1 This describes the all sub-modules in the architecture

2.1 Microarchitecture

The slave configuration is introduced in this DMA because whenever the DMA needs to send the data to the processor which is connected at that the core processor will be acting as the master and the DMA as slave.

The micro-architecture of the DMA controller is shown below (Fig. 1):

3 Modes of Operation

3.1 General Mode

The mode of operation determines the mode in which these channels is to use the AXI read and write buses. Each of the core has been configured to operate in an "Independent" or "Joint" mode. When using the independent mode, each of the channel can operate in the normal mode or the outstanding. When using the "Joint mode," each of the channel can work in either normal or joint mode.

3.1.1 Independent Mode: Normal Channel Mode

When using this mode, each core will have to make use of independent arbiter for read and write operations. Read operations on the AXI bus and will be carried out without taking into account of write operations.

Each channel works in the following manner:

1. According to AXI, the first channel calculates the next read and write burst sizes.
2. As long as the data buffer can contain the data, the channel will emit read bursts.
3. As long as there is write data in the data buffer, the channel will send write bursts.

In the CORE0 JOINT MODE and CORE1 JOINT MODE registers, the mode is turned on.

3.1.2 Independent Mode: Outstanding Channel Mode

If the core operates in independent mode, each channel can be set to "read outstanding," "write outstanding," or both. This mode works as a normal independent mode; the only *difference* is that the use of a read outstanding the read command to be issued *after* the *write* command has been issued to what are the items in the right way are recorded. With the "write outstanding," the write command will be issued *after the read* command is issued, as long as the data are read, it is actually a read from the data buffer. It works on the assumption that the AXI slave, it will respond quickly, else the data buffer is filled or overflowing.

This mode is configured with the help of the RD_OUTSTANDING and WR_OUTSTANDING registers in channel. This mode is efficient while working with the fast responding channels.

One advantage is that latency is improved and throughput for small data buffers.

3.2 Joint Mode

When using this mode, each core will have to make use of a *single arbitrator, both* for read and write operations. In the Joint Mode of operation, the channels are carried out at the same rate, for read and write operations. The present channel locks both the AXI read and write buses, and it is going to read all of the data in the write data. To the running of this channel configuration mode, this *will only happen* in the read registers, and they can have an impact on read operations, as well as the write operations.

This mode can be configured with help of CORE0_JOINT_MODE, CORE1_JOINT_MODE registers and JOINT_MODE register in the channel.

Fig. 2 This is the Round Robin model how channel is allocated based on request

4 Fsm Logic

4.1 Channel Arbiter FSM

This FSM implements the Round Robin logic in the next state logic (Fig. 2).

Initially, it will be in the grant0 state, and based on the valid request and priority, the algorithm is executed. For the channels having the same priority, then, the allocation of channel will be done with help of the Round Robin algorithm.

req and **advance** signals are used for the checking of the valid request for a channel, and **grant** signal will be used for the allocation of channel.

The priority arbitration has been depicted below with the three different levels.

4.2 Priority of the Channel

See Fig. 3.

Three different types of priority have been implemented based on the necessity of the channel usage. In the first priority mode (Normal priority), the channel is allocated to the peripherals based on the Round Robin algorithm. In the high priority mode, the preferred channel is allocated in alternate intervals of time, and in the

Fig. 3 Three different levels of priority are implemented, and the pattern is described with the yellow color

top priority mode, the preferred channel is allocated continuously till the end of its operation.

5 Block Diagram with AXI4

See Fig. 4.

The master and slave of the DMA controller have been connected to the AXI interconnect with the separate clock wizard and reset block. This can be packaged to

Fig. 4 With the on-chip bus, the pins are extended, and connections are shown

Fig. 5 Connection of the DMA controller with A2O core is depicted

use as an IP for the user. There are so many optional features for this DMA controller like outstanding mode, peripheral control, and multi-processor support.

5.1 A2O Core Interface

See Fig. 5.

This core is open sourced, and the developed controller has been integrated with the smart connect which is present in this by default.

6 Results

Unit testing of the design which is done in Questasim is attached below.

At the rising edge of the clock, the AXI interface signals AWVALID and AWREADY are getting asserted. These signals infer that the design has been checked correctly. 216 bits of data have been transferred total, so there are eight locations shift in the next, whenever the next transaction is occurring (Fig. 6).

At the rising edge of the clock, the AXI interface signals AWVALID and AWREADY are getting asserted. These signals are used to get the handshake correctly between the master as well as the slave. With the correct and accurate handshake, the transfers and the operation will be efficient. In the same way, for the read transaction, the interface signals ARVALID, ARREADY are also getting asserted at the rising edge of the clock only (Fig. 7).

WLAST and RLAST signals are asserted after the completion of the previous transaction. These AXI interface signals are useful in depicting the end of the transactions, while the current is going to end, and in other way, it is also depicting the information like in the next cycle the data being driven at the master, and for the read

Fig. 6 WRITE transaction handshaking signals are highlighted. AWVALID and AWREADY are the two important signals which determine the correctness of the controller with the on-chip AXI4 bus

Fig. 7 READ transaction handshaking signals are highlighted and the address also same format with WRITE transaction

transactions in the next cycle, the driven data are being sampled and at the end again these signals getting asserted as shown in the waveform below (Fig. 8).

Fig. 8 At the end of each and every transaction, the WLAST and RLAST signals are asserted for WRITE and READ transactions, respectively

7 Conclusion

This work developed, simulated, and synthesized the AXI4 DMA controller, which was utilized in A20 processor-based fabless SoC. The controller achieves the optional features like cyclic buffer mode also. This proposed architecture has the potential to improve data transmission performance using AXI4. This design is elaborated and interfaced with A2O in Xilinx Vivado, and Simulation is done in Mentor Questa using slave-based function unit testing with the write and read operations. This controller supports the slow communication devices with the help of the different core and the clock divider.

The future scope of this work is mentioned as currently in this work DMA has a direct interface with the on-chip bus only. But this can be interfaced directly with the peripheral interfaces respectively.

References

1. G. Ma, H. He, Design and implementation of an advanced DMA controller on AMBA-based SoC, in *2009 IEEE 8th International Conference on ASIC* (2009), pp. 419–422. https://doi.org/10.1109/ASICON.2009.5351258
2. K. Chen, L. Qi, H. Yu, Design of two-dimension DMA controller in media multi-processor SoC, in *2008 Second International Symposium on Intelligent Information Technology Application* (2008), pp. 708–711. https://doi.org/10.1109/IITA.2008.493
3. C. Yu, C. Liu, C. Kang, T. Wang, C. Shen, S. Tseng, An efficient DMA controller for multimedia application in MPU based SOC, in *2007 IEEE International Conference on Multimedia and Expo* (2007), pp. 80–83. https://doi.org/10.1109/ICME.2007.4284591

4. Y.J.M. Shirur, K.M. Sharma, A. Aishwarya, Design and implementation of efficient direct memory access (DMA) controller in multiprocessor SoC, in *2018 International Conference on Networking, Embedded and Wireless Systems (ICNEWS)* (2018), pp. 1–6. https://doi.org/10.1109/ICNEWS.2018.8903991
5. H. Yuan, H. Chen, G. Bai, An improved DMA controller for high-speed data transfer in MPU based SOC, in *2004 Proceedings of the 7th International Conference on Solid-State and Integrated Circuits Technology*, vol 2 (2004), pp. 1372–1375. https://doi.org/10.1109/ICSICT.2004.1436811

A Comparative Study Using Numerical Simulations for Designing an Effective Cooling System for Electric Vehicle Batteries

Akshat Maheshwari, Vineet Singhal, Ajay Kumar Sood, and Meenakshi Agarwal

Abstract As technology is advancing rapidly in electric vehicle (EV), numerous automobile industries are focusing on shifting a part of their catalogue or have already started production of EVs. One of the most significant parts of an electric vehicle is battery which powers the vehicle, and it usually contains thousands of cells, which can be arranged in various combinations of series and parallel arrangements to form a module, many such modules together make a battery. Lithium-ion batteries are one of the preferred types for use in electric vehicles. The efficiency and life cycle of a lithium-ion battery depend on several factors. Battery temperature is one such critical factor and needs to be monitored to avoid early failure of the battery. Therefore, advanced cooling technology is to be incorporated into a battery to lower the temperature. This paper presents a comparative study of direct and tube cooling methods in a battery for effective cooling of an EV using ANSYS Fluent. These cooling systems are designed keeping an easier process of manufacturing in mind.

Keywords Electric vehicle · Lithium-ion battery · Battery cooling

1 Introduction

Lithium-ion batteries have been in extensive use in case of several electric vehicles because of their high energy density, operational temperature range being wider, less toxicity, and better charge cycles performance. In modern electric vehicles, there are around 5000–8000 cells (cell spec: 18,650), and these thousands of cells are divided into many modules of 200–500 cells [1]. These modules after long charge cycles and during acceleration of the vehicle produce large amount of heat which has to be removed to maintain the efficiency of the cells, retaining the capacity and number of

A. Maheshwari (✉) · V. Singhal · A. K. Sood · M. Agarwal
BML Munjal University, Gurugram, Haryana 122413, India

A. K. Sood
e-mail: ajay.sood@bmu.edu.in

M. Agarwal
e-mail: meenakshi.agarwal1@bmu.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_18

life cycles. The optimum temperature for the well-functioning of the cell is around 15–35 °C [2]. Air has always been a good option when it comes to cooling, but as the demand for more power increases more heat is generally generated; as a result, air having a low heat capacity and thermal conductivity is then not a very good coolant [3, 4], and hence, we stick to water cooling in our cases of study. We need to implement a cooling system to remove the heat produced by the batteries as fast as possible using an efficient pump. Many small parts of such battery module are designed, and a cooling system is designed for this part of a module. This cooling system uses submerged water to cool down the cells and is simulated to find results and differences in flow and heat extraction. A fine mesh is generated in Ansys, and we use a micro-pump pumping 7 l/min of distilled water at temperature of 300 K and the outlet is an open outflow. The outer wall of cell is made of aluminum, and the entirety of the cell uses a custom-made Ansys material with a specific heat capacity of 998 J/Kg K and thermal conductivity in axial direction of 35 W/mK, radial direction 0.2 W/mK, and density of 2018 kg/m^3 [2, 5]. We compute steady flow, with energy equation and K epsilon standard viscosity equations on, each wall and interference are coupled to provide the heat exchange that is expected. The simulation is run for each design at two different discharge rates which supply us with different results to analyze.

2 Methodology

After exploring various design changes (distance between cells, distance from the shell), solution methods and boundary conditions two different cooling system models are designed. Each module is made using multiple lithium iron phosphate battery (LiFePO) [6]. This kind of battery uses graphite with metallic backing as anode and lithium iron phosphate as cathode.

The lithium cells referred in this research are of type 18,650, and each cell is created as a solid piece of a generalized matter named e-material which has generalized properties; this is done to not overcomplicate the individual design of a cell while avoid designing interior and exterior of the cell (standard size: 18 mm diameter and 65 mm length) with the following parameters [3, 7]:

Nominal voltage = 3.2 V, Density = 2018 kg/m^3, Reference temperature = 300 K, Specific heat capacity = 1018 J kg^{-1} K^{-1}, Thermal conductivity in radial direction = 1.5 Wm^{-1} K^{-1}, Thermal conductivity in axial direction = 25 Wm^{-1} K^{-1}, Energy density = 325 Wh/L (1200 kJ/L) and Cycle durability = 2000–12,000 cycles.

Pump properties are [8]: Max flow rate = 7, 9, and 12 L/min, Rated voltage = 12/24 V, Impeller type is closed impeller, Motor type is brushless DC motor, Liquid temperature = 0–100 °C, Ambient temperature = −40–70 °C, Media is pure water, antifreeze, working voltage range = 6–18 V or 12–28 V and power consumption = 1–32 W.

The cells are arranged in different orders to produce small modules, and each module has an outer shell made up of aluminum. Two different types of cooling

Fig. 1 20S 4P, 80-cells (direct cooling) (dimensions are in mm)

strategies are considered for simulations here in this paper, namely direct cooling and tube cooling which are presented in the following sections.

Boundary Conditions used for the simulations are:

- Inlet mass flow rate is 0.115 kg/s and 0.015 kg/s for direct cooling and tube cooling, respectively ([4, 9–11]).
- All cells are considered as a constant source of heat with heat generation rates corresponding for 1 and 2 °C discharge rates, (5318 W/m^3 and 19,452 W/m^3, respectively) ([12, 13]).
- All cell walls in contact with the fluid are thermally coupled with the system.

Simulation for two different discharge rates 1 and 2 °C was performed.

1. Discharge rate 1 °C is used to represent the ideal discharge rate during normal driving operation.
2. Discharge rate 2 °C is used to represent when there is a higher demand in energy during acceleration and hill climbing.

The heat generation rate is figured out using practical experiments on a set of twenty 18,650 lithium-ion cells at 1.35 A for 1 °C and 2.70 A for 2 °C, and the observed heat generation rate is 5318 W/m^3 and 19,452 W/m^3 [5] (Fig. 1).

2.1 Direct Cooling

Meshes for 20S 4P, 80-cells case, and straight tube case are as shown in Figs. 2 and 4, respectively. Orthogonal quality is >0.95 in all the cases. Quality of mesh was ensured by performing a grid independence test also to ensure that the final results

Fig. 2 Mesh for 20S 4P, 80-cells

are independent of the further refinement of the mesh. For direct cooling, a mass flow rate inlet is used with 0.115 kg/s, that is, 7 L/min, and for tube cooling, the inlet mass flow rate is 0.015 kg/s ([4] and [14]), the flow rate here is decreased because the area of cross section in case of tube cooling is very small, multiple tube-cooled modules can be cooled using the same pump to prevent heavy pressure differences in the tubes. The inlet temperature is maintained at 300 k. The outlet is made to be an outflow-type outlet. To achieve this flow rate, a readily available mini-DC water pump can be used with the specifications as shown in Table 2. We have used a pressure-based, steady-state solver with gravity set at 9.8 m/s^2, energy equations are turned on for the simulations.

2.2 Tube Cooling

80-cell module (20S 4P) straight tubes (see Fig. 3).

3 Results and Analysis

Velocity Contours for 1 °C discharge rate (see Figs. 5 and 6)

Fig. 3 20S 4P, straight tube

Fig. 4 Mesh for 20S 4P, straight tube

The velocity contours clearly show that straight tube model has a higher flow rate of liquid surrounding individual cells. Turbulence isn't very high in any of the cases. The velocity contours and types of packing are important because in real conditions, acceleration, and jerks have a huge impact on the actual flow rate, backflow is also a certain occurrence and can be prevented using backflow prevention valves at bottleneck regions.

Temperature Contours for 1 °C discharge rate (see Figs. 7 and 8)

A simple look at the temperature gradient contours shows that direct cooling is the superior form of cooling, these types of modules are easier to manufacture by using a punching press and boring process, seals can be made using rubber rings. The major drawback of this cooling method is the probability of a leakage, and this drawback is more observable when the entire cooling system is in an EV which is on move,

Fig. 5 20S 4P, 80-cells (direct cooling) velocity contour

Fig. 6 80-cell module (20S 4P) straight tubes (tube cooling) velocity contour

continuously accelerating and deaccelerating. Indirect cooling solutions like tube cooling when compared to direct cooling have moderate execution but have higher complexity when it comes to manufacturing. The temperature rises up drastically with the C rating and discharge amperage.

Tables 1 and 2 show the results for the two different charges, i.e., 1 °C and 2 °C, respectively. For indirect tube cooling method, the temperature rises much more as compared to the case with direct cooling with the increase in C rating from 1 to 2 °C.

Fig. 7 20S 4P, 80-cells (direct cooling) temperature contour

Fig. 8 80-cell module (20S 4P) straight tubes (tube cooling) temperature contour

Table 1 Results for 1C

Module	Maximum cell temperature (K)	Maximum fluid temperature (K)
80-cell module (20S 4P)	300.056	300.017
80-cell module (20S 4P) Straight tubes	300.3	300.13

Table 2 Results for 2C

Module	Maximum cell temperature (K)	Maximum fluid temperature (K)
80-cell module (20S 4P)	300.198	300.094
80-cell module (20S 4P) straight tubes	301.178	300.656

4 Conclusion

To recapitulate, two approaches are presented here. In case of indirect cooling, temperature rise is observed to be more with increase in C rating from 1 to 2 as compared to that in case of direct cooling. Direct cooling as observed from temperature contours seems to be a better option for providing more effective cooling as compared to the indirect cooling method. When it comes to designing a faction of a module, the results can help to fulfill the necessities of little and medium-sized companies. The exploration and approach prompt a decrease in the expense and the complexity of designing battery cooling systems thereafter improving the project lead time.

References

1. H. Liu, Z. Wei, W. He, J. Zhaoa, Thermal issues about Li-ion batteries and recent progress in battery thermal management systems: a review. Energy Convers. Manag., pp. 304–330 (2017)
2. L.H. Saw, Y. Ye, A.A.O. Tay, W.T. Chong, S.H. Kuan, M.C. Yew, Computational fluid dynamic and thermal analysis of Lithium-ion battery pack with air cooling. Appl. Energy **177**, pp. 783–792 (2016)
3. L.H. Saw, Y. Ye, A.A.O. Tay, Electrochemical-thermal analysis of 18650 lithium iron phosphate cell. Energy Convers. Manage. **75**, 162–174 (2013)
4. J.A. Brumley, A study of the energy consumption of a battery cooling system by different cooling, Graduate Theses, Dissertations, and Problem Reports, p. 5273 (2016)
5. Y. Huang, Y. Lu, R. Huang, J. Chen, Z. Liu, X. Yu, T. Roskilly, Study on the thermal interaction and heat dissipation of cylindrical Lithium-Ion Battery cells. Energy Procedia, pp. 4029–4036 (2017)
6. P. Cicconi, D. Landi, M. Germani, Thermal analysis and simulation of a Li-ion battery pack for a lightweight commercial EV. Appl. Energy, pp. 159–177 (2017)
7. W. Tong, K. Somasundaram, E. Birgersson, A.S. Mujumdar, C. Yap, Thermo-electrochemical model for forced convection air cooling of a lithium-ion battery module. Appl. Therm. Eng. **99**, 672–682 (2016)
8. X. Li, J. Zhao, J. Yuan, J. Duan, C. Liang, Simulation and analysis of air-cooling configurations for a lithium-ion battery pack, J. Energy Storage, pp. 102–270 (2021)
9. Q. Wang, B. Jiang, B. Li, Y. Yan, A critical review of thermal management models and solutions of lithium-ion batteries for the development of pure electric vehicles. Renew. Sustain. Energy Rev. **64**, 106–128 (2016)
10. S. Park, D. Jung, Numerical modeling and simulation of the vehicle cooling system for a heavy-duty series hybrid electric vehicle, SAE Technical Paper, p. 13 (2008)
11. A. Jarrett, I.Y. Kim, Design optimization of electric vehicle battery cooling plates for thermal performance**.** J. Power Sour. **196**(23), 10359–10368 (2011)
12. M. Rajadurai, S. Ananth, Battery thermal management in electrical vehicle-review article. IJISET—Int. J. Innov. Sci., Eng. Technol. **7**(2), (2020)
13. Z. Tang, X. Min, A. Song J. Cheng, Thermal management of a cylindrical lithium-ion battery module using a multichannel wavy tube. J. Energy Eng. 145(1) (2019)
14. Y. Lyu, A.R.M. Siddique, S.H. Majid, M. Biglarbegian, S.A. Gadsden, S. Mahmud, Electric vehicle battery thermal management system with thermoelectric cooling. Energy Rep. **5**, 822–827 (2019)

Lexical Resource Creation and Evaluation: Sentiment Analysis in Marathi

Mahesh B. Shelke, Saleh Nagi Alsubari, D. S. Panchal, and Sachin N. Deshmukh

Abstract In India, the raise of regional language contents over social media, websites, blogs and news article are exponentially increasing because of ease of use of technology, and people are expressing their thoughts, opinion more conveniently and powerfully over Internet. In this paper, we evaluate the challenges of sentiment analysis in Marathi by setting up a baseline, where we produced an annotated dataset, however, initially, we created an annotated dataset consisting of Marathi news scraped from various newspaper/channel websites. Furthermore, domain experts annotated Marathi news with positive, negative and neutral polarity. And we used machine learning models such as logistic regression, Stochastic Gradient Decent (SGD), support vector machine (SVM), nearest neighbour, neural network, decision tree (DT), Naïve Bayes (NB) and proposed ensemble-based model for sentiment analysis to demonstrate effectiveness. In experimentation, the proposed ensemble classifier outperforms other classifiers with an accuracy of 94.16% and an F-score of 97.02% for fivefold validation. Also, for tenfold validation, the accuracy is 95.07%, and the F-score is 96.93%.

Keywords Marathi · Sentiment dataset · Machine learning · Indian language

1 Introduction

It is very easy at present to express thoughts on the web or on social media. After viewing films, using a product or visiting a locality, we can write movie reviews, product reviews or tourist reviews. This opinion-rich data will be interested both in the decision makers for the entities concerned and in enterprises seeking to improve their products or services. It provides people the opportunity to express themselves rather than media personalities speaking on behalf of the general majority of the

M. B. Shelke (✉) · S. N. Alsubari · D. S. Panchal · S. N. Deshmukh
Department of Computer Science & IT, Dr. Babasaheb Ambedkar, Marathwada University, Aurangabad, M.S., India
e-mail: mahesh_shelke21@hotmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_19

population. You can make your feelings heard by posting your thoughts on the Internet. As a result, vast amounts of information, including the author's point of view, are available on the Internet. It is now a challenge that these documents contain useful information. This enhances the application of sentiment analysis/ opinion mining.

Emotions play an important role in the communication and decision-making processes in our intellectual activities. Emotions are a succession of events made up of feedback loops. Feelings and behaviours, like cognition, may aspect cognition. In addition, the detection and interpretation of emotional data are vital in a wide range of fields, including human–computer interactions, e-learning, e-Health, automotive, security, user profiling and customisation. The analysis of textual sentiment is one of the popular computer linguistic methodologies.

Sentiment analysis is a process of recognising and categorising views expressed in a piece of text in a computational way, particularly to assess whether the person has a positive, negative and neutral opinion towards any certain topic, product, etc. Product reviews, Film reviews, blog and articles are the popular and available opinion-rich contents. There are three levels of sentiment analysis can be performed: Document level, Phrase/Sentence level and Entity/Aspect level. In document level sentimental analysis, the polarity is determined for the complete document. In sentence level sentiment analysis, polarity is determined for the individual sentences of the document. In the aspect level sentiment analysis, the polarity is determined for the entity/aspect of the document. Following methods can be used for sentiment analysis:

- **Using Lexicon-based Sentiment Analysis**: It is a polarity-based information dataset of phrases or word, in which scores are assigned to each word. This score describes the characteristics related to the term as positive, negative or neutral sentiment expressed in text.
- **Using N-Gram Model-based Sentiment Analysis**: It forms and uses the N-Gram model (unigram, bigram, trigram or mixed) using the categorization trainings data.
- **Using Machine Learning-based Sentiment Analysis:** Supervised and Unsupervised learning model can be used to perform prediction on data by extracting features from text.

The main objective of this paper is to develop and evaluate lexical resources for sentiment analysis in Marathi, as there are few lexical resources, libraries, Corpus and tools available in Marathi, which signifies that Marathi has not been explored in the field of sentiment analysis. The proposed approach is designed in combination with machine learning-based algorithm. The lack of Marathi preprocessing tools and annotated dataset for training the model was the main challenge during model development. As a result, we created a preprocessing tool using the Indic NLP Library and inltk tools [1, 2], as well as collecting Marathi news from various online newspapers/channels and manually annotating and validating it with domain

experts. A comparison of all classification algorithms was also evaluated and discussed. In addition, an annotated dataset of Marathi news with sentiment orientations of positive, negative and neutral was developed [3, 4].

2 Related Work

Sentiment analysis (SA) and Opinion Mining are become popular in the field of natural language processing for Indian languages such as Hindi, Bengali, Tamil, Telugu and so on, but there are still some languages that remain unexplored due to a lack of lexical resources, such as Marathi, Gujarati, and Punjabi [5]. Three types of sentiment classification techniques are machine learning approaches, lexicon-based approaches and hybrid approaches. Machine learning approaches are classified into three types: supervised learning, semi-supervised and unsupervised learning. When there is annotated dataset unavailable to train the model then unsupervised learning is used, whereas supervised learning is used when there is a substantial annotated dataset available. The authors proposed a sentiment analysis (SA) method for Gujarati tweets that used POS tagging to extract features and SVM to classify the tweets. They have collected 40 tweets as dataset [6].

For performing sentiment analysis of Malayalam movie reviews, the authors devised a hybrid approach based on fuzzy logic. This consists of a tagging machine learning system and a fuzzy logic approach to determining review membership. TnT Tagger was also used to train datasets that had been manually tagged [7]. For Tamil and Bengali tweets, the authors developed machine learning-based sentiment analysis model based on probabilistic and decision tree classification. The amount of the dataset used, variations in writing style, and the incorrect use of punctuation marks all have an impact on the model's performance [8]. Using machine learning approaches such as support vector machine, maximum entropy, decision tree and Naïve Bayes, the author developed a system for predicting emotion in Tamil films. When it comes to accuracy, SVM outperforms other algorithms [9].

3 Proposed Methodology

This section describes the corpus creation process in detail, beginning with data collection, preprocessing, corpus annotation and inter-annotator agreement for measurement of agreement between annotators. And also, this section covers sentiment classification approach.

3.1 Lexical Corpus Creation

Web Scraping
We created a python-based web scrapper to collect Marathi news from various online newspapers/channels and collected 1649 news headlines from categories such as general news, current affairs, sports, entertainment, art, culture, science and technology, health and medicine and so on.

Data Preprocessing
We pre-processed the crawled data into the desired forms, following steps are carried out in preprocessing:

- Manually identify and remove duplicate and irrelevant news.
- Identify news items that contains English words and transliterated them.
- Remove improper punctuation marks, smileys, hashtags and photo tags.
- Remove complex sentences because they are unsuitable for sentiment analysis.

Data Annotation and Inter-annotator agreement
For manual data annotation, we chose three Marathi language domain experts who are academicians and researchers to annotate a Marathi news dataset with polarity scores −1, 0 and 1 (Negative, Neutral and Positive, respectively). For evaluation of manual data annotation, we have used the Fleiss' Kappa inter-annotator agreement score. Following formula is used to calculate Fleiss' kappa score [10].

$$k = \frac{\overline{P}_x - \overline{P}_x}{1 - \overline{P}_x}. \tag{1}$$

where the factor $1 - \overline{P}_x$. represents the degree of agreement that can be obtained other than by chance, The degree of agreement that was achieved above chance is given by $\overline{P}_x - \overline{P}_x$. and if the evaluators are totally in agreement, Kappa $k = 1$ and $k = 0$ if there is no agreement amongst the evaluators (other than what would expected by chance).

And the inter-annotator agreement score for Marathi news dataset is $k = 0.957$, which is almost perfect agreement. Above Table 1. Fleiss's Kappa Inter-annotator agreement score. And following Table 2 shows examples of data annotation. Table 3 shows the statistics for Marathi news dataset after preprocessing and data annotation.

Table 1 Fleiss's Kappa inter-annotator agreement score

Notator (i with j)	K-score
R_{12}	0.953
R_{23}	0.965
R_{13}	0.954
Fleiss's Kappa score (R_{123})	**0.957**

Table 2 Shows examples of data annotation

News	Polarity
पकिअप पलटी होऊन भीषण अपघात, ११ जणांचा मृत्यू ७ जखमी	−1
ऐतहिासकि अर्थसंकल्प, स्क्रॅप पॉलसिीमुळे देशात ५० हजार नवे जॉब	1
वॉर्ड आरक्षणासंदर्भात औरंगाबाद महापालकिने शपथपत्र दल्लिीला पाठवलि	0

Table 3 Shows the statistics for Marathi news dataset after preprocessing and data annotation

S. No.	Statistics	No. of news
1	Initial	1649
2	After preprocessing	1321
3	Positive	538
4	Negative	536
5	Neutral	237

3.2 Feature Extraction

To train the model, supervised machine learning algorithms require a text document in the form of a feature vector. Feature extraction techniques reduce the length of the feature vector by transforming all of the features in a lower-dimensional feature vector. Unigram features are bag-of-words (BoW) features obtained by removing unnecessary spaces and noisy characters between two words.

3.3 Sentiment Classification

Experiment work is based on supervised machine learning algorithms such as logistic regression, Stochastic Gradient Decent (SGD), support vector machine (SVM), nearest neighbour, neural network, decision tree (DT), Naive Bayes (NB), and the proposed ensemble-based model.

Individual algorithms compose an ensemble-based sentiment analysis in order to develop a highly accurate predictive model that classifies Marathi news in terms of sentiment orientation. The ensemble classifier system uses the average predicted probability, which is a soft voting approach, to determine the sentiment orientation. Figure 1 shows sentiment classification approach for Marathi news.

Fig. 1 Shows sentiment classification approach

4 Experimental Evaluation

In the experiment, we focused on the three types of class problems: positive, neutral and negative. We collected Marathi news headlines from a variety of newspaper and channel websites. In addition, the Marathi news dataset is divided into three categories based on the sentiment expressed in the sentences. If the expressed opinion is positive, it is labelled as 1, neutral, 0, and negative, it is labelled as −1. For training and testing samples, the dataset is divided into 80:20 ratios and various preprocessing techniques, such as data cleaning, URL and Hashtag removal, extra blank spaces, punctuation mark removal, emoticons, transliteration of English words into Marathi, Stopword removal and lemmatization, are used on the dataset. In experimentation validation, we used k-fold cross validation with $k = 5$ and $k = 10$. A classification model can be evaluated using a variety of measures, with accuracy being one of the most straightforward. The number of correctly classified examples divided by the total number of examples is the definition of accuracy. Accuracy is useful, but it ignores the complexity of class imbalances and the different costs of false negatives and false positives. Following Table 4. Shows performance evaluation of individual classifier with k-fold validation.

We performed fivefold cross validation on training dataset, and obtained highest accuracy for ensemble classifier as 95.19%, and also performed tenfold cross

Table 4 Shows performance evaluation of individual classifier with k-fold validation

S. No.	Classifier	$k = 5$		$k = 10$	
		Accuracy	F-score	Accuracy	F-score
1	Logistic regression	86.15	92.53	86.72	92.84
2	Stochastic Gradient Decent	94.12	96.84	94.44	95.32
3	SVM	88.78	93.87	90.95	95.01
4	Nearest neighbour	92.56	95.81	93.01	96.02
5	Neural network	94.16	97.02	95.07	96.93
6	Decision tree	93.12	94.16	94.21	95.16
7	Naïve Byes	87.30	93.11	88.44	93.71
8	Ensemble classifier	95.19	97.15	96.33	97.92

validation on training dataset, and we obtained highest accuracy for ensemble classifier as 96.33% for all Marathi news dataset. Figures 2 and 3 shows performance evaluation of individual classifier with fivefold and tenfold validation.

Fig. 2 Shows performance evaluation of individual classifier with fivefold cross validation

Fig. 3 Shows performance evaluation of individual classifier with tenfold cross validation

5 Conclusions and Future Work

This paper presents a baseline model for sentiment analysis in Marathi. We created an annotated dataset of Marathi news scraped from various newspaper/channel websites, and domain experts annotated Marathi news with positive, negative and neutral polarity. For sentiment analysis, we used machine learning models such as logistic regression, Stochastic Gradient Decent (SGD), support vector machine (SVM), nearest neighbour, neural network, decision tree (DT), Naive Bayes (NB) and the proposed ensemble-based model. The proposed ensemble classifier outperforms other classifiers in experiments, with an accuracy of 94.16% and an F-score of 97.02% for fivefold validation. In addition, the accuracy for tenfold validation is 95.07%, and the F-score is 96.93% and this dataset will be made available publicly for advancement in research.

In the future, we can create more domain-specific annotated lexical resources to expand resources for Indian languages, increase dataset size and implement using a deep learning-based model.

Acknowledgements Authors acknowledges the Chh. Shahu Maharaj National Research Fellowship (CSMNRF-2019), Pune, Maharashtra.

References

1. G. Arora, iNLTK: natural language toolkit for Indic languages, in *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, 2020
2. A. Kunchukuttan, *The IndicNLP Library* (2020)
3. S. Badugu, Telugu movie review sentiment analysis using natural language processing approach, in *Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing*, 2020
4. A. Rajan, A. Salgaonkar, Sentiment analysis for Konkani Language: Konkani Poetry, a case study, in *ICT Systems and Sustainability. Advances in Intelligent Systems and Computing*, 2020
5. S. Rani, P. Kumar, A journey of Indian languages over sentiment analysis: a systematic review, pp. 1415–1462 (2018)
6. V.C. Joshi, V.M. Vekariya, An approach to sentiment analysis on Gujarati Tweets. Adv. Comput. Sci. Technol. **10**(5),1487–1493 (2017)
7. M. Anagha, R.R. Kumar, K. Sreetha, P.C. Reghu Raj, Fuzzy logic based hybrid approach for sentiment analysis of Malayalam movie reviews, in *IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, Kozhikode, 2015
8. S. Se, R. Vinayakumar, M. Anand Kumar, K.P. Soman, Predicting the sentimental reviews in tamil movie using machine learning algorithms. Indian J. Sci. Technol. **9**(45), 1–5 (2016)
9. S.S. Prasad, J. Kumar, D.K. Prabhakar, S. Tripathi, Sentiment mining: an approach for Bengali and Tamil tweets, in *2016 Ninth International Conference on Contemporary Computing (IC3*, Noida, 2016
10. "Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Fleiss%27_kappa. [Accessed May 2021]

11. M.G. Jhanwar, A. Das, An ensemble model for sentiment analysis of Hindi-English code-mixed data, in *Workshop on Humanizing AI (HAI)*, Stockholm, Sweden, 2018
12. M.A. Ansari, S. Govilkar, Sentiment analysis of transliterated Hindi and Marathi Script, in *Sixth International Conference on Computational Intelligence and Information*, Cochin, India

Artificial Neural Network Method for Appraising the Nephrotic Disease

K. Padmavathi, A. V. Senthilkumar, Ismail Bin Musirin, and Bınod Kumar

Abstract In the shape of an affected person's proof, the medical report is an ever-developing supply of record for a medical institution. One of the complex issues that get up in the transplanted kidneys is glomerulonephritis. In AI, there are two methodologies: managed and solo mastering. Characterization is a method that falls underneath controlled learning. Out of numerous arrangement models, the maximum prevalently applied is the artificial neural community. While neural networks turn out tremendous in characterization and preparing a device, the precision of the outcome may also anyways be beneath inquiry. The enhancement of the artificial neural networks is completed by using the exactness and space of the result. For this, ANN may be hybridized with a metaheuristic algorithm referred to as the cat swarm optimization (CSO) set of rules. The benefits of optimization artificial neural community are normally the development in the precision of the order, translation of the statistics, and reduction in fee and time utilization for buying real outcomes and so forth within the prevailing study, a correlation between the aftereffects of an ANN decrease again propagation version and the proposed ANN-CSO version is carried out for medical assessment.

Keywords Machine learning · ANN · CSO algorithm

1 Introduction

The measure of information is developing dramatically and it is important to break down this immense measure of information to separate helpful data from it. This

K. Padmavathi (✉) · A. V. Senthilkumar
Computer Applications, Hindusthan College of Arts and Science, Coimbatore, India
e-mail: padhukarups3@gmail.com

I. Bin Musirin
Faculty of Electrical Engineering, Centre of Electrical Power Engineering, Universiti Teknologi, Skudai, Malaysia

B. Kumar
Department of MCA, Rajarshi Shahu College of Engineering, Pune, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_20

prompted the development field of information mining. Information mining alludes to removing information from such enormous measures of data sets. Information mining is the center of knowledge discovery data measures. Knowledge discovery data is a coordinated cycle of the recognized, substantial, novel, valuable and justifiable examples from huge and complex datasets [1]. Data mining assignments may be partitioned into two instructions: expressive and predictive. Prescient undertaking incorporate grouping, relapse, time arrangement examination and elucidating task incorporate bunching, affiliation rule [2]. These procedures can be utilized in explicit zones [3]. Discuss this process and their utility in a different area. Statistics mining applications contain special discipline, as an example, deals, medicine, account, advertising, clinical services, banking and safety [4, 5]. Order is information—digging method utilized for the expectation of a class of items. It is an illustration of directed learning. Grouping anticipates clear cut marks (discrete, requested). Information arrangement includes two stages. The preliminary step is the mastering step (getting ready step) wherein a classifier is labored to depict a foreordained association of data training. Within the 2nd step the model that's the inherent preliminary step is utilized for the characterization of obscure records as an example test records is applied for assessing the classifier exactness. There are various characterization calculations like preference tree, gullible classifier and artificial neural network gives the relative research of these order calculations [6, 7]. Synthetic neural community is a machine gaining knowledge of approach utilized for classification troubles.

2 Related Work

Various analysts executed different strategies to distinguish CKD and anticipate the endurance of patients. Various strategies, execution apparatuses and techniques give various outcomes and give alternate point of view on this specific sickness [8].

In [6], the creators thought the precision and execution time is based on the performance of the ocular networks. In light of the examination results, ocular network performs better than the other techniques.

In [7], artificial neural network (ANN) is executed with the back-propagation network, summed up generalized forward neural networks (GRNN) and modular neural network (MNN) (BPN) for early recognition of CKD. In light of the exhibition as far as precision, affectability and explicitness, the best model is picked to create of the smart framework. The framework is lined up with Google cloud stage by utilizing Google Application Engine. Subsequently, CKD can be recognized in the beginning phase all the more effectively. It is likewise valuable for general individuals to self-identify the danger of contact to CKD.

In [9], three information mining procedures have been analyzed for kidney sickness examination. These strategies are artificial neural network, decision tree and logical regression. From the computational outcome it is seen that ANN performs better and gives higher exactness.

An examination between Back spread fake neural network, PSO neural network and a Bat calculation streamlined neural network was made by Golmaryami et al. [10] on stock value forecast. The outcomes indicated that bat calculation changed the weight lattice more precisely than the other two utilized. The bat calculation was assessed alongside its the present variations and was introduced by Induja et al. [11]. Certain contextual investigations including bat calculation alongside its different applications were generally chosen and surveyed. The surveys and synopses dependent on these investigations have been quickly advanced in [11].

3 Optimization-Based Artificial Neural Network Classification

This segment depicts the fundamental commitment of this whole work. The subsection talks about the means remembered for the proposed mixture highlight choice strategy. As a rule, by decreasing the quantity of highlights, it will lessen the preparation time to assemble and limit the unpredictability of the classifier model. Highlight choice is to choose the most significant highlights before it adjusted as a contribution to the classifier to get upgrade arrangement. Thus, the less pertinent or less critical highlights will be dispensed with or taken out from the rundown. Half and half highlights determination technique utilizing highlight positioning procedure specifically alleviation f with ground-breaking transformative calculation improvement calculation is proposed. The detail clarification of our proposed strategy is depicted in the tracing after subsection [12].

3.1 Cat Swarm Optimization

The optimization technique unites two modes, looking for search mode and the tracing mode. Taking modes into the count, a mixture ratio of consolidating seeking for mode with tracing mode. While they might be resting, they move their position purposely and steadily, every now and then even live inside the main job. No matter what, for utilizing this lead into CSO, we utilize searching for mode to adapt to it.

The direct of looking for after focal points of cat is associated with tracing mode. for this reason, it is miles clear that MR have to be extremely a motivation with a selected final goal to guarantee that the cats make a contribution most of the strength in search of for mode, tons equal to this present day truth. The technique of CSO can be created into the following steps.

Step 1: Based on the identical time create N cats.

Step 2: Arbitrarily sprinkle the number of cats into the M-dimensional arrangement space and erratically pick qualities, which are in-degree of the most incredible speed, to the speeds of each cats. By utilizing then aimlessly

choose amount of felines and set them into following mode as demonstrated by means of MR, and the others set into looking for mode.

Step 3: Assess the well-being appraisal of each feline through utilizing the spots of felines into the wellness work, which tends to the proportions of our goal, and keep up with the quality feline into memory. Know that we most certainly need to think about the state of the quality feline (x best) because of it tends to the top notch plan as of as of now.

Step 4: Stream the felines as demonstrated by their flags, if feline k is looking for mode, apply the feline to the looking for mode measure, via and enormous use it on the following mode measure. The procedure steps are shown ahead.

Step 5: Again the select amount of felines and set them into following mode as affirmed through MR, with the guide of then set trade felines into searching for mode.

Step 6: test the give up circumstance, at whatever factor satisfied, prevent this machine, and for the maximum factor reiterate step 3 to step 5.

3.2 Artificial Neural Network

A network is prepared for appearing conceptual acknowledgment in la small amount of time. A ocular community abuses the non-linearity of an trouble to symbolize a group of wanted statistics sources. Ocular networks were massive in expertise a superior manner for grouping in AI and discovers utility in one of a kind fields, as an instance, statistics mining, design acknowledgment, crime scene investigation and so on.

An artificial neural network structure comprises of essentially 3 parts, they are input layer, hidden layer and output layer. There are the accompanying kinds of artificial neural network. Least difficult neural network structures is accompanied by feed—forward. It contains of a data film in which the records enters and is inferred out on the grounds that the yield from the yield film. The presence of shrouded layers is discretionary. It uses an incitation work for affiliation. Lower back-multiplication neural organization. A returned unfurl ocular local area is indistinguishable as the feed forward neural organization beside one little change.

4 Proposed Method

Proposed include determination method in CKD dataset. The cycle of proposed highlight determination methods in CKD dataset. In this cycle, we have utilized two covering techniques in particular Genetic Search and Greedy stepwise [13, 14]. In Genetic hunt approach, overlaying subset evaluator with development primarily

based is applied to select the vast 18 highlights from CKD informational collection. Moreover, if there should be an prevalence of keen stepwise, covering subset evaluator with enhancement-based classifier is applied to pick out the good sized 15 highlights from CKD informational series.

5 Result and Discussion

UCI machine learning repository CKD dataset was used for this proposed approach. Attributes are displayed accordingly. Classification algorithms are applied to the dataset for performance testing and assessment training [15].

Table 1 presents the performance of the various classification techniques. The accuracy of the ANN peaks high (Fig. 1)

Table 2 shows the various levels of classification techniques using the Cat Swarm Optimization techniques.

Table 1 Classifiers KNN, SVM, ANN performing their classification levels during their performance

S. No.	Algorithm	Accuracy	Precision	Recall
1	KNN	89	86	87
2	SVM	92	90	91
3	ANN	94	92	94

Fig. 1 Show the classification comparison of KNN, SVM, ANN classification algorithm ranges

Table 2 Performance of the classification techniques of Cat Swarm optimization techniques levels of KNN, SVM, ANN

Algorithm	Accuracy	Precision	Recall
CSKNN	92	90	91
CSSVM	95	93	95
CSANN	98	96	97

Fig. 2 Shows the classification comparison of optimization-based classifier algorithms. The comparison between existing methods (CSKNN, CSSVM) and proposed (CSANN) method

6 Conclusion

This examination built up an adjusted form of (CSO) to improve seeking through capacity by amassing look in the region of the best arrangement sets. We at that point consolidated this with an ANN to create the CSO-ANN strategy for include determination and ANN boundary streamlining to improve arrangement exactness. Assessment utilizing UCI datasets exhibited that the CSO-ANN technique requires less time than CSO-ANN to acquire grouping aftereffects of prevalent exactness.

References

1. J. Han and M. Kamber, *Data mining: concepts and techniques*, 2nd ed. Amsterdam; Boston: San Francisco, CA: Elsevier; Morgan Kaufmann, 2006.
2. P. Sondwale, Overview of Predictive and Descriptive Data Mining Techniques. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **5**(4), 262–265 (2015)
3. B. Ramageri, “DATA MINING TECHNIQUES AND APPLICATIONS,” *Indian J. Comput. Sci. Eng.*, vol. 1, no. 4, pp. 301–305.
4. P. Chouhan and M. Tiwari, “Image Retrieval Using Data Mining and Image Processing Techniques,” *Image (IN)*, vol. 3, no. 12, 2015.
5. Simmi bagga and g. n. singh, “applications of data mining,” *Int. J. Sci. Emerg. Technol. Latest Trends*, pp. 19–23, 2012.

6. N. Padhy, The Survey of Data Mining Applications and Feature Scope. Int. J. Comput. Sci. Eng. Inf. Technol. **2**(3), 43–58 (2012)
7. S.S. Nikam, A comparative study of classification techniques in data mining algorithms. Orient J Comput Sci Technol **8**(1), 13–19 (2015)
8. ArturQ.B.da Silva, Taina V. Delaware Sandes-Freitas, JulianaB.Mansur,Jose Osmar Medicina - Pestana,and Gianna Mastroianni-Kirsztajn."Clinical Presentation, Outcomes, and Treatment of Membranous renal disorder once Transplantation", International Journal of Nephrology, 2018.
9. Dr. S. Vijayarani and Mr. S. Dhayanand, "KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS," International Journal of Computing and Business Research (IJCBR), vol. 6, no. 2, 2015.
10. Ms. Sonali Sonavane, Prof. A. Khade and Prof. V. B. Gaikwad, "Novel Approach for Localization of Indian Car Number Plate Recognition System using Support Vector Machine," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 8, pp. 179–183, 2013
11. P.S. Devi, C.H. Sowjanya, K.V.N. Sunitha, A Review of Supervised Learning Based Classification for Text to Speech System. International Journal of Application or Innovation in Engineering & Management (IJAIEM) **3**(6), 79–86 (2014)
12. A.V. Senthilkumar, K. Padmavathi An efficient Meta classifier technique for Membranous Nephropathy Kidney disease. ISSN:227-8616, 2020
13. R. K. Chiu, R.Y. Chen, S. A. Wang "Intelligent systems on the cloud for the early detection of chronic kidney disease," in International Conference on Machine Learning and Cybernetics, Xian, 2012(July).
14. K.R. Lakshmi, Y. Nagesh, M.V. Krishna, Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability. International Journal of Advances in Engineering & Technology (IJAET) **7**(1), 242–254 (2014)
15. A.V. Senthilkumar, K. Padmavathi, A proposed method for Prediction of Membranous nephropathy Disease. International Journal of Innovative Technology and Exploring Engineering", ISSN:22783075, Volume -8 Issue-10,August 2019

Artificial Neural Network and Math Behind It

Harshini Pothina and K. V. Nagaraja

Abstract Article is structured in such a way that the reader could easily understand from the roots of an artificial neuron to its applications in between gaining knowledge on how an artificial neuron exists. It starts with the history of neuron, and the explanation goes on by describing the architecture of artificial neuron and on to the functioning of neuron. It also explains the basic math involved in constructing an artificial neuron by describing the activation functions used in the network, and the mathematical models helps in building the neural network. It also talks about the structure of feed forward network, loss function imposing an error on the output, also the gradient descent, and the back propagation method involving the computation of cost function with respect to the network parameters. As a conclusion, paper describes the importance of artificial neural networks in the upcoming artificial intelligence and many more applications behind it.

Keywords Artificial neural network · Back propagation · Feed forward network · Gradient descent · Perceptron · Sigmoid neuron

1 History

Neural networks that we are going to consider are artificial neural networks, which are based on human brains structure and its function. The first step toward artificial neural networks came in the year 1943; Warren McCulloch and Walter Pitts released a journal on working of neurons so as to depict its capacity in human brain. They demonstrated a basic neural system utilizing electric circuits.

H. Pothina (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: BLENU4EIE18031@bl.students.amrita.edu

K. V. Nagaraja
Department of Mathematics, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: kv_nagaraja@blr.amrita.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_21

Later, in 1949, Donald Heb composed a book named "The Organisation of Behaviour," which brought up the way that the pathways of neurons are reinforced each time they are utilized, an idea in general by which humans learn. He contended that if two nerves fire simultaneously, the association between them is upgraded.

In 1962, Widrow and Hoff assembled a studying approach that inspects the yield before the weight alters it (that is 0 or 1) as per the standard: change in weight = (pre-weight line yield)*(error/number of input values). It depends on the possibility that the major error of one perceptron can be modified by distributing it over the system while other perceptrons modify their weights. Even after applying this standard, despite everything brings about an error, however, it will run right independently [1].

2 Introduction

Neural networks process the data as like the human brain does. Network is structured with a huge number of interrelated processors called neurons that work along to deal with certain problem. *What are artificial neural networks?* ANNs are popular machine learning techniques that simulate the mechanism of learning in biological organisms. The cells present in the human nervous system are referred to as neurons. Axons and dendrites are used to connect these neurons to one another using a connecting region called synapses. The change in the response of external stimuli depends upon the strength of the synaptic connections, and due to these changes, learning in living organisms takes place. This biological procedure is replicated the same way in artificial neural networks, which contain processing elements called neurons. These processing elements are joined to each other by means of weights, which play the similar part as the power of synaptic connections in biological structures [2–4]. Each input to a neuron is associated with a weight which acts on the function operated at that subdivision.

3 Artificial Neuron

See Fig. 1.

An ANN computes a function of the input neurons to the output neurons using an intermediate parameters called weights and biases. The data from the input layer is taken into the hidden layer and the information is seen at the output layer after an amazing process performed in the hidden area. It is within the hidden area where all the process really happens through a system of connections specified by "weights" and "biases" (generally referred as W and b). Implementation and optimization of a neural network lies in understanding what exactly happens in the hidden area. The neuron receives the input and computes the weighted sum and by adding the bias it decides whether neuron should be "fired" or "activated" based on the result and a pre-set "activation function." After then, the neuron sends the data to the next

Fig. 1 Artificial neuron [5]

Fig. 2 Different layers of a network [9]

attached neuron in a process called "forward pass." At the termination of this action, output layer with one neuron for every one feasible output is connected to the hidden layer [6–8] (Fig. 2).

4 Perceptron

We have a general thought of the architecture of a neural network. Now, let's continue our discussion with the explanation of working of a neural system. *How does a neural network run?* To understand its working, we need to know the different kinds

of neurons present in our neural systems. Perceptron is one of its kind. *What is perceptron and what can it do?* A basic artificial neuron with input and output layers is called perceptron. It is a mathematical model of a biological neuron which talks about weights and biases. In an actual neuron, the electrical signals are received by the dendrites, whereas in perceptron, these electrical signals are given by numerical values. In a biological neuron, electrical signals at the synapses between dendrites and axons are modulated in various amounts. It can be modulated in perceptron by multiplying each input value by weights and adding a bias to it. Bias is a constant value which helps in better performance of the neuron, and it is given a constant value 1 by default. A neuron discharges an output only if the strength of the input neuron exceeds a certain threshold. Total strength of the input signals can be modeled in a perceptron by computing its weighted sum and applying its step function to determine its output value. As in biological neural systems, this can be fed to the other perceptrons. Perceptron takes the following steps:

(i) Multiplying all the input samples with their weights w, (ii) Summation of the weighted sum of all the input samples: $\sum w_j\ x_j$, (iii) Apply the activation function, at the end decide if the weighted total is more than a threshold, where threshold is proportionate to bias and allocate 1 or 0 as an output.

We can likewise compose the perceptron work in the accompanying terms: $F(x) = \begin{cases} 0 \text{ if } w.X + b < 0 \\ 1 \text{ if } w.X + b \geq 0 \end{cases}$. Here, $w.X$ is the dot product of the vector with components of weight and input vector matrix, b is the bias and is proportional to the $-$threshold (bias $= -$threshold) [10].

5 Activation Function

The major purpose of the activation function is to transform the weighted sum of the inputs of every neuron at the input side to the output signal, and these output values are distributed as inputs to the next layer neurons. Any activation function that we consider must be differentiable in view of the fact that a back propagation technique is used to reduce the error and run the weights accordingly. One of the foremost advantages of the activation function [2] is that we can get various models of the output of our choice by varying the weights and bias. According to the mathematical equation of the perceptron that we have seen above, the output can be made more likely with the large positive value of bias, and a large negative value of bias makes the output more unlikely, and weights can be changed accordingly to get the desired output. A perceptron is capable of decision-making by analyzing the given data according to the training set. Complex networks with lot many layers of perceptrons can be made possible where the output from previous layer is taken by the next successive layer as input signal, and weights it up to yield. We moderately vary weights and bias to observe the performance of the neural network, but for a perceptron, small variations in the weights and bias lead output shift from 0 to 1 and

vice-versa. Due to this drawback of perceptron, network is replaced with many other activation functions like tanh, rectified linear unit (ReLU), and Sigmoid functions, also called logistic function. The present time activation function which is used more is sigmoid function due to the fact that output can be any continuous value between 0 and 1.

Sigmoid function: It is demonstrated by the following equation: $\sigma(x) = \frac{1}{(1+e^{-x})}$, and the output:

$F(x) = \sigma(wX + b)$ So, the formula of output becomes $\sigma = \frac{1}{\left(1+e^{-\sum w_j X_j + b}\right)}$.

Artificial neurons receive multiple inputs as like biological neurons receives from pre-synaptic neurons. These artificial neurons take all the inputs, adds them up, and process using sigmoid function only if the summation exceeds the threshold limit, and then, the processed data is served as output.

Derivatives of common activation functions:

1. **Sigmoid activation function:** $\sigma(x) = \frac{1}{(1+e^{-x})}$

Quotient rule is used to calculate its derivative: $\sigma'(x) = \frac{\partial\left(\frac{1}{1+e^{-x}}\right)}{\partial(x)} = \text{Sigmoid}(x)(1 - \text{sigmoid}(x))$.

If we have a careful look at its derivative, we can observe that the outcome of its derivative is simply $\text{sigmoid}(x)$ weighted by $1 - \text{sigmoid}(x)$. This is very convenient to compute effective gradients in a neural network [7, 11]. The output of the sigmoid function ranges between 0 and 1, but for very large negative or positive input sample, the output at the sigmoid function becomes saturated at the rear end of 0 or 1. Hence, the derivative of the function becomes less in value preventing the updates of the weight. Even though the sigmoid function has a very good interpretation, it can cause neural network to get stuck at their current state during training. Hyperbolic tangent function is the substitute to the sigmoid function.

2. **Hyperbolic Tangent function**: $\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{(e^x + e^{-x})}$

Like sigmoid function, tanh is also a sigmoidal function which is "*S*" shaped. The output of tanh function ranges from -1 to $+1$, thus the very large negative input maps to the negative [2] output, very large positive input maps to the positive output, and the zero valued input maps to the near zero output. This makes the function less likely to get stuck while training [12].

Quotient rule is used to calculate its derivative:

$$\tan h'(x) = \frac{\partial \tanh(x)}{\partial x} = 1 - (\tanh(x))^2$$

3. **Identity Activation Function:** Linear $(x) = x$

The output values from this function ranges from $-\infty$ to $+\infty$. It maps the pre-activation to itself. Linear activation function helps to formulate a multilayered neural

network with linear activation into a single-layered linear network. The derivative of linear activation function is unity in case of one-dimensional inputs.

Linear'$(x) = 1$, So after having a knowledge of how neurons work in a network, let us now know how such multiple of neurons can be brought together to form a feedforward neural network.

6 Structure of Feedforward Neural Network

Feedforward neural network consists of three different layers, namely input layer, hidden layer, and output layer. Data is first given at the input layer which is being sent to the hidden layer for further process. Hidden layer is present with actual neurons, and after a magic trick, output is shown at the output layer. The information flows only in one direction which is forward. Hence, it is known as feedforward neural network [13]. If the information is not passed in one direction and the output of neuron is feedback into previous layer in a cycle, then it is known as recurrent neural network called back propagated neural network. Let us now consider a three-layered network with sigmoid neurons in a training set. The following network has four input signals, two neurons at the hidden layer, and three at the output layer (Fig. 3).

We first have to assign weights and biases to the connections between neurons at each layer [6]. In general, weights and biases are all arranged randomly in a matrix form and vector form, respectively.

The total weighted sum of the parameters is given by: $Y = wX + b$

Fig. 3 Structure of feedforward neural network [14]

$$Y = \begin{bmatrix} w_{(1,1)} & w_{(1,2)} & w_{(1,3)} & \cdots & w_{(1,n)} \\ w_{(2,1)} & w_{(2,2)} & w_{(2,3)} & \cdots & w_{(2,n)} \\ w_{(3,1)} & w_{(3,2)} & w_{(3,3)} & \cdots & w_{(3,n)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{(m,1)} & w_{(m,2)} & w_{(m,3)} & \cdots & w_{(m,n)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

In a weight matrix, number of columns is given same as the number of neurons existing at the previous layer, and number of rows is given same as the number of neurons existing at the current layer. In a bias vector, the number of vector components is given same as the number of neurons existing at the present layer. Now, we need to build these different layers in a neural network by assigning weights and biases to the neurons starting with the feedforward step to predict the output [7]. First layer is the input layer which takes the input samples represented by the vector "X." This vector X is multiplied with the weighted matrix "w" and sums up with bias at the connections between the layer1 and layer2, and the result is passed to the second layer as input. Input to the layer1 is given by $X^{(4)}$ which represents four input samples from four neurons.

The input to the layer2 is given as: $Y_1 = w^{(2,4)} X^{(4)} + b^{(2)}$.

Here, $w^{(2,4)}$ represents weighted 2×4 matrix with two neurons at the second layer and four neurons at the first layer.

$b^{(2)}$ represents bias vector of two neurons at the 2nd layer.

In this layer, the weighted input is computed under sigmoid activation function to form an output, and it is weighted up with the weights and bias present at the connections between second and third layer. The final output is taken by the third layer as input. Output at layer2 is given as

$$F_1(x) = \sigma(Y_1)$$
$$F_1(x) = \sigma(w^{(2,4)} X^{(4)} + b^{(2)})$$
$$\mathrm{F}_1(\mathrm{x}) = \sigma\left(\mathrm{w}^{(2,4)}\mathrm{X}^{(4)} + \mathrm{b}^{(2)}\right)$$

The input to the third layer is given as

$$Y_2 = w^{(3,2)} F_1(x) + b^{(3)}$$
$$Y_2 = w^{(3,2)}(\sigma(w^{(2,4)} X^{(4)} + b^{(2)}) + b^{(3)})$$

Here, $w^{(3,2)}$ represents weighted 3×2 matrix with three neurons at the third layer and two neurons at the second layer.

$b^{(3)}$ represents the bias vector of three neurons at the third layer.

This input to the third layer to computed under sigmoid activation function to form the final result which is given as

$$F_2(x) = \sigma(Y_2)$$
$$F_2(x) = \sigma(w^{(3,2)}(\sigma(w^{(2,4)}X^{(4)} + b^{(2)}) + b^{(3)}))$$

In this concept of neural networks, weights and bias values are picked in such a way that the output value of the network is approximated to the desired value of the function $F(x)$ for all inputs in the training set. Things being what they are, how would we evaluate how far our expectation is from our desired output with the goal for us to know whether we have to continue scanning for more exact boundaries? For this point, we need to compute an error or such characterize a cost function also called as loss function [2, 7]. It is the deduction between the desired output and the actual output. The most ordinarily utilized one in our neural networks is the quadratic cost function, likewise called mean squared error, characterized by the equation:

$$C(w, b) = \frac{1}{N}\sum_{i=1}^{N}\left|F_{i\,(\text{desired})}(x) - F_{i\,(\text{actual})}(x)\right|^2$$

"w" refers to the weight, "b" refers to bias, "N" is the total number of input samples in the training set, "$F(x)$" is the output of the network, "i" refers to the sample number, and "$\sum$" refers to the summation over all the inputs in the training set.

In order to modify the variables, weights and biases to get the approximated output, mean squared error method is preferred over linear error because tiny modifications in weights and biases do not produce any change within the range of desired outputs; therefore, employing a quadratic function wherever massive variations have additional impact on the loss function than small ones helps in deciding the way to modify these variables. To minimize the error, we have a new algorithm called gradient descent. It is used to minimize the loss function by finding the new set of values of the input parameters at each layer as the loss function gets reduced whenever the output approaches the desired value of Y. Adjusting the weights by calculating the gradient of a loss function [11] helps in reducing the error of an output. In general, gradient is a measure of change in output value due to the change in input parameter values. This can be achieved by finding a global minimum of the loss function. Mathematically, gradient is nothing but the partial derivative of the function with respect to its input value (Fig. 4).

As we see from the above graph between cost function and weight, weight is changed to a certain value in a direction to the negative gradient to get the minimum value of a cost function. Now, we present a new parameter called "Learning rate" which has the greatness of explaining what number of units should we move from a current position to the heading that makes our work quicker. We have known how to calculate the error and the derivative of a loss function to update weights and biases. Now, we will move further to know how to compute the gradient descent in order to minimize the error [16–18]. Here is where a new method called "Back Propagation" arises. The word back propagation says that the gradient of the error at each layer is propagated back starting from the output side to the input side by making changes

Fig. 4 Graph of gradient descent with respect to weight [15]

in the weights and bias values for each layer in a network. The gradient at each layer with respect to all the parameters is computed according to the chain rule by updating weights and bias values in the direction of negative gradient toward its minimum point.

Now, let's consider the equation of loss function or cost function to compute the gradient. Gradient with respect to weights is

$$\begin{aligned}\frac{\partial C}{\partial w_{(L)}} &= \frac{\partial C}{\partial F_{L(\text{actual})}(x)} \frac{\partial F_{L(\text{actual})}(x)}{\partial Y_{(L)}} \frac{\partial Y_{(L)}}{\partial w_{(L)}} \\ &= 2\big(F_{\text{desired}}(x) - F_{L(\text{actual})}(x)\big)\sigma'(Y_{(L)})(F_{L-1(\text{actual})}(x)\end{aligned}$$

where "L" is the last layer and "$L-1$" is the previous layer or second last layer, and we require three main equations to back propagate the error gradient with respect to weight and bias at current layer (L).

$$\begin{aligned}\frac{\partial C}{\partial w_{(L)}} &= \frac{\partial C}{\partial F_{L(\text{actual})}(x)} \frac{\partial F_{L(\text{actual})}(x)}{\partial Y_{(L)}} \frac{\partial Y_{(L)}}{\partial w_{(L)}}, \\ \frac{\partial C}{\partial b_{(L)}} &= \frac{\partial C}{\partial F_{L(\text{actual})}(x)} \frac{\partial F_{L(\text{actual})}(x)}{\partial Y_{(L)}} \frac{\partial Y_{(L)}}{\partial b_{(L)}}.\end{aligned}$$

Gradient with respect to the activation function at previous layer or second-last layer ($L-1$)

$$\frac{\partial C}{\partial F_{L-1(\text{actual})}(x)} = \frac{\partial C}{\partial F_{L(\text{actual})}(x)} \frac{\partial F_{L(\text{actual})}(x)}{\partial Y_{(L)}} \frac{\partial Y_{(L)}}{\partial F_{L-1(\text{actual})}(x)}$$

The above gradient equations are to measure how a weight and bias parameters affect the cost function which we optimize. Now, learning rate comes into picture while updating weights and bias.

$w_{i\,(\text{new})} = w_{i\,(\text{old})} - \alpha\left(\frac{\partial C}{\partial w_i}\right)$, where $\frac{\partial C}{\partial w_i}$ is the gradient of loss function with respect to weight at a specific layer.

$w_{i\,(\text{new})}$ is the updated weight at a specific layer, $w_{i\,(\text{old})}$ is the old weight at a specific layer and "α" is the learning rate.

At the last layer or third layer in our neural network, refreshing the weights and biases relies on cost function, weights, and biases given to that layer, but when comes to hidden layer or second-last layer, updating weights and biases depends upon the computations involved in the last layer and weights and biases given to that second-last layer. These dependencies would add up if we have more hidden layers. Let's see how the dependency happens by considering the gradients with respect to weights and biases at third layer

$$\frac{\partial C}{\partial w_{(3)}} = \frac{\partial C}{\partial F_{3(\text{actual})}(x)} \frac{\partial F_{3(\text{actual})}(x)}{\partial Y_{(3)}} \frac{\partial Y_{(3)}}{\partial w_{(3)}},$$
$$\frac{\partial C}{\partial b_{(3)}} = \frac{\partial C}{\partial F_{3(\text{actual})}(x)} \frac{\partial F_{3(\text{actual})}(x)}{\partial Y_{(3)}} \frac{\partial Y_{(3)}}{\partial w_{(3)}}.$$

To compute the updates for weights and biases in the second-last layer, we would have to utilize some of the computations from last layer [7].

$$\frac{\partial C}{\partial w_{(2)}} = \frac{\partial C}{\partial F_{3(a)}(x)} \frac{\partial F_{3(a)}(x)}{\partial Y_{(3)}} \frac{\partial Y_{(3)}}{\partial F_{2(a)}(x)} \frac{\partial F_{2(a)}(x)}{\partial Y_{(2)}} \frac{\partial Y_{(2)}}{\partial w_{(2)}},$$
$$\frac{\partial C}{\partial b_{(2)}} = \frac{\partial C}{\partial F_{3(a)}(x)} \frac{\partial F_{3(a)}(x)}{\partial Y_{(3)}} \frac{\partial Y_{(3)}}{\partial F_{2(a)}(x)} \frac{\partial F_{2(a)}(x)}{\partial Y_{(2)}} \frac{\partial Y_{(2)}}{\partial b_{(2)}}.$$

$\frac{\partial C}{\partial F_{3(a)}(x)}$, $\frac{\partial F_{3(a)}(x)}{\partial Y_{(3)}}$ are the re-utilized partial differential terms from last layer or third layer.

7 Applications of Artificial Neural Network

1. **Neural Network approach to an Autonomous Car Driver***:* In this scenario, neural network is trained under Gazelle system to determine the movement of the car. Network is trained with back propagation algorithm and uses five neurons to assign five different input values at input side, hidden layer, and output layer. Hyperbolic tangent activation function is used to train the neuron. The output yields the target angle for steering the car. Five input values are given by target angle, current angle, lateral offset of the car, and the distances which

are the difference between the distance ahead in the direction of movement and the distance ahead at a certain angle [19].

2. **Neural Networks for modeling Brain Wave Data:** Wisconsin Card Sorting test is used to record the signal waves produced by the brain. It uses feed forward neural network and Elman recurrent neural network to analyze the accurate brain wave forecasts with the help of linear activation function and logistic activation function. The signal waves are generated by varying number of input neurons and number of nodes at the hidden layer, and the best architecture of the neural network is established by the root mean square error (RMSE) and mean absolute percentage error (MAPE) [20].
3. **Artificial Neural Networks-based controllers for biped Robots:** Biped robots are the ones which have two legs, same like humans. Different kinds of artificial neural network controllers are used to control and for the stability of the robot. Recurrent neural network (RNN), self-recurrent wavelet neural network (SRWNN), and Recurrent wavelet Elman neural network (RWENN) are the three main ANN-based controllers [8] used to stabilize and improve the response of the robot. These controllers work the same like a neural network with input, hidden layer, and output layer neurons with a transfer function. These controllers are compared based on the result by computing root mean square error (RMSE) [21].
4. **Artificial Neural Network for predicting Crop Yield***:* Earlier to understand the weather condition toward the crop yield, crop modeling was used as a tool which is based on linear methods. These methods were neither enough to know the interconnections between these external factors and crop yield. To get the better of these linear methods, nonlinear methods like neural networks, fuzzy logic, and many more are used. Application of mathematical models and artificial neural network approach are used to predict the crop response as they give accurate outputs with minute error. Multilayered feed forward neural network is chosen to know the interactions between the soil layer and the crop yield. Based on the crop, the input parameters given to the network changes like planting date, minimum and maximum temperature conditions at which crop grows, maturity group etc., to leave leaf area index, mass and temperature stress factor, relative growth rate etc., as outputs [22].
5. **Neural Networks to estimate participation in elections:** Artificial neural networks are widely used to anticipate polity issues with a best accuracy. To select the members in an election, multilayered neural network and tan sigmoid activation function are used to predict the selected people in elections. Data such as qualification, age, opinion about the government services, and political orientation of people are given as inputs to the network with hidden layers and gives the target members as output of the network. Weights and biases are chosen as number and percentages at each stage of the network. Best accuracy of the data is finalized based on the error graphs left by the network [23].
6. **Neural Networks in Electric Power Industry:** Optimization techniques are very much used to deal with problems related to power systems, controlling, and many such complex parameters. Economic load dispatch, load forecasting,

and security analysis are the three main constraints in electric power industry, which requires mathematical optimization technique methods to solve such complex and large problems. One such optimization tool to dispatch required number of units to the load of the system, to estimate the load and to supply the power with required number of volts and frequency limits is artificial neural network. Multilayered perceptron with back propagation algorithm is used by giving security indicators such as fuel mix diversity and energy intensity as inputs to the network [24].

7. **Neural Network to predict the yield of enzymatic synthesis of betulinic acid ester:** Synthesis of betulinic acid and phthalic anhydride gives a $3\beta - O -$ phthalic ester which acts as an anticancer agent. Esterification is carried out using feedforward neural network with sigmoid function as transfer function and gradient descent algorithm. Amount of ester, reaction time, reaction temperature, and number of moles of the substrate are given as data at the input neurons to the network, and the number of neurons at the hidden layer depends upon the complexity of the network. The output of the neuron gives the percentage of yield ester. Weights and biases values are assumed accordingly, and gradient descent back propagation technique is used to predict the yield by training the network [25].
8. **Neural Networks for Face Recognition:** Face recognition is a kind of pattern recognition problem which has face detection, face alignment, face extraction, and face matching. It uses three-layered feedforward network to detect the human faces and involves multilayered perceptron for face alignment and other models of artificial neural networks combined with geometric feature-based methods for face extraction and face matching. Length of the extracted image vector, that is, the number of pixel boxes of the image becomes the number of neurons at the input side, whereas number of neurons at the hidden layer depends upon the complexity of the network and the output neuron as one or many depend on the kind of task to be solved. Length of the pixel of the image is given at the input neurons, and the artificial neural network filters out the false negative images and yields out the true image in the form of 0 or 1 if it contains human face [26].
9. **Neural Networks in geotechnical engineering:** Artificial neural networks in this field are used to estimate the properties of soil and to predict the liquefaction which causes the ground failure during earthquakes. ANNs are also used to retain walls, mining, blasts, and dam. Many other mathematical models fail to encounter these complex problems in geotechnical areas, but ANN helps in solving those by providing an input and output architecture with appropriate yield as the desired output can be obtained by changing the input parameters using back propagation algorithm. Multilayered feedforward network with logistic activation function is used, and number of hidden layers is found by hit and miss method which depends on the complexity of the network, but, however, scientists had found that two hidden layers are sufficient to model the algorithm to obtain liquefaction potential and other such related parameters as output [27].

10. **Neural Networks in HIV Modeling:** Artificial neural networks are used for HIV modeling which includes stages like prediction, function approximation, and classification of the pattern. Multilayered perceptron is used for prediction and function approximation of HIV which takes HIV drug resistance mutations and treatment change episodes (TCEs) as input variables and for the prediction of HIV protease cleavage sites in proteins. It also uses feedforward neural network along with hyperbolic tangent and sigmoid activation functions. ANNs are also used in prediction of creatinine clearance in patients with input variables as their age, actual and ideal body weight, body surface area, and some blood chemistries like RBC count and WBC count, platelet count, albumin, blood urea nitrogen, and many more such variables. Multilayered perceptrons are also used for the classification of HIV patients into positive and negative result, which takes input data as their age, weight, CD4, CD8, gender, HB, and TB and also used to classify the health and symptomatic status, visual fields (VFs) between eyes of a normal person and HIV person [28].
11. **Neural Networks for Industrial drying:** Artificial neural networks are used to increase the quality in the drying process of grated coconut by controlling the product humidity level. Multilayered perceptron with back propagation method and hyperbolic tangent function are used to maintain the required humidity level. Number of neurons at the input and output layer depends upon the data given and the yield. Number of hidden layer neurons depends upon the selected architecture of the problem with less error. The seven temperatures of the product dryer, final temperature of the product, and the initial moistures of the crushed coconut and pasteurized grated coconut are considered as input data which yields the final moisture of the grated coconut [29].
12. **Neural Networks for analysis of path planning of Humanoid Robots:** Artificial neural networks are used for the path making of humanoids which reach the target place by avoiding the obstacles on their path. Feedforward neural network with an input, two hidden and output layers are considered, and a hyperbolic tangent activation function is used along with back propagation algorithm. The network takes heading angle, distance to the right, left, and front obstacles are considered as input data and gives the steering angle at the output layer. Simulation process using ANN and experimental process of a humanoid are compared to observe the errors in the length of the path and the time taken for the robot to reach the final point by avoiding the obstacles. Simulation process using ANNs shows less error compared to the experimental process. ANN works efficiently in the path planning of humanoid robots in a cluttered environment [30].
13. **Neural Networks to early Lung Cancer detection:** Artificial neural networks are used efficiently to predict the patients with and without lung cancer. Multilayered perceptron with gradient descent algorithm and other two such algorithms for error measurement are used. Hyperbolic tangent activation function and sigmoid activation function are used to train the network. Variables like age, weight, gender, pneumonia, cough, bone pains, lung cancer in the family,

WBC and RBC count, fever, cigarettes, hemoglobin, platelets, Na, K, vital capacity, creatinine, cardiovascular diseases, appetite etc. Such kind of 48 parameters count of the patients are taken as the input data and gives an output value as 0 or 1, indicating with and without lung cancer of the patient. The final architecture of the network is with 48 input neurons, 9 hidden layers, and 2 output neurons are considered based on the experiment [31, 32].

14. **Neural Networks for Medical Diagnosis:** Artificial neural networks are used in diagnosis of diseases like acute nephritis disease and heart disease. Feedforward neural network along with back propagation method is used to differentiate the infected and non-infected people. Disease symptoms, images, and signals related to the test part are used as input data; hidden layer consists of 20 neurons, and output layer has two neurons which shows 0 or 1 identifies as non-infected or infected person. Sigmoid activation function is used to process the data, and the performance of the network is increased by the back propagation method and the calculation of loss function (mean square error of the samples) [33].
15. **Neural Networks to simulate performance of Osmosis membrane:** Artificial neural networks are used in desalination process of seawater using reverse osmosis membrane method. It uses feedforward network with sigmoid activation function and back propagation algorithm to calculate the error. Desalination process is done for different sea water reverse osmosis membranes (SWRM) which calculates the fresh flow concentration and total suspended solids rate (TDS). Number of input layer and hidden layer neurons depends upon the membrane that we are selecting. Parameters like temperature (T), influent flow (q), percentage of recovery (Re), and saline water with influent concentration (So) are given at the input nodes to yield fresh flow concentration (Se) [34].
16. **Neural Networks to predict moisture in Halite:** Artificial neural networks are used to foresee the presence of water on the exterior of halite rocks. Multilayered perceptron with back propagation algorithm and sigmoid activation function are used to direct the neural network with the given input data. Presence or absence of water on the halite rocks is found by the yield of electrical conductivity of the surface of a halite rock using conductivity sensors. Relative humidity and temperature of the air are given as input parameters to yield electrical conductivity. Number of neurons at the hidden layer is estimated by trial and error method based on the complexity of the algorithm, weights and biases are assigned randomly. Error between the actual and desired value is calculated using mean square error method at the output layer to further proceed for the increments in weights to get the desired yield [35].
17. **Neural Networks for inverse design and simulation of Nano-photonic particle:** Artificial neural networks are used to approximate the simulation of problems on light scattered by the nanoparticle and to solve problems on inverse design in physics like quantum theory, photonic devices, and photovoltaic materials. Inverse design problems are related to the geometry of a spectrum closest to any arbitrary spectrum which was chosen. Lossless silica shells

and lossless TiO_2 were considered as nanoparticles, and the neural network is trained with 250 neurons at each layer. Thickness of the nanomaterial shells was given as the input data, and the output was the spectrum which is sampled at a certain range of wavelength. Training of network is made by changing the learning rate and number of neurons at each layer. Neural network approach was compared with the simulation model and observed that it works well in approximating nanoparticle scattering phenomena. Neural network produces an accurate geometry of an arbitrary spectrum for inverse design problems. Inputs to the network were given by the trainable variables with fixed weights, and the network is trained with the back propagation algorithm by fixing the output node with the desired output to propose the geometry of the spectrum [36].

18. **Finite element method based Neural Networks:** Finite element method is generally used to tackle mathematical problems related to optimization techniques, partial differential equations, fluid flow, heat analysis, electromagnetism, and many more such problems from different branches in engineering. This method serves as a tool in dealing complex algorithms like solving partial differential equations in back propagation algorithm in ANN. Finite element analysis is performed at the beginning for the parameters required for a network construction. These results are passed through the network layers to proceed with further actions. FEM is a powerful tool to deal with boundary values in a neural network by updating the necessary parameters like weights, inputs etc. This results in accurate output in less time with minimal error [37–41].

8 Conclusion

Artificial neural networks have a wide range of applications in artificial intelligence, cloud computing, cyber security, and many more areas. ANNs are widely used for face recognition, image processing, language processing, text classification, and so on. Nowadays, handling problems in medical science, financial analysis, agriculture, education, banking systems, marketing, and many more has become an easy task for neural networks. Applications of ANN have a great impact in the current trend because of its analysis factors like processing speed, tolerance, performance etc. ANNs take part in control systems and robotics. Mathematics is the key step behind all these applications to structure a neural network. Mathematical model provides the means to implement the application and to reach the goal. Some of the topics from mathematics like calculus, linear algebra, probability, statistics, vectors, matrices etc., play a dominant role to architect any type of neural network. This article on artificial neural networks helps the reader to get a basic idea of artificial neuron. It draws the reader's attention by giving the knowledge of math required to build a neural network and makes their task easy in solving such algorithms. This paper can also help the reader in implementing the network algorithm in a Python code, and finally, it can summarize its applications in different areas.

References

1. History of Neural Networks, Stanford computer science
2. C.C. Aggarwal, Neural Networks and Deep learning, IBM T. J. Watson Research Center International Business Machines Yorktown Heights, NY, USA
3. R. Singh, Basics of Artificial Neural Network: International Journal of Engineering Research and Technology (IJERT), Department of Mechanical Engineering, Ganga Institute of Technology and Management, Kablana, Jhajjar, Haryana, India
4. T.J. Griinke, Development of an artificial neuralnetwork(ANN) for predicting tribiological properties of KENAF FIBRE REINFORCED EPOXY COMPOSITES (KFRE)
5. A. Choudhary, D. Pandey, S. Bhardwaj, Artificial neural network based solar radiation estimation: a case study of Indian cities. Int. J. Emerg. Technol. **11**(4), 257–262 (2020)
6. S. Haykin, Neural Networks, A Comprehensive Foundation", 2nd edn, Mc Master University, Hamilton, Ontario, Canada, Published by Pearson Education (Singapore) Pte. Ltd., Indian Branch, 482 F.I.E. Patparganj, Delhi 110092, India
7. C.F. Higham, D.J. Higham, Deep Learning: An Introduction for Applied Mathematicians, January 19, 2018
8. O.I. Abiodun A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State of the art in artificial neural network applications
9. M. Devlin, B.P. Hayes, Non-intrusive load monitoring using electricity smart meter data: a deep learning approach, November 2018. https://doi.org/10.13140/RG.2.2.29463.42402
10. M. Nielsen, Neural Networks and Deep learning
11. S.-H. Han, K.W. Kim, S.Y. Kim, Y.C. Youn, Artificial Neural Network: Understanding the basic concepts without mathematics: Dementia and Neurocognitive disorder(DND)
12. A.S. Jacobs, R.J. Pfitscher, R.L.D. Santos, M.F. Franco, E.J. Scheid, L.Z. Granville, Artificial neural network model to predict affinity for virtual network functions, in *NOMS 2018—2018 IEEE/IFIP Network Operations and Management Symposium*, Taipei, 2018, pp. 1–9. https://doi.org/10.1109/NOMS.2018.8406253
13. M.H. Sazli, *A Brief Review Of Feed Forward Neural Networks* (Ankara University, Faculty of Engineering, Department of Electronics Engineering 06100 Tandoğan, Ankara, Turkey)
14. http://www.geocomputation.org/1998/05/gc05_01.gif
15. https://www.oreilly.com/library/view/learn-arcore-/9781788830409/assets/f3899ca3-835e-4d3e-8e7f-fd1c5a9044fb.png
16. L. Fridman, Recurrent Neural Networks for Steering through time, MIT 6.S094: youtube video: https://youtu.be/nFTQ7kHQWtc
17. A. Karpathy, Stanford University: Back Propagation and Neural Networks1, youtube video: https://youtu.be/i94OvYb6noo
18. S. Muhammad, O. Tomio, K. Kazuhiko, N. Kunihiko, The Development of Anomaly Diagnosis Method Using Neuro-Expert for PWR Monitoring System
19. K. Albelihi, D. Vrajitoru, An Application of Neural Networks to an Autonomous Car Driver, Computer and Information Sciences Department, Indiana University South Bend, South Bend, IN, USA
20. C.H. Aladag, E. Egrioglu, C. Kadilar, Modeling Brain Wave Data by using Artificial neural networks
21. N.M. Mirza, Al Ain University and Al Ain, Comparison of Artificial Neural Networks based on Controllers for Biped Robots**,** United Arab Emirates
22. S. Khairunniza-Bejo, S. Mustaffha, W.I.W. Ismail, Application of Artificial Neural Network in Predicting Crop Yield
23. S.R. Khaze, M. Masdari, S. Hojjatkhah, Application of Artificial neural networks in estimating participation in elections
24. M. Mohatram, P. Tewari, Shahjahan, Applications of Artificial Neural Networks in Electric Power Industry

25. M.G. Moghaddam, F. Bin H. Ahmad, M. Basri, M. Basyaruddin Abdul Rahman, Artificial neural network modeling studies to predict the yield of enzymatic synthesis of betulinic acid ester
26. T.H. Le, *Applying Artificial Neural Networks for Face Recognition* (Department of Computer Science, Ho Chi Minh University of Science, Ho Chi Minh City 70000, Vietnam)
27. A. Mohamed Shahin, M.B. Jaksa, H.R. Maier, State of the Art of Artificial Neural Networks in Geotechnical Engineering
28. W. Sibanda, P. Pretorius, Artificial Neural Networks—A Review of Applications of Neural Networks in the Modeling of HIV Epidemic, North West University, Vaal Triangle Campus Van Eck Blvd, Vanderbijlpark, 1900, South Africa)
29. E. Assidjo, B. Yao, K. Kisselmina, D. Amane, Modleing of an industrial drying process by Artificial neural networks
30. B. Sahoo, D.R. Parhi, P. Biplab Kumar, Analysis of Path Planning of Humanoid Robots using Neural Network Methods and Study of Possible use of Other AI Techniques
31. K. Goryński, I. Safian, W. Grądzki, M.P. Marszałł, J. Krysiński, S. Goryński, A. Bitner, J. Romaszko, A. Buciński, Artificial neural networks approach to early lung cancer detection
32. L. Bertolaccini, P. Solli, A. Pardolesi, A. Pasini, An overview of the use of artificial neural networks in lung cancer research
33. Q. Kadhim Al-Shayea, MIS Department, Artificial Neural Networks in Medical Diagnosis, Al-Zaytoonah University of Jordan Amman, Jordan
34. E.S. Salami, M. Ehetshami, A. Karimi-Jashni, M. Salari, S. Nikbakht Sheibani, A. Ehteshami, A mathematical method and artificial neural network modeling to simulate osmosis membrane's performance
35. K. Wierzchos, J.C. Cancilla, J.S. Torrecilla, P. Díaz-Rodríguez, A.F. Davila, C. Ascaso, J. Nienow, C.P. McKay, J. Wierzchos, Application of artificial neural networks as a tool for moisture prediction in microbially colonized halite in the Atacama Desert
36. J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B.G. DeLacy, J.D. Joannopoulos, M. Tegmark, M. Soljacic, Nanophotonic particle simulation and inverse design using artificial neural networks
37. L. Neamt, O. Matei, O. Chiver, *Finite Element Method Combined with Neural Networks for Power System Grounding Investigation* (Electrical, Electronic and Computer Engineering Department Technical University of Cluj-Napoca Cluj-Napoca, Romania)
38. T. Guillod, P. Papamanolis, J.W. Kolar, Artificial neural network (ANN) based fast and accurate inductor modeling and design. IEEE Open J. of Power Electron. **1**, 284–299 (2020). https://doi.org/10.1109/OJPEL.2020.3012777
39. Padmasudha Kannan, K.V. Nagaraja, An efficient automatic mesh generator with parabolic arcs in Julia for computation of TE and TM models for waveguides. IEEE Access **8**, 109508–109521 (2020)
40. T.V. Smitha, K.V. Nagaraja, An efficient automated higher-order finite element computation technique using parabolic arcs for planar and multiply-connected energy problems. Energy **183**, 996–1011 (2019)
41. T.V. Smitha, K.V. Nagaraja, Application of automated cubic-order mesh generation for efficient energy transfer using parabolic arcs for microwave problems. Energy **168**, 1104–1118 (2019)

Machine Learning-Based Multi-temporal Image Classification Using Object-Based Image Analysis and Supervised Classification

Swasti Patel, Priya Swaminarayan, Simranjitsingh Pabla, Mandeepsingh Mandla, and Hardik Narendra

Abstract During last decade, there has been tremendous research related to the image-based technique in remote sensing; object-based classification is one of the popular techniques due to its capacity of promising results. This paper presents a novel approach where a hybrid method of object-based image analysis and supervised classification is used. The data used in this study is high-resolution multi-spectral 4-band images from 2017 to 2019 provided by the PlanetScope satellite of region Chandigarh, India. First, the data has been pre-processed through passing it in a pipeline of steps followed by a multi-resolution segmentation algorithm and classifying the image into seven classes based on the spectral signature using algorithms like maximum likelihood (ML), support vector machine (SVM), Mahalanobis distance (MD). Comparing the three algorithms, it was observed that SVM and ML have given the highest overall accuracy of 95.21% and kappa coefficient = 0.9159. Also, the overall accuracy 91.91% and kappa coefficient = 0.8860 were achieved.

Keywords Object-based approach · Machine learning · Land cover classification · Change detection · Remote sensing

1 Introduction

There are three techniques for image classification in remote sensing [3]: unsupervised [13] classification, supervised [3, 7, 14] classification, and object-based classification. Multi-temporal images are combined imagery of multiple images of the same place or location taken at different points of time. For research purpose,

S. Patel (✉)
Parul Institute of Technology, Parul University, Vadodara, India
e-mail: swasti.patel20247@paruluniversity.ac.in

P. Swaminarayan
Parul Institute of Computer Applications, Parul University, Vadodara, India
e-mail: priya.swaminarayan@paruluniversity.ac.in

S. Pabla · M. Mandla · H. Narendra
Parul Institute of Engineering and Technology, Parul University, Vadodara, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_22

two multi-temporal images of Chandigarh were taken into consideration, from April 2017 to May 2019. In the traditional approach, land classification [15] and changes were detected with pixel-based [12] image detection. This was replaced by object-based image detection due to higher accuracy compared to former method. Object-based image analysis (OBIA) includes pixels first being grouped into objects based on spectral similarity as geological unit. In OBIA, there are two main approaches segmentation and classification.

Considering all this, a new approach is introduced where OBIA [6] is used for segmentation and supervised learning is used for classification of the classes, instead of using OBIA for both purposes. For this research, the following classes were classified: road, urban area, barren land, water, wooden land, grassland, and fields. While doing this process, it was discovered that if two classes are very near to each other like grassland and fields, then it is recommended to make a new class. Kappa coefficient is used for calculating the accuracy of the classification done by supervised learning algorithms. Furthermore, to identify change detection over time, the thematic change technique is utilized. This new method gives promising results with high-resolution and medium-resolution images.

2 Study Area, Dataset, and Software Used

2.1 Study Area

This study focuses on Chandigarh City located in India which is a perfect suitable for because of its geo-location and well-structured planning. This region has vast variety due to its combination of urban area, rural area, water bodies, mountains, and agricultural areas.

2.2 Data

The high-quality data is provided by Planet [10] via Planet's Education and Research Program. For performing chance detections, two images were acquired; April 11, 2017 and May 4, 2019. All the images are of high resolution. Table 1 provides details for the image acquired. In order to perform the classification, the images need to be resized equally into a shapefile that is being used with all the coordinates of the location as shown in Fig. 1.

Table 1 Summary of the remotely sensed datasets along with the satellite used for this study

Satellite name	Sensor type	Acquisition date	Spectral band (nm)	Pixel size (m)
PlanetScope analytic ortho scene	4-band natural color (red, green, blue) and near infrared (NIR)	04/11/2017–05/04/2019	Blue 455–515 Green 500–590 Red 590–670 NIR 780–860	3 3 3 3

Fig. 1 Multispectral four-band images acquired from PlanetScope of Chandigarh

2.3 Software

For image pre-processing, GIS-software [11] is used, and for OBIA, eCognition Developer 9.0 software is used, and eCognition Developer 9.0 makes it easy for the region of interest (ROI) creation. ROI plays a vital role in training the machine learning algorithms; for thematic change Envi 5.1 software is used.

3 Methodology

The flow of the system is depicted in Fig. 2. The process starts with attaining the images from PlanetScope. This is followed by the process of combination and clipping of the images. Once this is done, the image segmentation is performed using OBIA. After the segmentation, it is essential to define ROI for classification purpose. ML algorithms run on the classes as training data. Final step includes determining the accuracy of the classification. This process is repeated until desired accuracy is achieved.

Fig. 2 Flow diagram that shows OBIA and classification workflow

Fig. 3 Pre-process of image shown here uses mosaicking to merge the image and shapefile to clip the image

3.1 Pre-processing Image

A pipeline is being created for pre-processing the image where initially the image is being merged using mosaicking and then resized using the shapefile [13, 16]. Furthermore, to make the shapefile, the coordinates of the desired location have been pin-pointed by user preference, so both the images are resized evenly shown in Fig. 3.

3.2 OBIA

In this research, segmentation [5, 13] is difficult to perform directly because of the processing that is done in order to get hand-like image layer mixing using equalizing histogram with three-layer mix. This is the pre-processing of the images in order to increase the color vibrancy. Defining the rule-set for segmentation where a process tree should be created is shown below.

Fig. 4 Multi-resolution segmentation with hyper-parameter for composition of homogenous criterion where shape value is set to 0.2 and compactness to 0.8

In this process, selection of the multi-resolution algorithm where first image object domain is to be set to pixel level because there are not any objects for now. Secondly, in the segmentation settings, the hyper-parameter and image layer weights of the algorithm should be blue: green: NIR: red to 1:1:2:1. Nonetheless, the last hyper-parameter is composition of homogenous [1, 12] criterion where the shape is set to 0.2 and compactness [11] to 0.8. In the end, the image is exported with all the 4-band with a file extension of.tif file. This enables the process tree to run and obtain a segmented image as shown in Fig. 4.

3.3 Classification

In earlier studies, the classification part was done with OBIA; thus, machine learning [2, 14] has been incorporated to take care of this part. The segmented image is taken as an input and load it in Envi as a raster file after this makes the ROI (i.e., region of interest) file also referred to as training data in terms of ML in which seven specified classes are used as shown in Table 2. While specifying classes [4, 15], use of ROI type is chosen as a polygon to choose the pixels from the image to get better training

Table 2 Overview about the ROI classes with color of each class used for the classification and brief class description

ROI Classes	Classes Colour	Class Description
Wooden land		Non-wetland which contains large presence of medium to large trees.
Urban land		Area which has artificial surfaces such as houses and buildings
Mountain		Landform which rises above or area which is elevated from the surrounding land
Field		Non-wetland class where farming is present
Grassland		Non-wetland where small grass or bushes
Water		Exposed fresh or saline surface water
Roads		Artificial surface to travel on roads, foot-paths and Highways
Barren land		Non-wetland bare land with no or very less vegetation cover

Fig. 5 Land cover classification results using maximum likelihood algorithm of subset image of April 2017 and May 2019 (i.e., upper half) and support vector machine algorithm of subset image of April 2017 and May 2019 (lower half)

data. The supervised algorithm on both the images has been performed that leads to mimic the classification that is being done by OBIA. Moreover, the input file and ROI are to be passed together when choosing any supervised algorithm.

If the user did not get the desired results from visualizing the image, then it is recommended to start with less training data at first and then increase it accordingly as it is completely trying and viewing method. One can see the figure after the classification with the desired output in Fig. 5, then move on to the next step.

Now to extract the data from the classified image, it has to be converted from raster to vector [7], so the database file will be created that is.vcf file.

3.4 Change Detection

The.vcf file which consists of class_name, class_id, parts, length, and area as headers makes new filtered database file. The details of the training data are given in Table 3. For research purpose, two databases have been used created from both the images. ROI is to be created for ground truth [13, 14] that is used for accuracy assessment. For identifying changes, it is required to provide both the classified images from the timeline and thus calculate the thematic change in order to find out the change.

Table 3 Selection of training data

ROI classes	Number of objects (i.e., pixels) classified	Actual objects (i.e., pixels) taken into consideration
Wooden land	5110	104
Urban land	1406	285
Mountain	21,243	118
Field	10,761	216
Grassland	642	65
Water	4051	193
Roads	1070	33
Barren land	989	193

4 Results

This novel approach for classification is promising because in the first place, the preprocessing of the data is done by setting certain parameters and weights in image bands [6] before proceeding it for segmentation. After this process, an overall accuracy for both the images with a kappa coefficient of 0.915 for support vector machine (SVM) (Ma et al., n.d.), [14] was achieved. For the classified images, the kappa coefficient 0.886 for maximum likelihood (ML) April image is as shown in Table 3. And, both the algorithms that are used in this research have provided the best results. Thus, SVM is great for urban classification and wooden land, while ML is good for classifying road and grassland. One can say that the data here is a delicious cake and the machine learning algorithm provides the icing on this cake. This approach takes less time than the general approach where the classification is done in OBIA. Not only it is taking less time but also gives great accuracy result. The output of change detection over a duration of time is shown in Table 4, and for thematic change, refer to Fig. 6.

5 Discussion

From the past studies, it has already been known that object-based approach provides superior results over pixel-based. The main reason behind this is it allows to incorporate the textual and spatial features [4]. However, the approach our team developed is not only giving better results than traditional pixel-based approach but also gives a competitive performance to object-based approach. Moreover, the multi-resolution (Fig. 4) algorithm cannot be ignored which plays an important role in the object-based approach. So, decided to keep the multi-resolution [5, 13] algorithm because of it distinct benefits, but the software that does it is quite expensive (i.e., eCognition Developer 9.0 [6], (Ma et al., n.d.)), although there are always other options

Table 4 Summary of results achieved by thematic change detection algorithm's used for land cover change detection from 2017 to 2019

Water	Mountain	Field	Urban Land	Road	WoodenLand	Grassland	Barren Land	Row Total	Class Total	
Unclassified	0	0	0	0	0	0	0	0	0	0
Water	246,567	87,581	23,776	5898	28,775	6589	2502	9446	411,134	411,134
Mountain	16,028	7,185,805	3,710,606	37,959	308,036	1,542,371	223,365	973,046	13,997,216	13,997,216
Field	1864	382,384	11,443,395	11,586	9293	639,630	378,427	784,651	13,651,230	13,651,230
Urban Land	4603	115,505	310,648	2,589,912	1,162,597	2882	28,978	889,623	5,104,748	5,104,748
Road	11,147	417,475	134,649	487,379	1,400,830	3421	11,868	423,399	2,890,168	2,890,168
WoodenLand	855	452,152	1,047,513	82	1068	2,393,803	51,187	29,023	3,975,683	3,975,683
Grassland	2238	232,317	2,639,939	34,063,768	14,125	135,545	738,764	679,254	38,505,952	38,505,948
Barren Land	3341	774,413	4,278,263	292,788	351,504	56,081	712,914	2,660,497	9,129,801	9,129,801
Class Total	286,643	9,647,632	23,588,788	37,489,372	3,276,228	4,780,322	2,148,005	6,448,939	0	0
Class Changes	40,076	2,461,827	12,145,393	34,899,460	1,875,398	2,386,519	1,409,241	3,788,442	0	0
Image difference	124,491	4,349,584	−9,937,558	−32,384,624	−386,060	−804,639	36,357,944	2,680,862	0	0

Fig. 6 Change detection results using thematic change detection approach for image of year April 2017 and final image of May 2019

like open-source or economical ways, then, there is a trade-off as the algorithm will perform differently and may not give the desired result.

Here comes the twist after performing the OBIA, decided to replace the traditional classification process and change it with ML (Fig. 2). So, the challenge was how to replace it and get the whole thing working. Accordingly, the defining of rule-set in the object-based approach for the classification is got finally replaced by ROIs (Table 2) where the training dataset was collected and classified into all the classes of one's

interest. Here, the recommendation would be to start with small training data and change accordingly for the classes that do not perform well. It was noted that there may be conflicts in the classes when they are classified by the ML algorithm. In order to overcome this limitation, it was decided to add the classes separately. After all this, once the classification process is over, then do test visually that the algorithm is giving the desired results.

Here, if the algorithm is performing well, then proceed further for the accuracy. But if not then, increase the training data for the classes that one thinks is not performing well. Now, the question is how to select the ML algorithm that is being used in this research. By doing comprehensive [3, 14] exploration, decided to work with support vector machine and maximum likelihood [3, 15] and both performed well (Fig. 5). From the research, it was found/categorized that SVM is giving good results for classes like urban land, fields, water, while maximum likelihood is giving good results for the small classes like road and wooden land. The key feature that really stands out in this approach is the classification part in which the model is doing great with a small amount of training data. Now, referring to the data (i.e., image quality), any medium-quality image [9, 15] to high-quality image will work with this approach. And, if one cannot get hands-on high-resolution [5, 11] data, then go with sentinel-2 [13] data rather than going for landsat-2 [14]. This study was done keeping mainly India in mind as India is a developing nation, so the Chandigarh area was selected for this research. All in all, this approach opens new avenues to explore rather than to only use OBIA.

6 Conclusion

This is a novel approach that is introduced after the object-based approach and pixel-based approach. And, the main motto is to get more accurate results that will benefit the monitoring of change detections and classification. Already object-based approach was giving great results, but it is largely limited to high-resolution data. In order to overcome this limitation, this hybrid approach is developed. Future research has scope to conduct studies by experimenting on different ML algorithms and even apply deep learning algorithms. There is certain deep learning research already underway, but it will take time to achieve the accuracy that this approach and object-based approach provide.

Acknowledgements The support of Planet, Planet Labs, Inc. (formerly Cosmogia, Inc.) is an American private Earth-imaging company based in San Francisco, California, for providing us the high image satellite image data for our research via their Education and Research Program.

References

1. I.L. Castillejo-González, J.M. Pe, F.J. Mesas-carrascosa, F. López-granados, Evaluation of pixel- and object-based approaches for mapping wild oat (Avena sterilis) weed patches in wheat fields using QuickBird imagery for site-specific management. **59**, 57–66(2014). https://doi.org/10.1016/j.eja.2014.05.009
2. G. Chen, Q. Weng, G.J. Hay, Y. He, Emerging trends and future opportunities Accepted us crt. GIScience Remote Sens. 0(0) (2018). https://doi.org/10.1080/15481603.2018.1426092
3. H. Costa, G.M. Foody, D.S. Boyd, Remote sensing of environment using mixed objects in the training of object-based image classifications. Remote Sens. Environ. **190**, 188–197 (2017). https://doi.org/10.1016/j.rse.2016.12.017
4. H.Y. Gu, H.T. Li, L. Yan, X.J. Lu, C. Vi, W.G. Vi, A framework for geographic object-based image analysis (GEOBIA) based on geographic ontology. XL(July), 21–23 (2015). https://doi.org/10.5194/isprsarchives-XL-7-W4-27-2015
5. G.J. Hay, G. Castilla, M.A. Wulder, J.R. Ruiz, An automated object-based approach for the multiscale image segmentation of forest scenes. **7**, 339–359. https://doi.org/10.1016/j.jag.2005.06.005
6. X.U. Jingping, Z. Jianhua, L.I. Fang, W. Lin, S. Derui, W.E.N. Shiyong, W. Fei, Object-based image analysis for mapping geomorphic zones of coral reefs in the Xisha Islands, China. 41201328 (2016). https://doi.org/10.1007/s13131-016-0921-y
7. D. Jovanovi, M. Govedarica, Đ. Ivana, V. Paji,. Object based image analysis in forestry change detection, pp. 231–236 (2010)
8. L. Ma, T. Fu, T. Blaschke, M. Li, D. Tiede, Z. Zhou, Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers (n.d.). https://doi.org/10.3390/ijgi6020051
9. K.V, Mitkari, S. Member, M.K. Arora, R.K. Tiwari, Extraction of Glacial Lakes in Gangotri Glacier Using Object-Based Image Analysis, pp. 1–9 (2017)
10. Planet Team, Planet Application Program Interface: In Space for Life on Earth. (San Francisco, CA, 2017). https://api.planet.com
11. K. Verbeeck, M. Hermy, J. Van Orshoven, An hierarchical object based image analysis approach to extract impervious surfaces within the domestic garden. March 2009, 2009–2012 (2011)
12. T.G. Whiteside, G.S. Boggs, S.W. Maier, International Journal of Applied Earth Observation and Geoinformation Comparing object-based and pixel-based classifications for mapping savannas. Int. J. Appl. Earth Obs. Geoinf. **13**(6), 884–893 (2011). https://doi.org/10.1016/j.jag.2011.06.008
13. A. Whyte, K.P. Ferentinos, G.P. Petropoulos, Environmental modelling & software a new synergistic approach for monitoring wetlands using Sentinels-1 and 2 data with object-based machine learning algorithms. Environ. Model. Softw. **104**, 40–54 (2018). https://doi.org/10.1016/j.envsoft.2018.01.023
14. M. Wieland, Y. Torres, M. Pittore, B. Benito, M. Wieland, Y. Torres, M. Pittore, B. Benito, Object-based urban structure type pattern recognition from Landsat TM with a Support Vector Machine. 1161(July) (2016). https://doi.org/10.1080/01431161.2016.1207261
15. W. Yu, W. Zhou, Y. Qian, J. Yan, Remote sensing of environment a new approach for land cover classification and change analysis: Integrating backdating and an object-based method. Remote Sens. Environ. **177**, 37–47 (2016). https://doi.org/10.1016/j.rse.2016.02.030
16. L. Zhong, L. Hu, H. Zhou, Deep learning based multi-temporal crop classification. Remote Sens. Environ., **221**(March 2018), 430–443 (2019). https://doi.org/10.1016/j.rse.2018.11.032

A Model to Generate Benchmark Network with Community Structure

Shyam Sundar Meena and Vrinda Tokekar

Abstract The studies of social networks focus on the structure and components of networks at different levels. To identify the component in a network, researchers have developed various community detection algorithms. To test the quality of community detection results, networks with well-known community structures are used. But, a very few networks are available for this purpose. Researchers have suggested some models that generate artificial networks with the community. However, most of the proposed models are unable to produce benchmark networks similar to the real-world network. We propose a model that generates benchmark networks for the evaluation of community detection algorithms. The proposed model has been compared with well-known LFR Lancichinetti et al. (Phys Rev 78(4):046110, 2008 [14]) and GLFR Le et al. (2017 26th international conference on computer communication and networks (ICCCN). IEEE, pp 1–9, 2017[15]) models. For performance testing, various structural properties have been analyzed, which are followed by real-world networks. The NMI scores achieved by well-known community detection algorithms were also compared. In experimental analysis, we found that networks generated by our model follow essential properties of real-world networks.

Keywords Community detection · Synthetic networks · Network analysis

1 Introduction

Networks are everywhere! Most of the networks are large, complex and dynamic. Examples of such networks are biological networks [10], social networks:

S. S. Meena (✉)
Department of Computer Science and Engineering, Shri Vaishnav Institute of Information Technology, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore, India
e-mail: ssmeena7@gmail.com

V. Tokekar
Department of Information Technology, Institute of Engineering and Technology, Devi Ahilya Vishwavidyalaya, Indore, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_23

Fig. 1 Community structure in a network

Twitter [12], Facebook [1] and infrastructural networks: electrical power grids [20] and internet routing paths [6].

Some queries related to the complex social networks analysis (SNA) are: How the structure of real-world networks looks? What types of rules the network formation process follows? How nodes make connections to each other in a network? To find answers to these and other related queries, component and node-level studies are required. In recent years, studies have been made to find the meaningful divisions (also known as communities) of networks. The main factors behind the complexity of community detection are the nature and size of networks [7].

Nodes have most of the connections within their community and very few connections to nodes that belong to other communities [8]. In Fig. 1, community structure of a network is shown. In community detection, we identify closely related nodes. It is helpful for epidemic outbreak prevention [11], viral marketing [21], etc. For analysis of the community detection task performance, predicted structure of communities is compared to the original structure of the community. The unavailability of appropriate benchmark networks generated the demand for the development of models to build artificial benchmark networks. We propose a model for building an artificial benchmark network with realistic properties. It works in two steps; first of all, Barabási–Albert preferential attachment model (BA) [2] is used for the generation of network of N nodes. In the second phase, new connections are introduced among nodes, to meet the desired characteristics of real-world networks.

The paper is structured as follows. A review of the popular strategies and models for generating synthetic networks is given in Sect. 2. In Sect. 3, proposed model is described. Experimental setup description is given in Sect. 4. Section 5.1 presents evaluation and discussions on the performance of a proposed model, followed by a conclusion in Sect. 6.

2 Related Work

In this section, a review of the popular strategies and models for the generation of synthetic networks is given. Girvan and Newman [8] presented a benchmark for building a network of 128 nodes. Each node in a group follows the same average degree. But, distribution of degrees of nodes and size of communities is different than real-world networks [14]. Danon et al. [5] proposed changes in Girvan and Newman [8] to generate a network with different sizes of communities. But, generated networks do not follow fat tails in their degree distributions. In the real world, the networks follow power-law degree and community size distribution [2].

LFR benchmark of Lancichinetti et al. [14] can generate networks that follow the power-law distribution of degrees and community sizes. But, it produces artificial networks with low transitivity [18]. According to Newman [17], real-world networks usually have high transitivity. HLFR benchmark [13] can generate networks with hierarchical communities' structure. It is also able to build directed and weighted networks.

In [22], a hierarchical LFR (HLFR) model for the generation of artificial networks with hierarchical community structures was proposed. It follows community structures similar to social networks, biological networks and technical systems. In [15], the authors presented a generalized version of LFR benchmark (GLFR). In place of similar internal and external connections of nodes for each community, the heterogeneous concept was applied.

Though various models for the artificial network have been introduced, it is still hard to achieve satisfactory accuracy and realistic aspects. The serious problem is dissimilarity in structural properties between the generated artificial benchmarks and the available real-world networks.

3 Proposed Benchmark

The possible causes of unrealistic properties observed in the networks created by the LFR [14] benchmark are: (a) initial network generation with the help of configuration model (CM) [16] and (b) assignment of a similar fraction of external links for each node. In the initial step of LFR [14] process, a very flexible model is applied to produce networks of any size, although generated networks do not follow real-world properties such as correlation [19] and transitivity [17]. We propose to adopt a separate generative model with various realistic features. The proposed work, first of all, uses Barabási–Albert preferential attachment model (BA) [2] for initial network generation. In the next step, the fraction of internal connection for each node of communities is selected randomly from a range $[\max(\mu_{\min}, \mu - \Delta\mu), \min(\mu_{\max}, \mu + \Delta\mu)]$ suggested in [15]. The rest of the steps goes similar to the LFR [14] model. Table 1 lists the input parameters of the proposed model.

Table 1 Parameters used in proposed model

Parameter	Description
N	Number of nodes
$< k >$	Average degree
$k_{\max}$	Maximum degree
μ	Mixing parameter
e_d	Minus exponent for the degree sequence
e_c	Minus exponent for the community size distribution
$C_{\min}$	Minimum size of communities
$C_{\max}$	Maximum size of communities
$\Delta\mu$	Mixing heterogeneity parameter

Algorithm 1 proposed Algorithm

Input: N, $< k >$, k_{max}, $C_{\min}$, C_{max}, e_d, e_c, μ and $\Delta\mu$
Output: G

1: Generate N nodes and, decide degree sequence $degSeq$ of each node based on the three parameters $< k >$, k_{max} ,e_d
2: Assign communities to all N nodes using parameters $C_{\min}$, C_{max} and e_c
3: Create a initial graph G using preferential attachment model (BA)
4: **for** each $c \in$ communities **do**
5: μ_c = randomChoice[max($\mu_{\min}$, μ - $\Delta\mu$), min(μ_{max}, μ + $\Delta\mu$)]
6: **for** each $u \in c$ **do**
7: **while** $degree(u) < (degSeq[u]) * \mu_c$ **do**
8: v = randomly select a $node$ from c
9: add an $edge(u, v)$ in G
10: **end while**
11: **while** $degree(u) < degSeq[u]$ **do**
12: v = select a $node$ randomly
13: **if** (v not in c) and ($degree(v) < degSeq[v]$) **then**
14: add an $edge(u, v)$ in G
15: **end if**
16: **end while**
17: **end for**
18: **end for**

The process of benchmark network construction is given in Algorithm 1. In line 1, for the generation of degree sequence of nodes, power-law distribution with exponent e_d is applied. The decision of $k_{\min}$ and $k_{\max}$ is done such that the average degree remains k. In line 2, community assignment for each node is performed such that sum of the sizes of communities equals the number N. The sizes of communities are chosen as per the power-law distribution with exponent e_c. Initially, the assignment of community to nodes is done randomly till the total nodes in a community have not reached the size of the community. The process finally stops when the assignment of the community for each node is done successfully.

In line 3, an initial graph of N nodes and $\alpha * N$ edges is generated using the Barabási–Albert [2] scale-free property, where the value of α is taken 0.5. From 4th to 17th line, rewiring is applied. Here, for each node of each community, new links are attached. The number of nodes that belong to similar community nodes is calculated by μ_c*(degree sequence of node), where μ_c is selected randomly from a range $[max(\mu_{\min}, \mu - \Delta\mu), \min(\mu_{\max}, \mu + \Delta\mu)]$ and remaining $(1 - \mu_c)$*(degree sequence of node) links are connected to the nodes that belongs to other communities.

4 Experimental Setup

The proposed benchmark for the generation of artificial networks is tested for important measures like transitivity, degree distribution and density. Two well-known community detection algorithms are also used, to analyze the behavior and ability of these methods. The selected algorithms are: CNM [4] and Louvain [3]. The implementation of LFR [14] provided by network library [9] is used for performing the all experiments. For the testing of the proposed benchmark, various artificial networks were generated. We used values of various controlling parameters as described below: (1) networks size N= 1000, (2) mixing parameter μ= (0.3, 0.4,0.5 and 0.6), (3) exponent for the degree sequence e_d = 3, (4) value of exponent of community size (e_c) = 1.5, (5) minimum sizes of communities $C_{\min}$ = 20, (6) average degree $< k >$ =20 and (7) mixing heterogeneity parameter $\Delta\mu$ = 0.2.

4.1 Structural Measures of Networks

The proposed model was tested on various network properties that most networks seem to share such as density, transitivity and right-skewed degree distributions.

- **Density**: quantifies the fraction of links that belong to the same group from all possible links [7].

$$D(C) = \frac{|\{(i, j) \in E : i \in C, j \in C\}|}{n(n-1)/2} \tag{1}$$

  where n = total nodes in community C and an edge between node i and j is represented by E.
- **Degree distribution**: networks in real-world follow right-skewed degree distribution [2]. It is described by the following equation.

$$P(k) \approx k^{-\gamma} \tag{2}$$

Fig. 2 Degree distribution in network generated by proposed model

where $P(k)$ represents fraction of nodes having degree k and parameter γ generally falls in the range $2 < \gamma < 3$.

- **Transitivity**: Nodes are said in transitive relation if friends of a node are also friends. This effect is quantified by the transitivity C [20], defined by

$$C = 3 * \frac{t_G}{T_G}. \tag{3}$$

where t_G is a count of triangles and T_G represents a number of connected triples of vertices. C values fall in the range [0.3, 0.6] for most of the networks. For a fully connected graph (everyone friend of everyone else), C will be 1.

5 Results and Discussion

5.1 Evaluation of Structural Properties

First, we studied the structure of real-world networks and their essential properties such as degree distribution, transitivity, density.

Fig. 2 shows the degree distribution of networks generated by the proposed model. From the experimental analysis, it is clear that for different values of the mixing parameter (μ), degree follows power-law (see Eq. 1). It illustrates that generated

Fig. 3 Transitivity of networks

networks by the proposed model maintain small-world and scale-free characteristics. The consistency of the result ensures that the proposed modification in the LFR [14] benchmark did not affect degree distribution structural characteristics.

Figures 3 and 4 show the impact of μ on transitivity and density. For the LFR [14] network, both decrease as the value of μ increases. The transitivity and density of the GLFR [15] network increase when the value of μ grows. The proposed benchmark shows consistent results for different value of μ. The results state that model proposed in this paper is less sensitive in terms of transitivity and density.

5.2 Community Detection Performances

Normalized mutual information (NMI) [5] values are computed to quantify the similarity between resultant communities—assignment to nodes by an algorithm and known communities of nodes. First, different networks were generated by the proposed model, LFR [14] and GLFR [15], for different values of μ. Communities were detected in these networks by two well-known algorithms. The result of Louvain [3] and CNM [4] is presented respectively in Figs. 5 and 6. The NMI values reveal that the performance of both algorithms decreases for larger values of μ as

Fig. 4 Density of networks

Fig. 5 Performance of Louvain algorithm on the three benchmarks

expected. Although, it was noticed that networks generated for various values of mixing parameter μ by LFR [14] are more susceptible than the proposed model and GLFR [15] benchmarks. It concludes that the proposed model can maintain similar structures and properties for the different values of μ.

Fig. 6 Performance of CNM algorithm on the three benchmarks

6 Conclusion

In this paper, we studied various properties of real-world networks and proposed a framework for generating artificial networks. We have analyzed various structural properties of networks generated by the proposed model such as transitivity, degree distribution and density. The suggested model is capable of building different sizes of networks akin to real-world networks. Note that our aim is not to prove that the proposed model is better than the LFR [14] and GLFR [15] models. The basic fact we want to highlight is that these models' networks and community structures are topologically different, and algorithms used for community detection should consider these variations when processing benchmark networks.

Acknowledgements I am thankful to my Ph.D. supervisor Dr. Vrinda Tokekar, Professor, Department of Information Technology, Institute of Engineering & Technology, Devi Ahilya Vishwavidyalaya, Indore, India, for her valuable guidance, motivation and support. I am also thankful to each member of my family for their care and support without which this work would have not been feasible for me.

References

1. L. Backstrom, P. Boldi, M. Rosa, J. Ugander, S. Vigna, Four degrees of separation, in *Proceedings of the 4th Annual ACM Web Science Conference* (2012), pp. 33–42
2. A.L. Barabási, R. Albert, Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
3. V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. **2008**(10), P10,008 (2008)
4. A. Clauset, M.E. Newman, C. Moore, Finding community structure in very large networks. Phys. Rev. E **70**(6), 066111 (2004)
5. L. Danon, A. Diaz-Guilera, J. Duch, A. Arenas, A, Comparing community structure identification. J. Stat. Mech. Theory Exp. **2005**(09), P09,008 (2005)
6. M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology. ACM SIGCOMM Comput. Commun. Rev. **29**(4), 251–262 (1999)

7. S. Fortunato, Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
8. M. Girvan, M.E. Newman, Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)
9. A. Hagberg, P. Swart, D. Chult, Exploring network structure, dynamics, and function using network (2008)
10. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.L. Barabási, The large-scale organization of metabolic networks. Nature **407**(6804), 651–654 (2000)
11. S. Kitchovitch, P. Liò, Community structure in social networks: applications for epidemiological modelling. PloS One **6**(7), e22,220 (2011)
12. H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? in *Proceedings of the 19th International Conference on World Wide Web* (2010), pp. 591–600
13. A. Lancichinetti, S. Fortunato, Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. Phys. Rev. E **80**(1), 016118 (2009)
14. A. Lancichinetti, S. Fortunato, F. Radicchi, Benchmark graphs for testing community detection algorithms. Phys. Rev. E **78**(4), 046110 (2008)
15. B.D. Le, H.X. Nguyen, H. Shen, N. Falkner, Glfr: a generalized lfr benchmark for testing community detection algorithms, in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. (IEEE, 2017), pp. 1–9
16. M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence. Random Struct. Algorithms **6**(2–3), 161–180 (1995)
17. M.E. Newman, The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
18. G.K. Orman, V. Labatut, A comparison of community detection algorithms on artificial networks, in *International Conference on Discovery Science*. (Springer, 2009), pp. 242–256
19. M.Á. Serrano, M. Boguñá, Weighted configuration model, in *AIP Conference Proceedings*, vol. 776 (American Institute of Physics, 2005), pp. 101–107
20. D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world'networks. Nature **393**(6684), 440–442 (1998)
21. L. Weng, F. Menczer, Y.Y. Ahn, Virality prediction and community structure in social networks. Sci. Rep. **3**, 2522 (2013)
22. Z. Yang, J.I. Perotti, C.J. Tessone, Hierarchical benchmark graphs for testing community detection algorithms. Phys. Rev. E **96**(5), 052311 (2017)

Human Activity Recognition Using LSTM with Feature Extraction Through CNN

Rosepreet Kaur Bhogal and V. Devendran

Abstract Human activity recognition is important for detecting anomalies from videos. The analysis of auspicious activities using videos is increasingly important for security, surveillance, and personal archiving. This research paper has given a model which can recognize activities in random videos. The architecture has been designed by using BiLSTM layer which helps to learn a system based on time dependencies. To convert every frame into a featured vector, the pre-trained GoogLeNet network has been used. The evaluation has been done by using a public HMDB51 data set. The accuracy achieved by using the model is 93.04% for ten classes and 63.96% for 51 classes from same data set only. Then, this network is compared with other state-of-the-art method, and it proves to be a better approach for the recognition of activities.

Keywords Action recognition · HMDB51 · Neural network · CNN · LSTM · BiLSTM · Video frames

1 Introduction

Computer vision is a multi-stage domain, which essential framework for automatic extraction, analysis, and comprehension from a single image or image sequence. Man-made visualization from videos is one of the real-world challenges of computer vision in various sectors like industrial, educational, security, consumer etc. In the process of recognition, detection, or tracking, there are two broad terms "actions" and "activity", which are generally used in the vision of the survey. And, there is difference

The original version of this chapter was revised: The incorrect last line in the Abstract has been removed. The correction to this chapter can be found at https://doi.org/10.1007/978-981-16-9967-2_76

R. K. Bhogal (✉)
School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, India
e-mail: rosepreetkaur12@gmail.com

V. Devendran
School of Computer Science and Engineering, Lovely Professional University, Phagwara, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023, corrected publication 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_24

Fig. 1 Steps for activity recognition system [1]

between two words, the action refers to simple movement patterns performed by one person and in a short time such as "raise your hand", "bend", or "swim". They are involved with each other in some way. On the other hand, the activities are in which actions are for long duration, such as when two people shake hands, a soccer team scores a goal, or a group of people hijack a plane, or rob. [1].

An action or action recognition program may include a series of steps. First: insert a video or photo sequence. Second: extraction of feature and action descriptions. Last: interpretation from primitive actions [1] as shown in Fig. 1. Figure 1 shows the different steps to design any activity recognition system. Recognition of human activity remains a challenging research problem because of the following causes. First: there are wide variations of intra-class reason due to the difference in pattern of speed and movement, visual perception, and background clutter. Second: perception of action is closely related to details of complex situations; for example, the characteristics of a scene, the interaction between people and objects. Finally: the diversity and changing nature of the action phase makes it difficult to model the actions (lasting for short time) between successive actions [2].

It is intriguing to know the recent developments in activity recognition which includes its numerous applications. The researchers had heavily relied on clever human and computer-assisted communication programs such as managing a presentation slide or seeing steps to help employees learn and improve their skills. For example, a memory-based attention system can be used to design large public places [3]. There are various other applications which include human activity recognitions like gaming, surveillance, intelligent robots etc. [4]. With that, there is also need to create a smart system for older people that can automatically and quickly detect falls and inform family [5]. There is use of a directionality-base algorithm to track multiple identities and individuals in special care homes [6]. Using deep learning, various frameworks are given by Hasan and Roy-Chowdhury [7] in which the system designed that is calculated hierarchical feature and model which detect actions based on calculated. The researcher is also working on view-oriented activity recognition by pointing to a multi-dimensional hash table followed by a randomized voting system [8]. Niu and Abdel-Mottaleb [9] view invariant recognition is now at a consistent way of seeing human activity using details of movement and structure [3]; it has different sub-systems for behavioural detection and crowd measurement in a single learning mode and did not require camera measurement. Singh et al. [6] used adaptive background–foreground separation and derive directionality-based feature vector for recognition. Maddalena and Petrosino [10] proposed an approach which can handle videos of having moving background, illumination and system execution time also

is less for high resolution. In this [11], use the low-level feature vector that contains the action-wide profile prediction depending on the action and simple temporary objects. In [12], the multiview-based method using deep network has been given and designed using LSTM and CNN models. Fan et al. [13] proposed model multiple features expression and improve long short-term memory (LSTM) network performance. The LSTM-based network is the most acceptable network and many HAR system using RNN because this network is working based in past and future time also. Similarly, in this paper also, the network has been designed using LSTM.

This paper presents a learned system using deep learning network layers which can identify the action from videos. The paper is having various sections; Sect. 2 includes explanation regarding Network model for HAR. Section 3 is about the result and discussion. Then, in the last, there is the conclusion based on work performed.

2 Network Model for HAR

Building video recognition system can be divided into two parts as follows:

2.1 Feature Extraction

The feature extraction is the first step for classification. By using pre-trained network, each video has been converted to sequence vector. The Google network is used as pre-trained network in feature extraction process. GoogLeNet is a 22-layer deep convolutional neural network with overall number of layers is 144. This design network has 224-by-224 image input size. GoogLeNet achieves optimal performance by representing the input frame of video in vector while simultaneously storing important location information. With the concept of transfer learning, the global average pooling layer is used for extraction of features for each video. The last four layers are not used here, because of the fact this network is used only for feature extraction not for classification. Each video represents in the sequence has size of [1024 $\times$ number of frames]. Where, 1024 are the number of features for each frame from a video. The feature extraction process is shown in Fig. 3.

2.2 BiLSTM Network for Prediction

To design an activity recognition system, the deep learning has been emerging area. The deep learning-based techniques are very famous and proved as to be the effective way for automatic action recognition. Earlier models use backpropagation which increases the time duration of the processing because of insufficient backflow error.

The RNN is the model which works using gradient descent algorithm with less-computational complexity [14]. The long short-term memory (LSTM) is one of the recurrent neural networks (RNN) which is given by Hochreiter [14]. Basically, the LSTM network is designed by using various cells, and it also contains an input gate, output gate as well as a forget gate. The cells are used to remember values over some intervals and gates regulate the pass of an information from one cell to another cell. For understanding of the LSTM network, assume the network contains N blocks and M inputs [15]. It has the various elements as stated below:

Block Input: This is used to combine the current input x^t and output of the LSTM y^{t-1} in the last iteration. The corresponding calculation can be done using below equation:

$$z^t = g\left(W_z x^t + R_z y^{t-1} + b_z\right) \tag{1}$$

where W_z and R_z are called as weights with input and output, respectively, and b_z acts as bias weight vector.

Input gate: This is required to update the input gate that combines the input and output of the last iteration and cell value in the last iteration. Equation used to find the input gate value is given below:

$$i^t = \sigma\left(W_i x^t + R_i y^{t-1} + p_i \Theta\, c^{t-1} + b_i\right) \tag{2}$$

In Eq. 2, the W_i, R_i and p_i are the weights which are used to update x^t, y^{t-1} and c^{t-1}, respectively, and b_i is called bias vector corresponding to the input. In Eq. 2, Θ is the pointwise multiplication process of vector.

Forget gate: It determines which information should be retained in cell states. For that, activation function of the gate is calculated based on input and output of state of memory cell at previous step. For that calculation, equation is stated below:

$$f^t = \sigma\left(W_f x^t + R_f y^{t-1} + p_f \Theta\, c^{t-1} + b_f\right) \tag{3}$$

In Eq. 3, the W_f, R_f and p_f are weights which are used to update the x^t, y^{t-1} and c^{t-1}, respectively, and b_f is called bias vector corresponding to the input. In Eq. 2, Θ is the pointwise multiplication process of vector.

Cell: Equation 4 given states how to calculate the value for cell.

$$c^t = (z^t \Theta i^t + f^t \Theta c^{t-1}) \tag{4}$$

Output gate: After combining the input, output, and cell value, the given Eq. 5 used to calculate the output of the gate.

$$o^t = \sigma\left(W_o x^t + R_o y^{t-1} + p_o \Theta c^t + b_o\right) \tag{5}$$

Fig. 2 The LSTM block architecture [15]

Block output: In the last, to calculate the block output using cell value and output gate value, the corresponding Eq. 6 is used:

$$y^t = g\left(c^t\right)\Theta o^t \tag{6}$$

where σ, h and g used to pointwise non-linear activation functions. The sigma is logistic sigmoid function, g and h is $\tanh(x)$ functions, respectively [15] (Fig. 2).

Pipelined two LSTM in an opposite direction together in a network to make BiLSTM network.

In the BiLSTM, the output of a single layer of LSTM during the "t" combined, results of the BiLSTM layers are determined not only by traces from previous vectors but also by indications from future vectors. In this way, more details are keeping by network which is continue from the future data. This network is capable of providing a good level of interaction between consecutive data with video signals [16].

In this work, the BiLSTM network has been used, and it has given good accuracy for activity. In other words, using the BiLSTM network in our model, results in long-term bidirectional relationships which have been subtracted by going back and forth several times in the vector sequence embedded in all parts of the video [16].

The extracted features, as given in Fig. 3, have been given to sequence layer of the network which is used for action recognition. The layout of the LSTM network with layers is given in Fig. 3. The network has various layers named as sequence input, BiLSTM, dropout, fully connected, Softmax, and in the last classification layer, of which the output is class labels. Sequence input layers with input size corresponding to the size of the feature vectors. BiLSTM layer with 200 hidden units with a dropout layer after that. For output, only one label for each sequence is used by setting the "output mode" of the BiLSTM layer to the "final", demonstrated in Fig. 3. The BiLSTM layer is used to study the long-term dependence between the time series of each video. Dropout layer is a type of standard method in which the inputs and reconnection of LSTM units are seamlessly extracted from renewals

Fig. 3 Network model for HAR

and weight recovery during network training. This layer basically helps to reduce the overfitting and improves the overall model with respect to performance. The Softmax layer used to run neural network for multi-class function used to help us determine multiple classes in the images or videos.

3 Experiments Results

To convert video frames into vectors, pre-trained network configuration, i.e. GoogLeNet is used. In this case, a large database of human movements has been used. The database contains 2 GB of video data with 7000 clips over 51 classes. The CNN network has been used to calculate feature output for each video frame. Each videos in the form of sequence as feature vector has extracted from global average layer of the GoogLeNet. After extraction of the feature, then designing of system which will recognize activities. The network which is train and validate with all videos. For the training and validation of the designed network for HAR, the videos are divided into parts. For the training, took 90% of the videos for the total amount of recognition and validation rest 10% of videos.

The BiLSTM hidden units used are 200 units in number with batch size of 16. The Adam optimizer has been used to train the entire network from end-to-end with the initial reading level of 0.0001. The dropout layer of 0.5 is applied to avoid overfitting. The implementation has been done on i5 processor NVIDIA GEFORCE, 8 GB RAM, and 1 TB SSD.

Fig. 4 Distribution of various conditions for the HMDB51, **(a)** visible body part, **(b)** camera motion, **(c)** camera view point, and **(d)** clip quality [17]

3.1 HMDB51 Data set

HMDB51 (human movement database) contains 51 different categories that are captured with considering the complexity of human actions. The database contains 6849 clips divided into 51 action categories, each containing a minimum of 101 clips. The action categories can be grouped into five categories:

(i) Ordinary facial expressions: smile, laugh, chew, and speak.
(ii) Facial actions by deception of something: smoke, eat, and drink.
(iii) Typical body movements: cartwheel, clapping, climbing, climbing stairs, diving, falling down, backhand flip, handstand, jump, lift up, run, run, sit down, sit down, distract, get up, turn, go, and wave.
(iv) Physical movement with object contact: hair brushing, gripping, drawing a sword, pulling a golf ball, hitting something, kicking a ball, picking, pouring, pushing something, cycling, horseback riding, shooting a ball, shooting a bow, shooting a gun, swinging a baseball bat, exercise with the sword, and throwing.
(v) Physical movement of human contact: fencing, hugging, kicking, kissing, punching, shaking hands, and fighting with a sword.

The detailed distribution of database is given as in Fig. 4. Statistically, distribution is divided into various parameters such as visual body, camera movement, camera view point, and clip quality in Fig. 5. This database has been used for training and validation of a system which will predict class.

3.2 Experiment on 10 Classes of HMDB51

The result has been obtained in 10 Classes which are *brush_hair, climb_stairs, cartwheel, catch, chew, clap, climb, dive, draw_sword and driddle.*

Fig. 5 **a** Training progress using HMDB (10 classes) (blue coloured solid line). **b** Variation of loss (red coloured solid line)

In the training and testing process, total 1150 video have been used. The training progress has been shown in Fig. 5. In the same figure, the first graph is between accuracy and iteration. Initially, there is prompt increase in the accuracy until 100 iterations. After 100 iterations, there is very gradual increase and with maximum value reaching at 93.04%. In Figure 6, the second graph is variation of loss with iterations. The validation loss is very high at 0th iteration, and there is a very sharp reduction in loss from 0 to 200th iteration. After that, it reduces very slowly and end up at 0.3245.

3.3 Experiment on 51 Classes of HMDB51

In Fig. 6, the evaluation has been done on complete data set by taking 51 classes. The observation graph has been given in Fig. 6 same as Fig. 5. There are two graphs; first is about accuracy, and second is about validation loss with respect to iterations. With a greater number of classes, the resultant accuracy is not similar to the previous experiment. Till, 200th iteration, the accuracy become more than 40%, and then afterwards, it increases in small percentage and end up at 63.96%. Similarly, in the validation loss-based observation, there is gradual reduction after 2000th iteration and ends up at approx. 1.56. If we consider all the classes of data set HMDB51, then accuracy is 63.96%. The comparison with other state-of-the-art method has been given in Table 1. Based on comparison, it concluded that BiLSTM-based network can also be used for activity recognition.

Fig. 6 **a** Training progress using HMDB (51 classes) (blue coloured solid line). **b** Variation of loss (red coloured solid line)

Table 1 Comparison with other method for HMDB51

Reference for comparison	Accuracy (%)
Zhu et al. [18]	53.8
Tran et al. [19]	54.9
Simonyan and Zisserman [20]	55.4
Srivastava et al. [21]	44.1
Sun et al. [22]	59.1
Wang et al. [23]	57.2
Bilen et al. [24]	57.2
Pan et al. [25]	49.2
Our method	*63.96*

4 Conclusion

In the paper, the model based on BiLSTM has been demonstrated. This model has been compared with other methods, given in Table 1. Every system needs the improvement same as this model also. When the number classes taken are 10, this model gives very good recognition rate. When 51 classes are taken, the accuracy ends up at 63.96%. There are various conditions on which accuracy is depending that are how calculated feature vectors, hyperparameters, and number of layers of the network. The most important concept which whole thing is depending upon is the tuning of hyperparameters of the model. It has significant impact on the model which is based on deep learning concepts.

In future work, the system can be designed after tuning hyperparameter with multiple training videos. This model is based on the concept of supervised learning. The model can also be evaluating on unsupervised learning. The data set used in this paper is single view data set. Ideal model can be created by extending this model which can work for multiview data set.

References

1. P. Turaga, R. Chellappa, V.S. Subrahmanian, O. Udrea, Machine recognition of human activities: a survey. IEEE Trans. Circ. Syst. Video Technol. **18**(11), 1473–1488 (2008)
2. A.-A. Liu, N. Xu, W.-Z. Nie, Y.-T. Su, Y. Wong, M. Kankanhalli, Benchmarking a multimodal and multiview and interactive dataset for human action recognition. IEEE Trans. Cybern. **47**(7), 1781–1794 (2017)
3. K.F. MacDorman, H. Nobuta, S. Koizumi, H. Ishiguro, Memory-based attention control for activity recognition at a subway station. IEEE Multimed. **14**(2), 38–49 (2007)
4. M.B. Holte, C. Tran, M.M. Trivedi, T.B. Moeslund, Human pose estimation and activity recognition from multi-view videos: comparative explorations of recent developments. IEEE J. Sel. Top. Signal Process. **6**(5), 538–552 (2012)
5. N. Lu, Y. Wu, L. Feng, J. Song, Deep learning for fall detection: 3D-CNN combined with LSTM on video kinematic data. IEEE J. Biomed. Heal. Inform. **2194**(c), 1–1 (2018)
6. M. Singh, A. Basu, M.K. Mandal, Human activity recognition based on silhouette directionality. IEEE Trans. Circ. Syst. Video Technol. **18**(9), 1280–1292 (2008)
7. M. Hasan, A.K. Roy-Chowdhury, A continuous learning framework for activity recognition using deep hybrid feature models. IEEE Trans. Multimed. **17**(11), 1909–1922 (2015)
8. J. Ben-Arie, Z. Wang, P. Pandit, S. Rajaram, Human activity recognition using multidimensional indexing. IEEE Trans. Pattern Anal. Mach. Intell. **24**(8) 1091–1104 (2002)
9. F. Niu, M. Abdel-Mottaleb, View-invariant human activity recognition based on shape and motion features, in *IEEE Sixth International Symposium on Multimedia Software Engineering* (IEEE, 2004), pp. 546–556
10. L. Maddalena, A. Petrosino, A self-organizing approach to background subtraction for visual surveillance applications. IEEE Trans. Image Process. **17**(7), 1168–1177 (2008)
11. S. Cherla, K. Kulkarni, A. Kale, V. Ramasubramanian, Towards fast, view-invariant human action recognition, in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (IEEE, 2008), pp. 1–8
12. C. Li, Y. Hou, P. Wang, W. Li, Multiview-based 3-D action recognition using deep networks. IEEE Trans. Hum. Mach. Syst. **49**(1), 1–10 (2018)
13. Z. Fan, X. Zhao, T. Lin, H. Su, Attention-based multiview re-observation fusion network for skeletal action recognition. IEEE Trans. Multimed. **21**(2), 363–374 (2019)
14. S. Hochreiter, Long Short-Term Mem. **9**(8), 1735–1780 (1997)
15. G. Van Houdt, C. Mosquera, G. Nápoles, A review on the long short-term memory model. Artif. Intell. Rev. **53**(8), 5929–5955 (2020)
16. X. Wu, Q. Ji, TBRNet: two-stream BiLSTM residual network for video action recognition. Algorithms **13**(7), 1–21 (2020)
17. H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, T. Serre, HMDB: a large video database for human motion recognition. Proc. IEEE Int. Conf. Comput. Vis. 2556–2563 (2011)
18. Y. Zhu, Y. Long, Y. Guan, S. Newsam, L. Shao, Towards universal representation for unseen action recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 9436–9445 (2018)
19. D. Tran, J. Ray, Z. Shou, S.-F. Chang, M. Paluri, ConvNet architecture search for spatiotemporal feature learning, no. section 3 (2017)

20. K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos. Adv. Neural Inf. Process. Syst. **1**, 568–576 (2014)
21. N. Srivastava, E. Mansimov, R. Salakhutdinov, Unsupervised learning of video representations using LSTMs, in *32nd International Conference Machine Learning ICML 2015*, vol. 1 (2015), pp. 843–852
22. L. Sun, K. Jia, D.Y. Yeung, B.E. Shi, Human action recognition using factorized spatio-temporal convolutional networks, in *Proceedings of IEEE International Conference Computer Vision*, vol. 2015 Inter (2015), pp. 4597–4605
23. H. Wang, C. Schmid, A. Recognition, T. Iccv, Action recognition with improved trajectories to cite this version : action recognition with improved trajectories (2013)
24. H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, Action recognition with dynamic image networks. IEEE Trans. Pattern Anal. Mach. Intell. **40**(12), 2799–2813 (2018)
25. T. Pan, Y. Song, T. Yang, W. Jiang, W. Liu, VideoMoCo: contrastive video representation learning with temporally adversarial examples (2021)

Sentiment Classification of Higher Education Reviews to Analyze Students' Engagement and Psychology Interventions Using Deep Learning Techniques

K. R. Sowmia, S. Poonkuzhali, and J. Jeyalakshmi

Abstract Globally, higher education institutions are closed due to the COVID-19 pandemic. The sudden shift to online education excites most teachers and students. The professors are researching online learning platforms. They are only involved in face-to-face teaching in traditional teaching platforms. There are many concerns about the quality of online education. This paper proposes a framework for comparing online learning with traditional learning using emotions, learner perception, instructors, student engagement, understanding, effectiveness, learning outcome, peer collaboration, constraints, and comparisons. Deep learning algorithms like LSTM, GRU, and RNN classify the reviews. Students are positive during online learning in higher education according to LSTM, GRU, and RNN experimental analysis. Students are becoming more comfortable with online learning environments for higher education, according to detailed survey results.

Keywords Student engagement · Psychology interventions · Sentiments · Reviews · Online learning · Higher education

1 Introduction

Online education relies heavily on ICT to improve the activities and tools. Sentimental classification is used in higher education to analyze student comments and categorize them as positive, negative, or neutral towards online learning. In higher education, it is common practice to collect student feedback via comments or Google Forms to analyze their attitudes towards various subjects.

To analyze the reviews, machine learning techniques such as multinomial naïve Bayes, stochastic gradient descent, SVM, random forest, and multilayer perceptron classifier can be used. The results can be used to improve online education.

K. R. Sowmia (✉)
Department of IT, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India
e-mail: sowmia.kr@rajalakshmi.edu.in

K. R. Sowmia · S. Poonkuzhali · J. Jeyalakshmi
Department of CSE, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_25

The responses from Google Forms help understand student collaboration. Students actively participate in online learning, create activities, discuss and evaluate them, thus improving their overall performance.

Deep learning now plays a major role in improving sentiment analysis results in fields like academic research, healthcare, tourism, and textile industries. The proposed sentiment classification system addresses student psychology and online learning engagement. Sentimental classification can be used to classify student feedback using deep learning neural network algorithms like LSTM, GRU, and RNN. Sentimental classification using deep neural networks will provide reliable data to improve the quality of learning. The overall contribution of this work can be summarized as follows as:

1. The extraction of reviews about online learning in higher education based on parameters such as emotions, learner's perception, instructors, student engagement, student understanding, student effectiveness, learning outcome, peer collaboration, constraints, and comparison with traditional learning.
2. Several kinds of data preprocessing steps have been carried out namely tokenization, stopword removal, and lemmatization
3. In the processed dataset, class imbalance problem has been addressed and rectified by using data augmentation techniques
4. The classification of the positive and negative sentiment using deep learning techniques such as LSTM, GRU, and RNN
5. The visualization of the resultant dataset has been done using word cloud for online education for higher education.

The organization of the paper can be described as follows as Sect. 2 explains about the relevant work, Sect. 3 describes the proposed model for Sentiment Classification, then Sect. 4 demonstrates about the obtained results, and finally, Sect. 5 concludes the proposed work.

2 Related Work

A detailed discussion of the literature review that has been conducted for the sentimental classification of higher education is as follows:

There is no simple definition to reveal the concept of student engagement in higher education. Researchers have come up with various suggestions to identify positive psychology interventions in higher education to improve student engagement in online learning. Research has been done on the positive effects of psychology interventions to enhance student engagement. The parameters that relate to psychology interventions and student engagement are self-perceptions, motivation, happiness, sentimentally connected with the organization, positive feedback in LMS etc [1]. To make students active in an online learning environment, the personal and academic data of students must be related to avoid student disengagement in online learning [2].

Hu et al. [3] conducted research on how and why students learn, as well as factors influencing student engagement. The model proposed for student engagement is based on two goals: future goals and school satisfaction. Kahu [4] frames student engagement in higher education based on four perspectives, namely the behavioral perspective, the psychological perspective, the sociocultural perspective, and the holistic perspective. The behavioral perspective is based on student behavior and institutional practice; the psychological perspective is based on the individual psychosocial process; the sociocultural perspective is based on the socio-political context; and the holistic perspective is based on the broader view of engagement [5].

The academic outcomes of students and their relationship with engagements is an open question for research in higher education [6]. Devito [7] suggested the factors influencing student engagement are grouped into five clusters: Peer collaboration among students and active involvement in learning, interaction between tutors and students, academic challenges faced by students and supporting faculty and family members. Apart from the psychological influence of institutions and tutors on student motivation and achievement, parents have a great influence on their learning styles and activities [8]. Boulton et al. come with a survey on student engagement with the results of positive interaction between student engagement and their happiness [9].

3 Framework of the Proposed Model for Sentiment Classification to Analyze Students' Engagement and Psychology Interventions

Figure 1 shows the architecture diagram about the proposed model for sentiment classification to analyze students' engagement and psychology interventions.

Fig. 1 Architecture diagram for sentiment classification to analyze students' engagement and psychology interventions

In the above depicted sentiment analysis model, the Twitter data collected based on student engagement, emotions, understanding, peer collaboration, psychological interventions pass through various data preprocessing techniques such as removing retweets, uppercase, numerals, special characters, and ascents. Feature extraction is being done and extracted data is classified using different classification algorithms like LSTM, GRU, and RNN. The elaborate process flow in listed below:

3.1 Dataset Generation

The dataset has been collected from Twitter API which consist of reviews based on student engagement and psychology interventions in online learning. The reviews are extracted from Twitter API based on the parameters such as student feedback, student emotions, learners perception, instructor feedback, student engagement, student understanding, student effectiveness, learning outcome, and peer collaboration. Nearly, 3978 tweets are stored as a .csv file for further analysis.

3.2 Data Preprocessing

Data preprocessing is done to make the dataset in a standard format to make learning algorithms more effective. It is done using natural language processing techniques. The preprocessing methods used here are tokenization, stopword removal, and lemmatization [10]. The reviews may be long sentences with a variety of phrases, emoticons, punctuation, and other elements. All non-alphanumeric characters must be omitted from the reviews. All retweets are also removed in this process. The process also includes the deletion of all uppercase letters and converting all of the texts to lowercase letters. Tokenization is applied where we split down long text strings into smaller chunks or tokens. It is also important to remove unnecessary vocabulary. Stopwords are the most common words in any language (for example, articles (a, an, the), prepositions (in, at, on, of, etc.), and these words do not add much detail to the text. Stopword removal is the method of eliminating certain words from a document so that the low-level details can be focused on. The process of lemmatization is also done to transform a term into its simplest form, such as 'learning' to 'learn.' All these processes are done on the tweet dataset to perform classification as positive or negative sentiments experienced by students in online learning.

3.3 Feature Extraction

In this process, the initial raw data is reduced to smaller units for making further processing easier. In this approach, these words are pre-trained, tokenized words are re-joined, and then it is replaced with the existing words.

3.4 Classification

In this step, the classification of the sentiments into two classes such as positive reviews and negative reviews using LSTM, GRU, and RNN. The structure and tuned hyperparameters of these algorithms are described below:

3.4.1 LSTM

The embedded layer, which uses 500 length vectors to construct the LSTM model, comes first. It is followed by the LSTM layer, which has 32 memory units (smart neurons). Finally, as it is a classification problem, the dense output layer with a single neuron and a sigmoid activation function makes a 0 or 1 prediction for two classes, namely positive and negative. Since this is a binary classification problem, log loss is used as the loss function, which is binary cross entropy in keras. The optimization algorithm ADAM is employed. With a batch size of 128 epochs, the model is then suited for 7 epochs. The evaluation metrics such as F1, precision, recall, and accuracy are estimated using the matplotlib library.

3.4.2 GRU

The embedded layer, which uses 100 length vectors to construct the GRU model, comes first. It is followed by the GRU layer, which has 32 memory units, (smart neurons). Finally, as it is a classification problem, the dense output layer with a single neuron and a sigmoid activation function makes a 0 or 1 prediction for two classes, positive and negative. Since this is a binary classification problem, log loss is used as the loss function, which is binary cross entropy in keras. The optimization algorithm ADAM is employed. With a batch size of 128 epochs, the model is then suited for 7 epochs. The evaluation metrics such as F1, precision, recall, and accuracy are estimated using the matplotlib library. The Algorithm 1: GRU for sentiment classification is depicted in (Fig. 2).

Algorithm 1: GRU for Sentiment Classification

Result: h_t is calculated

x_t is initialized

while *TRUE* **do**

  instructions;

  if *update signal* **then**

    update the gate $z_t : z_t = \sigma\left(W^{(z)}x_t + U^{(z)}h_{t-1}\right)$

    reset the memory : $r_t = \sigma\left(W^{(r)}x_t + U^{(r)}h_{t-1}\right)$

    new memory updated using tanh activation

    Final memory is calculated: $h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h_t'$

  else

    Compute: $z_t = \sigma\left(W^{(z)}x_t + U^{(z)}h_{t-1}\right)$

  end

end

Fig. 2 Algorithm 1: GRU for sentiment classification

3.4.3 RNN

The embedded layer, which uses 100 length vectors to construct the RNN model, comes first. It is followed by the RNN layer, which has 32 memory units, (smart neurons). Finally, as it is a classification problem, the dense output layer with a single neuron and a sigmoid activation function makes a 0 or 1 prediction for two classes, positive and negative. Since this is a binary classification problem, log loss is used as the loss function, which is binary cross entropy in keras. The optimization algorithm ADAM is employed. With a batch size of 128 epochs, the model is then suited for 7 epochs. The evaluation metrics such as F1, precision, recall, and accuracy are estimated using the matplotlib library.

4 Results and Discussions

The implementation of these dataset is done on Windows platform with 8 GB size of RAM, x-64-based processor with Intel CORE i5 10th Gen processor. The three deep learning techniques have been implemented in the Google Collaboratory environment.

Model accuracy graph is drawn to represent the number of classifications predicted correctly by a model to the total number of predictions done. Model loss graph is drawn to measure the error in our trained model.

Figure 3a depicts the increase in accuracy after every epoch as the LSTM model is being trained, and Fig. 3b depicts the decrease in loss after every epoch as the LSTM model is being trained.

Figure 4a depicts the increase in accuracy after every epoch as the GRU model is being trained, and Fig. 4b depicts the decrease in loss after every epoch as the GRU model is being trained.

Fig. 3 **a** Accuracy graph for the LSTM model. **b** Loss graph for the LSTM model

Fig. 4 **a** Accuracy graph for the GRU model. **b** Loss graph for the GRU model

Figure 5a depicts the increase in accuracy after every epoch as the RNN model is being trained, and Fig. 5b depicts the decrease in loss after every epoch as the RNN model is being trained.

Fig. 5 **a** Accuracy graph for the RNN model. **b** Loss graph for the RNN model

Fig. 6 Word cloud for online learning in higher education

Table 1 Comparing LSTM, GRU, and RNN models with performance metrics

Performance metrics	LSTM	GRU	RNN
Accuracy (%)	90.38	97.3	75.14
F1 score	0.89	0.9	0.71
Precision	0.89	0.9	0.71
Recall	0.89	0.9	0.71

Figure 6 is a word cloud that displays the most frequently used words in the higher education reviews. It can be inferred that the words "online," "technical," "career" etc. are frequently used in the reviews.

The performance metrics for the classification algorithms such as LSTM, GRU, and RNN are calculated and evaluated.

The performance metrics such as accuracy, F1 Score, precision, and recall are calculated for the different deep learning algorithms.

$$\text{Accuracy} = \frac{\text{Total number of Correct Predictions}}{\text{Total number of Predictions}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Total Positive} + \text{False Negative}}$$

$$\text{recision} = \frac{\text{True Positive}}{\text{Total Positive} + \text{False Positive}}$$

$$F1\ \text{Score} = 2 \times \left(\frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \right)$$

Table 1 compares the trained models, LSTM, GRU and RNN based on various performance metrics. It can infer that GRU performs well compared to LSTM and RNN as the accuracy is higher.

5 Conclusion

Sentiment analysis is performed using text classification algorithms like LSTM, GRU, and RNN. The models are trained to evaluate the student engagement and psychology interventions in online learning using ten parameters such as student

feedback, student emotions, learners perception, instructor feedback, student engagement, student understanding, student effectiveness, learning outcome, and peer collaboration. The results are compared with various performance metrics, and then, the best model is chosen. It can be inferred that the GRU model out performs the LSTM and RNN model in accurate text classification. The work can be further extended by taking online surveys from students based on the above-mentioned parameters. Statistical analysis can be done on the data to work on the analysis of student engagement and psychology interventions in higher education during online learning.

Acknowledgements I present my sincere thanks towards "The Centre for Data Science", Rajalakshmi Engineering College, Chennai, Tamil Nādu, India for extending support to the research work by providing necessary infrastructure and software facilities.

References

1. R. Cerezo, M. Sa´nchez-Santilla´n, M.P. Paule-Ruiz, J.C. Nu´ñez. Students' LMS interaction patterns and their relationship with achievement: a case study in higher education. Comput. Edu. **96**, 42–54 (2016)
2. J. Hammill, T. Nguyen, F. Henderson, Student engagement: the impact of positive psychology interventions on students, Act. Learn. High. Edu. SAGE. (2020)
3. Y-L. Hu, G. Ching, Factors affecting student engagement, in *2012 Conference on Creative Education* (2012)
4. E.R. Kahu, Framing student engagement in higher education. Stud. High. Edu. (2013)
5. G.V. Uma, R. Kumar, S. Poonkuzhali, R. Kishore Kumar, K. Sarukesi, Collapsed concept map game: an innovative online formative knowledge assessment method. Soc. Dev. Teach. Bus. Process. New Net Environ. B H **9**(4), 818–827 (2014)
6. K.S. Na, Z. Tasir, Identifying at-risk students in online learning by analysing learning behaviour: a systematic review, in *2017 IEEE Conference on Big Data and Analytics (ICBDA)* (16–17 Nov 2017)
7. M. DeVito, Factors influencing student engagement. Summer (2016)
8. C.A. Boulton, C. Kent, H.T.P. Williams, Virtual learning environment engagement and learning outcomes ata 'bricks-and-mortar' university. Comput. Educ. **126**, 129–142 (2018)
9. C.A. Boulton, E. Hughes, C. Kent, J.R. Smith, H.T.P. Williams, Student engagement and wellbeing over time at a higher education institution. PLoS ONE **14**(11), e0225770 (2019). https://doi.org/10.1371/journal.pone.0225770
10. K.R. Sowmia, S. Poonkuzhali, Artificial intelligence in the field of education:a systematic study of artificial intelligence impact on safe teaching learning process with digital technology. J. Green Eng. **10**(4), 1566–1583 (2020)

Strategic Network Model for Real-Time Video Streaming and Interactive Applications in IoT-MANET

Atrayee Gupta, Bidisha Banerjee, Sriyanjana Adhikary, Himadri Shekhar Roy, Sarmistha Neogy, Nandini Mukherjee, and Samiran Chattyopadhyay

Abstract In this paper, we discuss how to handle the issue of energy loss, delay during processing and transmission in IoT-MANET for real-time applications such as video in surveillance networks and also interactive applications in smarthome. Here, we propose a hierarchical strategic network model to reduce the cost of computation, latency and energy consumption. The proposed model when compared with random and normal grid topology gives better performance. Results of comparison show that the proposed hierarchical model provides an average of 166.29 Mbps for bandwidth utilization, 63.92 Mbps of throughput, 0.000095 s of jitter and 12.28 % of packet loss for transferring video data in mobile conditions. Also, each supernode receives an average 6.22 frames per second out of 15 transmitted frames.

Keywords Intelligent collaboration · Video streaming · Edge-MANET

1 Introduction

In contemporary researches, mobile edge computing components are integrated with MANET, WSN [6] and IoT to build up smartspaces [1]. IoT-MANET [1] has been built as a small but scalable network to support a variety of applications such as smart-building/homes, precision farming, defence surveillance, augmented reality. For our daily well-being, these technologies support things of various types to interact among themselves. For example, depending on the ambient temperature, the automatic air-conditioning in a smarthome system may adjust its cooling effect. These interactions may not involve exchange of a large amount of data or consuming enough network bandwidth, however, in some cases challenge is low latency and real-time decision

A. Gupta (✉) · B. Banerjee · H. S. Roy · S. Neogy · N. Mukherjee
Department of Computer Science and Engineering, Jadavpur University, Jadavpur, India
e-mail: atrayeegupta@gmail.com

N. Mukherjee
e-mail: nmukherjee@cse.jdvu.ac.in

S. Adhikary · S. Chattyopadhyay
Department of Information Technology, Jadavpur University, Jadavpur, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_26

making at the edge which involves the use of efficient network design and intelligent collaboration among nodes. Associated edge technologies include fog, cloudlets, edge-based MANET technologies [3]. Also, in these contexts, some processing is best done at the edge due to security reasons. Here, mobile edge computing [4] uses high computing resources at strategic locations to facilitate processing and storage in real time. It aids real-time application by enabling data fusion and intelligent decision making from multiple sensor devices. For example, in a tactile network, processing video surveillance data for target tracking in real-time requires high network bandwidth, low latency and low jitter which are best supported in edge rather than the cloud. Hence, in both surveillance and interactive cases discussed above, the challenge is to design a network that can reduce communication delay (which affects response time) and energy consumption (which directly affects network lifetime). Therefore, the purpose of this research work is to find a network design to reduce the energy consumption (due to complex computation and data transmission) and delay for resource-constrained IoT/edge-based MANET nodes which support different types of applications. Therefore, in this paper, we propose a hierarchical strategic network model and compare the same with random and grid topology. We also check the performance of a video streaming application and an interactive application in these three topologies over various layer-2 and layer-3 routing protocols of MANET. The results of the comparison are obtained in NS3 simulator with selected parameters such as average throughput, bandwidth consumption, packet delay, jitter, packet loss and also video frames per second (FPS) received at the sink.

The paper is organized as follows. Section 2 describes the hierarchical model of the network, Sect. 3 discusses the strategic approach, and Sect. 4 discusses the simulation parameters. Results are discussed in Sect. 5. Paper is concluded in Sect. 6.

2 Hierarchical Model

We consider a heterogeneous network cluster as shown in Fig. 1. Here, twelve ordinary nodes (red), R_{ij}, distributed throughout the network, with limited resources are used primarily for data collection and routing. These are equipped with cameras and stream video feeds to more powerful super nodes. These ordinary nodes may or may not have GPS, from which they ascertain their location. They communicate with other R_{ij} and also with supernodes to transmit video, image, audio, text data and also control information.

The other nodes, in the above scenario, responsible for intelligent processing are called supernodes (SuN) or cluster head (CH) used interchangeably in this paper. In Fig. 1, there are four supernodes (green) equipped with GPS (e.g. laptops). They have higher energy, storage, computation capability and high antenna gain to cover areas larger than R_{ij}'s. Each of them is equipped with a camera. Hence, they can also detect objects under the coverage area they are monitoring. They communicate with the ordinary nodes, collect data, aggregate and process them and send a collective decision to base station (BS). For example, CH can process the image frames of

transmitted video from R_{ij}s and decide the number and type of target. It can also find target's possible location and future direction collaborating with other CHs. If no such information is revealed, the supernode may refrain from sending to BS. Moreover, these CH's have no/lower mobility compared to R_{ij}s. The purpose of dividing the area in Fig. 1 is to reduce the number of hops and mitigate interference by restricting the number of control information within a MANET. We call this a hierarchical model for data collection and efficient processing.

3 Strategic Approach

In this approach, we divide the large network coverage area into small grids to reduce the number of hops and induce quicker communication between source to destination. The entire 1000 m × 1000 m area is subdivided into zones to place sixteen nodes as shown in Fig. 1. In a typical zone ($A1$, $B2$, $C3$ and $D4$), we place three mobile R_{ij} that can communicate with CH. We denote CH_A, CH_B, CH_C and CH_D as cluster heads in each zone. For example, R_{11}, R_{12}, R_{13} in zone $A1$ communicate with CH_A within four hops. However, if we used 12 randomly placed nodes, the number of hops and control message exchange could have been higher in case there is route reconfiguration due to mobility of nodes. Therefore, this placement strategy also reduces the number of control messages in the MANET. Here, each R_{ij} moves inside the (250 m × 250 m) region and does not move to other zones. Each CH collects zone information from R_{ij} and communicates among themselves and routes to BS also within four hops. Blue lines in Fig. 1 represent hop count. Hence, with this approach, the number of hops can be reduced from source to sink with less

Fig. 1 Strategic placement of nodes

Table 1 Simulation parameters-1

Parameter	Value
General	
Simulation time	300
Total number of nodes	16
Number of source	12
Number of sinks	4
Coverage area	$1000 \times 1000\ m^2$
Data rate	331,776 kbps(15 FPS)
Packet size	1472 Bytes
Physical layer	
Transmission power	30 dBm
Tx antenna gain	9 dBi
Rx antenna gain	9 dBi
Propagation loss model	FriisPropagationLossModel
PropagationDelayModel	ConstantSpeedPropagationDelayModel
MAC layer	
WiFi standard	WIFI_PHY_STANDARD_80211ac
RemoteStationManager	ns3::ConstantRateWifiManager
DataMode	VhtMcs8
ControlMode	VhtMcs8
MCS	8
ChannelWidth	80
Nodes 0–3 channel number	106 (80 Mhz) 5 Ghz frequency
Nodes 4–7 channel number	122 (80 Mhz) 5 Ghz frequency
Nodes 8–11 channel number	138 (80 Mhz) 5 Ghz frequency
Nodes 12–15 channel number	155 (80 Mhz) 5 Ghz frequency

than five hops reducing latency and data loss during transmission. We also used different frequencies or channel numbers, shown in Table 1, to mitigate inter-cluster interference among nodes during data transmission from R_{ij} to SuN.

4 Simulation

To check how our proposed grid performs compared to random and generic grid topology, we simulated the above scenario in NS3. Here, we perform three test simulations as scenarios-1 (random), 2 (grid) and 3 (proposed grid -p_grid). We have used scenario-1 for random topology as shown in Fig. 2a where all 16 nodes are randomly placed inside 1000 m × 1000 m area. Here, any source node can send

Fig. 2 Communication at different instants in random and grid topologies

Fig. 3 Communication at different instants in proposed grid topology

data to any other sink node. In scenario-2, we placed the nodes in a grid [4 × 4] as shown in Fig. 2b. The distance between each node is 250 m. Here, any node can send collected data to other sink nodes. Green nodes represent sink placed in the first column of the grid. Scenario-3 shows the proposed grid in Fig. 3a, where each node is restricted to move around a bounded rectangle of 250 m × 250 m. Nodes under the supervision of a particular SuN cannot move to other zones. When any node hits the boundary, it bounces back and chooses another direction. Simulation parameters are detailed in Tables 1 and 2.

Table 2 Simulation parameters-2

Parameter	Value
Mobility model	ns3::RandomWaypointMobilityModel
	ns3::ConstantPositionMobilityModel
Node speed	1.94 m/s
Node pause time	2 s
Topology (grid)	GridPositionAllocator
MinX, MinY	0.0, 0.0
DeltaX, DeltaY (Distance Between Nodes)	250.0, 250.0
GridWidth	4 nodes per column wise
Topology (Random)	RandomRectanglePositionAllocator
16 nodes placed randomly inside $1000 \times 1000\,m^2$	
X length	1000 m
Y length	1000 m
Topology (Proposed grid)	RandomRectanglePositionAllocator
Single node is randomly placed inside a restricted area of $250 \times 250\ m^2$ (For each 16 nodes, each rectangle is chosen so that they do not move to other zone)	
X length	250 m
Y length	250 m
Routing protocol	Global, AODV, OLSR, DSDV, HWMP
Packet interval	0.1 s
IP address	10.1.1.0/24
Transport layer	UDP
Application layer	
I-server and I-client	Sends and receives simple text-message
OnOff application	Video streaming
OffTime	ns3::LogNormalRandomVariable
	[Mu=0.4026|Sigma=0.0352]
OnTime	ns3::WeibullRandomVariable
	[Shape=10.2063|Scale=57480.9]
MaxBytes	10,989,173
Server start and stop time	0, 300 s
Client start and stop time	0.1 s after server, 299 s

Table 3 Bandwidth comparison for three scenarios

	Bandwidth (Random)	BandWidth (Grid)	Bandwidth (P_Grid)
Video (Static) (Mbps)	196.34	209.22	217.59
Video (Mobile) (Mbps)	150.86	160.65	166.29
Interactive (Static) (Kbps)	159.67	160.34	162.36
Interactive (Mobile) (Kbps)	158.71	158.12	160.36

Table 4 Comparing packet delay for three scenarios

	Delay (Random) (s)	Delay (Grid) (s)	Delay (P_Grid) (s)
Video (Static)	16.12	13.52	10.47
Video (Mobile)	21.78	22.25	18.99
Interactive (Static)	0.00230	0.00238	0.00237
Interactive (Mobile)	0.00353	0.00272	0.00261

5 Results and Discussion

A. Comparing Bandwidth Utilization

Tabulated results are average of all layer three (L3-AODV, OLSR, DSDV, etc.) and layer two (L2-Hybrid Wireless Mesh Protocol (HWMP)) routing protocols used in this paper. Table 3 provides overall results of the comparison of bandwidth in three scenarios for L3 and L2 protocols. Theoretically, we know a higher bandwidth network allows transferring more data. Also, we can observe in the case of video and interactive applications that random topology has the lowest, but proposed grid has the highest bandwidth utilization on average for all MANET protocols used above. These results show that bandwidth utilization of networks can vary by changing network topology, type of MANET routing protocol and mobility of nodes.

B. Comparing End-to-End Delay

Term delay is used in this paper to refer to packet delay or average end-to-end delay as per [2]. Table 4 shows packet delay for three topologies in static and mobile condition. In Table 4, packet delay in the proposed grid for both video and interactive applications shows minimum delay compared to other topologies.

Table 5 Comparing jitter for three scenarios

	Jitter (Random) (s)	Jitter (Grid) (s)	Jitter (P_Grid) (s)
Video (Static)	0.0001029	0.0000882	0.0000705
Video (Mobile)	0.0001108	0.0001137	0.0000946
Interactive (Static)	0.002133	0.001324	0.001481
Interactive (Mobile)	0.002034	0.001634	0.001435

Table 6 Comparing throughput for three scenarios

	Throughput (Random)	Throughput (Grid)	Throughput (P_Grid)
Video (Static) (Mbps)	74.57	83.78	104.45
Video (Mobile) (Mbps)	60.33	53.29	63.92
Interactive (Static) (Kbps)	81.24	80.72	81.27
Interactive (Mobile) (Kbps)	80.37	79.92	80.85

C. Comparing Jitter

Packets with high jitter can pause the video for few moments, and in the worst case, it may also induce packet loss at the receiver. We calculated mean jitter from paper [2]. In Table 5, mean jitter values of all three topologies are tabulated for video and interactive application where video application in proposed grid shows significantly lower jitter than interactive application.

D. Comparing Throughput

As bandwidth provides an estimate of how much data can travel, while throughput provides actual value of transmitted data. We calculated throughput using formula from paper [5]. Table 6 shows throughput decreases due to mobility in both video and interactive applications. Proposed grid shows a higher throughput in both video and interactive applications.

E. Comparing Packet Loss

Percentage of packet loss is calculated with respect to number of packets sent. Table 7 shows packet loss comparison for three topologies in static and mobile conditions for video and interactive data. We can observe that mobility-induced packet loss is slightly higher in all three topologies. Also, interactive application shows a lower

Table 7 Comparing packet loss for three scenarios

	PacketLoss (Random) (%)	PacketLoss (Grid) (%)	PacketLoss(P_Grid) (%)
Video (Static)	11.39	11.06	7.08
Video (Mobile)	13.92	13.60	12.28
Interactive (Static)	1.74	3.12	0.41
Interactive (Mobile)	4.86	5.87	2.75

Fig. 4 Comparing frames per second for video application

Table 8 Comparing FPS for three scenarios

	FPS (Random)	FPS (Grid)	FPS (P_Grid)
Video (Static)	7.28	8.37	10.03
Video (Mobile)	5.28	5.12	6.22

packet loss percentage compared to multimedia data transfer as expected. The proposed grid has a minimum packet loss percentage among three scenarios.

F. Comparing Frames Per Second

We measure the number of frames received at the sink for three topologies for video data transfer at 15 FPS from each source. Figures 4a, b show FPS in static and mobile conditions, respectively. HWMP protocol has 5.5 FPS on average in three topologies in mobile condition. For others, FPS drops in the range of 4.3–7.5 for all three scenarios during mobility. Comparison of FPS is shown in Table 8, which also depicts that the proposed grid has the highest FPS for video application tested here.

6 Conclusion

In this research work, we propose a hierarchical strategic network model to reduce the cost of computation, latency and energy consumption for IoT/edge- based MANET technology for real-time multimedia (surveillance in tactile networks) and interactive (smarthome) applications. We designed a cluster-based distributed network with IoT-MANET over 802.11 ac to mitigate the problem of computation and delay. This network topology considers the reduction of number of hops and number of messages aggregated at the cluster head by restricting the number of nodes inside the zone. Also to mitigate inter-cluster interference, each zone has been assigned a different channel. Results show that proposed hierarchical strategic network model performs better in transmitting both multimedia (streaming video) and text-based interactive applications for IoT-MANET and, hence, can be deployed in surveillance and smarthome applications.

Acknowledgements This research is supported by Centre for Artificial Intelligence and Robotics (CAIR)-the Defence Research and Development Organization (DRDO), India.

References

1. R. Bruzgiene, L. Narbutaite, T. Adomkus, *MANET Network in Internet of Things System* (2017)
2. G. Carneiro, P. Fortuna, M. Ricardo, Flowmonitor: a network monitoring framework for the network simulator 3 (ns-3), in *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools. VALUETOOLS '09, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)* (Brussels, BEL, 2009). https://doi.org/10.4108/ICST.VALUETOOLS2009.7493
3. H. El-Sayed, S. Sankar, M. Prasad, D. Puthal, A. Gupta, M. Mohanty, C.T. Lin, Edge of things: the big picture on the integration of edge, iot and the cloud in a distributed computing environment. IEEE Access **6**, 1706–1717 (2018)
4. P. Mach, Z. Becvar, Mobile edge computing: a survey on architecture and computation offloading. IEEE Commun. Surv. Tutor. **19**(3), 1628–1656 (2017)
5. R. Patidar, *Validation of wi-fi Network Simulation on ns-3* (University of Washington, Technical report, 2017)
6. R. Rao, G. Kesidis, Purposeful mobility for relaying and surveillance in mobile ad hoc sensor networks. IEEE Trans. Mobile Comput. **3**(3), 225–231 (2004)

Limitation for Single-equation Dependency

Michael Shell, John Doe, and Jane Doe

Abstract Linear regression models are trained with standard errors. We have observed that regression models often give a low accuracy as compared to other techniques. We have evaluated the regression technique and draw a conclusion that these standard errors are the reason for the lower accuracy. The main reason of producing low accuracy is dependency on single-equation denoting relation between the dependent and independent variables. In regression model, on the basis of least square method, the affect of one variable is studied on other variable. However, there comes a time oftenly where the affect of one variable on another drastically changes due to the type of values or range of values present in the data. Such changes in the values might increase the higher accuracy of regression model. So, highly accurate regression model might not work for datasets with such variations. In this paper, we have discussed the issue of single-equation dependency to improve the accuracy. By using the ridge regression technique, we have developed a model that can produce high accuracy for multiple datasets with different variants in the values and the ranges of the values.

Keywords Least square method · Regression model · Ridge regression · Single-equation dependency

1 Introduction

Regression is a statistical procedure that analyzes the relationship between a dependent variable (Y) and one or more independent variables (X_1, X_2...X_k) [1].

M. Shell (✉)
Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
URL: http://www.michaelshell.org

J. Doe · J. Doe
Anonymous University, Atlanta, GA, USA

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_27

If only one independent variable was used, it would be a simple regression analysis, and if more than one independent variable was used, it would be a multiple regression analysis. In order to study the relationship between these variables, it is necessary to establish the functional relationship between them. A first step in determining this possible relationship between the variables is to analyze the graph of observed data. This graph is called a scatter plot and allows you to visually determine whether the variables are related or not. If they are related, the intensity, the sense of the relationship between the variables (direct or inverse) and the type of relationship (linear or nonlinear) existing between them can be intuited.

1.1 Assumptions of the Linear Regression Model

In order to work with a linear regression model, the following assumptions are met [2]:

- The relationship between the variables is linear, and the model is correctly specified.
- The coefficients of the model are constant over time.
- The independent variables are deterministic, and there is no linear relationship between them (absence of multicollinearity).
- The errors are incorrect (there is no autocorrelation), they have constant variance (homocedasticity) and mathematical hope equal to zero, that is, the matrix of variances and covariances of the errors is a scalar matrix where the elements of the main diagonal are the variances of errors.
- The number of observations of the variables must be greater than the number of parameters to be estimated.

1.2 Simple Linear Regression Model

The simple linear regression model is characterized in that to estimate or predict the dependent or endogenous variable, only one independent or exogenous variable is used [3], through the following equation:

$$\gamma_i = \alpha + \beta X_i + \epsilon_i \qquad i = 1...N \tag{1}$$

where N is the number of observations of the variables; the coefficients α and β are the unknown parameters that indicate, respectively, the ordinate in the origin (or estimated value of Y when X = 0) and the slope or coefficient of the regression (or variation the dependent variable before unit variations of the variable independent); ϵ is the random disturbance that includes all those non-observable facts and, therefore, are associated with chance. This disturbance is what gives the model its stochastic character.

The problem of regression consists in estimating values for the unknown parameters α and β from the observations of the variables Y and X in such a way that the equation is completely specified.

Since the observed values of the dependent variable (Y) generally differ from those obtained through the regression line, an error is produced that will be calculated as the difference between the observed and estimated values. These errors are known as waste.

The objective of the regression analysis will be to obtain the estimates of the parameters that make the sum of the squares of the residues minimal. This estimation procedure is generally carried out using the ordinary least squares method (MCO). This term of "least squares" proceeds from the description given by Legendre in 1805 "carrés moindres" and by Gaus in 1809.

Where the estimate of the ordinate at the origin, is the estimate of the regression coefficient, is the sample mean of the independent variable and is the sample mean of the dependent variable.

For example, if we estimate by MCO the relationship between consumption (Y) in thousands of euros and income (X) in thousands of euros received in a given rgion, we could obtain the following results:

$$\hat{Y} = 18.1 + 0.18X \tag{2}$$

$$(3.3) \quad (0.03) \qquad R^2 = 0.84 \tag{3}$$

where in brackets, we have the standard deviations of each estimated parameter (useful especially for making inference); R^2 is the coefficient of determination; autonomous consumption would be 18,100 euros, and the slope that, in this case, is positive and equal to 0.18 would indicate that if the rent is increased by one thousand euros, consumption would only increase by 180 euros.

1.3 Multiple Linear Regression Model

Multiple linear regression is characterized in that to estimate or predict the dependent variable, (Y), two or more independent variables are used, ($X_1 ... X_k$):

$$\gamma_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + \epsilon_i \quad i = 1...N \tag{4}$$

where β_0 is the independent term that each of the coefficients of the independent variables and $\beta_i (i = 1k)$ measures the effect that each of these variables has on the dependent variable.

As in the simple regression, the estimation procedure is ordinary least squares, although it is true that, in this case, increasing the number of variables makes obtaining the estimator the parameters of the model more complicated. The matrix expression of this estimator is as follows:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{5}$$

where $\hat{\beta}$ is the vector of the estimated parameters of the model, X is the matrix of observations of the explanatory variables of the model, X^T its transposed, and Y is the vector of observations of dependent variable.

The estimator properties obtained by ordinary least squares are:

1. It is unbiased, that is, $E(\beta) = \beta$ if it is true that $E(\epsilon) = 0$
2. The variance of the estimator is: $Var(\beta) = \sigma_\epsilon^2 (X^T X)^{-1}$ if it is fulfilled that $Var(\epsilon) = \sigma_\epsilon^2 I$ I, where σ_ϵ^2 is the variance of the term error, e, I is the identity matrix
3. It is consistent
4. The ordinary least squares estimator is the optimal unbiased linear estimator (Gauss Markov theorem).

For example, if we estimate the relationship between consumption (Y), income (X 1) and inflation (X 2) of a giv- en region, we could get the following results:

$$\hat{Y} = 14.3 + 0.15X_1 - 0.07X_2 \tag{6}$$

$$(2.2) \quad (0.02) \quad (0.01) \qquad R^2 = 0.94 \tag{7}$$

where we can appreciate that according to the data obtained from each of the variables, autonomous consumption would be 14,300 euros; if there is an increase in income of one thousand euros (keeping the rest of the variables constant), consumption would increase by 150 euros; an increase in inflation by 1% would imply that consumption would decrease by 700 euros (if the rest of the variables remain constant).

1.4 Nonlinear Regression

Linear regression does not always provide adequate results since, sometimes, the relationship between the dependent variable and the independent variables is not linear. The estimation of the parameters of these nonlinear functions is a more complicated process than in the linear functions, since it can even happen that in the nonlinear models, the number of parameters does not match the number of explanatory variables.

However, there are times when the original variables can be transformed to convert the nonlinear function into linear and, thus, apply linear regression techniques. Therefore, if nonlinearity, for example, affects only the explanatory variables but not the coefficients, then new variables (depending on the nonlinear ones) can be defined to obtain the linear model.

An example of a nonlinear model in the parameters would be the one given by the following exponential function:

$$Y = \alpha X^{\beta} \tag{8}$$

where α and β are unknown constants.

However, this nonlinear model can be transformed into a linear one by applying logarithms on both sides of the equation, as follows:

$$\log(Y) = \log(\alpha) + \beta \log(X) \tag{9}$$

In this model, which is called double log regression, the relationship between the $\log(Y)$ and the $\log(X)$ would be estimated and the interpretation of the parameter β is interesting, since it measures the elasticity of Y with respect to X.

However, other times, nonlinear models cannot be transformed into linear ones. In that case, it would be necessary to apply other methods for the estimation of the parameters such as: nonlinear least squares, maximum likelihood, Monte Carlo simulations.

1.5 Applications

The use of linear regression is very widespread and can be applied to virtually any field that is related, among others, to economics, finance, medicine, biology, physics, engineering, etc.; that is, it allows us to relate variables in a rigorous and quantifiable way in different environments.

1.6 Limitations

One of the main limitations of the regression analysis is based on the fact that two variables grow or decay following the same guidelines does not necessarily imply that one causes the other, since it can happen that a spurious relationship occurs between them. Therefore, to establish the relationship between different variables, it is necessary that this relationship be based on a good theory, since this statistical analysis is adequate to quantify a known relationship between variables, but it is not the best instrument to find functional relationships between variables.

2 Literature Review

When addressing any type of machine learning problem, there are many algorithms to choose from, but there is something that we must be clear about and that is that no algorithm is the best for all problems, and each of them has its own pros and cons, which it does not serve as a guide to select the most suitable.

Although remember that the performance of the different machine learning algorithms depends largely on the size and structure of the data, linear regression defines as train the best line through all data points. This algorithm can be simple linear regression, where the prediction is made with a single variable, multiple linear regression, where a model is created for the relationship between multiple independent input variables and the polynomial regression in which the model becomes a nonlinear combination of the characteristic variables.

- Advantages of linear regression includes easy to understand and explain, which can be very valuable for business decisions.
- It is quick to model and is particularly useful when the relationship to be modeled is not extremely complex and does not have much information.
- It is less prone to overfitting. Along with the advantages, there comes the disadvantages too.
- You cannot model complex relationships.
- You cannot capture nonlinear relationships without transforming the input, so you have to work hard to fit nonlinear functions.
- It can suffer with outliers. Such type of regression technique is very useful for:
- Take a first look at a dataset.
- When you have numerical data with many features.
- Perform econometric prediction.
- Modeling marketing responses.

2.1 Related Work

Regression model is a kind of predictive model where the relationship between dependent and independent variable is established. Here, in this model, the predictor is the independent variable and target is the dependent variable. Regression model is used in time series models. Most of the time to present this relationship, a single equation is used, and we will be discussing limitations in the use of single equations in regressions model.

Multiple methods are used in estimating a single equation, and the one very common is ordinary least squares method which is used for linear regression estimation. Ordinary least squares method uses supervised discrete hashing (SDH) for hashing, where SDH solves the optimization in steps like most critical ones, binary hashing optimization in discrete fashion, discrete cyclic coordinate descent (DCC) used multiple iterations for solving SDH, and all this is a time consuming in relationship to

the fast supervised discrete hashing (FSDH) described in [4]. FSDH used in [4] is faster than traditional SDH in least square regression in solution to projection matrix.

Thanks to its efficacy and completeness, linear least square regression has fulfilled its purpose as the tool of choice for system optimization. Limitations in shapes that linear models can assume over long ranges, possibly poor extrapolation properties and susceptibility to outliers are the principal drawbacks of linear least squares.

In [5], authors have evaluated the traditional regression methods and prompted the cases where regression analysis failed to conclude with the appropriate results. To counter this, they have evaluated the propensity score analysis which produced much better results.

Addressing the problem of single-equation dependency, authors of [6] have proposed an effective piecewise linear regression with novel linear programming formulation. Their proposed methodology can overcome the issues of single-equation dependency. Their proposed methodology consists of dividing a single variable into multiple segments. After dividing it into pieces, their method fits multivariable linear regression function per segment.

3 Proposed Work

To demonstrate the proposed work, we have developed a problem statement and then demonstrate how our proposed solution works fine to elimate the single-equation dependency problem. In general, the proposed methodology can be explained as: We obtain the problem of optimizing the cost function (empirical risk) (we believe that among the signs, there is a constant, and therefore, a free coefficient is not needed)

$$\mathcal{L}(X, \overrightarrow{y}, \overrightarrow{w}) = \frac{1}{2n} \sum_{i=1}^{n} (\overrightarrow{y}_i - \overrightarrow{x}_i^T \overrightarrow{w}_i) \tag{10}$$

$$\frac{1}{2n} |\overrightarrow{y}_i - X\overrightarrow{w}|_2^2 \tag{11}$$

$$\frac{1}{2n} (\overrightarrow{y}_i - X\overrightarrow{w})^T (\overrightarrow{y}_i - X\overrightarrow{w}). \tag{12}$$

If we differentiate this functionality by vector wow, equal to zero and solve the equation, we get an explicit formula for the solution:

$$\overrightarrow{w} = (X^T X)^{-1} X^T \overrightarrow{y} \tag{13}$$

We have tested this proposed work with a Boston Housing Dataset and evaluated the results which are presented in the results section.

Fig. 1 Heat map

4 Experiments and Results

For experiments and evaluation of the proposed technique, we have used Boston Housing Dataset consisting of the prices of houses present at different places in the Boston. Moreover, along with the prices of houses, the dataset also provides a side information about the Crime, INDUS, AGE and other attributes too. The loaded dataset is Boston area estate details

MEDV—Mid value
RM—Room per dwellinh)
TAX—Tax rate per 10K $
INDUS—Proportion of non-retail business acres per town
LSTAT—% lower status of the population
NOX—Nitric oxides concentration

The first step of the experiments is to perform the required preprocessing of the given dataset. Initially as always, we start exploring a new dataset by looking at the graphs. We visualize the correlation matrix and in the form of a heat map presented in Fig. 1. It is a heat map that shows relation and dependency of each independent variable on to dependent variable that the larger the value, the greater the impact of the variable.

Table 1 Initial regression model constructed

Slope	−0.95
Intercept	34.55

Fig. 2 Regression model

We proceed to the construction of a regression model. Define dependent and independent variables. We have presented the slope and the intercept of the constructed regression model in Table 1.

How our model have performed a regression on the given dataset, we need to evaluate the performance. Before that, we have drawn a graphical representation of the regression model which is represented in Fig. 2.

Possible metrics for checking the quality of the model in sklearn can be found in the documentation , or you can calculate by yourself using the formulas. Since there are several independent variables in our model, we cannot display their dependence on two-dimensional space, but we can plot the relationship between the residual model and the predicted values, which will also help us diagnose the quality of the model. This is called a residuals plot. With it, we can see nonlinearity and outliers, andcheck the randomness of the error distribution. Outputs are represented in Table 2.

Regularization is a way to reduce the complexity of a model in order to prevent retraining or correct an incorrectly posed task. This is usually achieved by adding some a priori information to the condition of the problem (Fig. 3).

Table 2 MSE and R^2 values

MSE train	37.934
MSE test	39.817
R^2 train	0.552
R^2 test	0.522

Fig. 3 Unexpected outcomes

In this case, the essence of regularization is that we create a model with all the predictors, and then artificially reduce the size of the coefficients, adding a certain amount to the error.

Ridge regression or ridge regression (ridge regression)—this is one of the methods of reducing the dimension. Often, it is used to combat data redundancy when independent variables correlate with each other (i.e., multicollinearity occurs). The consequence of this is the poor conditioning of the matrix $XT\ X$ and instability of estimates of regression coefficients. Estimates, for example, may have the wrong sign or meanings that far exceed those that are acceptable for physical or practical reasons.

When we do a linear regression, the loss function looked like this:

$$\mathcal{L}(X, \overrightarrow{y}, \overrightarrow{w}) = \frac{1}{2n} \sum_{i=1}^{n} (\overrightarrow{x}_i^T \overrightarrow{w}_i - \overrightarrow{y}_i)^2 \tag{14}$$

Now in ridge regression, we add to it a parameter that will handle the affect of the independent variable on the dependent variable λ which denotes the size of the fine. If $\lambda = 0$, then this is an ordinary linear regression. Now, the formula looks like this:

$$\mathcal{L}(X, \overrightarrow{y}, \overrightarrow{w}) = \frac{1}{2n} \sum_{i=1}^{n} (\overrightarrow{x}_i^T \overrightarrow{w}_i - \overrightarrow{y}_i)^2 + \lambda \sum_{j=1}^{m} |w_j| \tag{15}$$

Table 3 MSE and R^2 values

MSE train	0.449
MSE test	0.472
R^2 train	0.552
R^2 test	0.522

Now, the output after applying the ridge regression becomes as represented in Table 3 with a ridge coefficient [0.75149352]

5 Conclusion

By using the ridge regression technique, we can easily tackle the single-equation dependency for the model and can improve its performance for multiple type of dataset ranging from low to high values and can obtain increase accuracy.

Acknowledgements The authors would like to thank...

References

1. N. Gogtay, S. Deshpande, U. Thatte, J. Assoc. Phys. India. Principles of regression analysis **65**, 48–52 (2017)
2. H.-F. Wang, R.-C. Tsaur, Fuzzy Sets Syst. Insight of a fuzzy regression model **112**(3), 355–369 (2000)
3. W. Donsbach, W. Donsbach, *The international encyclopedia of communication* (Wiley, 2008)
4. J. Gui, T. Liu, Z. Sun, D. Tao, T. Tan, IEEE Trans. Pattern Anal. Mach. Intell. Fast supervised discrete hashing **40**(2), 490–496 (2017)
5. J. Amoah, E.A. Stuart, S.E. Cosgrove, A.D. Harris, J.H. Han, E. Lautenbach, P.D. Tamma, Comparing propensity score methods versus traditional regression analysis for the evaluation of observational data: a case study evaluating the treatment of gramnegative bloodstream infections. Clin. Infect. Dis. (2020)
6. L. Yang, S. Liu, S. Tsoka, L.G. Papageorgiou, Expert Syst. Appl. Mathematical programming for piecewise linear regression analysis **44**, 156–167 (2016)

Autocorrelation of an Econometric Model

Preeti Singh and Sarvpal Singh

Abstract The treatment of an econometric model requires a clearly defined sequence of tasks. The identification of the model leads us to review the literature, to justify the defined relationship between the dependent variable and the independent variables. Model estimation uses the mathematical apparatus to find the equation of fit. Once the model has been estimated, it must be properly diagnosed using statistical tests. After the diagnosis phase, one can use the model to make predictions. This contribution deals with the identification, estimation, diagnosis, and prediction phases for the treatment of econometric models. Likewise, the diagnosis is deepened by developing the problems of autocorrelation, heteroscedasticity, residual normality, multicollinearity, endogeneity, and others.

Keywords Autocorrelation · Econometric model · Data analysis

1 Introduction

This topic addresses the problem of autocorrelation, a characteristic that the term can present error of an econometric model and that manifests itself more frequently when dealing with data observed over time. This type of data, called time series data, consists of observations of variables collected over time (days, months, quarters, years, etc.) of a certain economic unit (company, consumer, region, country, etc.). These data present a natural ordering [1], since the observations are ordered according to the moment of time in which they have been observed.

It is very important to keep in mind that when explaining the behavior of an observed variable over time, one should keep in mind that observations may be correlated throughout it. For example, a family's consumption [2] in one year is likely to be correlated with consumption from the previous year since, on average,

P. Singh (✉) · S. Singh
MMM University of Technology, Gorakhpur, India
e-mail: singh.preeti294@gmail.com

S. Singh
e-mail: spsingh@mmmut.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_28

ceteris paribus, one would expect a family to not to change much their behavior as far as their consumption is concerned from one year to the next. This would possibly involve introducing into the model, to explain the consumption of a year, for example, the consumption of the previous year [3]. Moreover, one can consider that a change in the level of an explanatory variable, for example the income of this family in one year, has effects on consumption in subsequent years. One way to keep account of this would be to introduce, as an explanatory variable of consumption in a year, not only the income of that year, but the income from previous years. Finally, in addition to observable factors [4] whose influence on consumption is distributed throughout the time, there can be unobservable factors, collected in the disturbance term of the model, influence of which is maintained for several periods, presenting what is called autocorrelation or serial correlation over time [5].

As it can be seen, there are different ways of introducing effects into a model that are distributed throughout it. One can introduce delays of both the endogenous variable, for example $Y\ t - 1$, and of certain explanatory variables, $X\ t - j, j = 1 \ldots J$, or to model the error term u t as a function, for example, from its own past $u\ t - j, j = 1 \ldots p$. That is, what is known as a specification model dynamics. In many cases, it is necessary to reflect on this specification and know to which part to assign, systematic or disturbance, it corresponds to. A dynamic bad specification or simply functional of the systematic part of a model can lead to detect autocorrelation in the disturbance term.

In this topic, this contribution will consider how to detect the presence of serial correlation in the error term, analyze what can be the source of this problem, and if it is not a bad specification problem, how to collect this dynamic characteristic in the error term in the model. We will also study the implications of the existence of autocorrelation in alternative estimation methods and in hypothesis tests on the parameters of interest of the model. In a later topic, we will deal with the dynamics in the systematic part.

1.1 Autocorrelation Tests and Residue Analysis

The disturbances are not observable, so the residuals will be our approximation to these variables, both in the graphic analysis and in the statistical contrasts. As in the subject of heteroscedasticity, one can use as a preliminary [6] to a statistical autocorrelation test, the graph of the time series of the residuals, obtained from estimating by OLS the specification of the model of interest. Here, two contrasts designed to test the null hypothesis of the absence of autocorrelation against various alternatives of autocorrelation in error will be considered.

1.2 Durbin and Watson contrast

This contrast is designed to test the null hypothesis of no autocorrelation

$$H0 : \rho = 0, ut = \varepsilon t \quad ^{iid}(0, \sigma 2\varepsilon) \tag{1}$$

versus the alternative of an AR process (1), where the alternative hypothesis can be

$$H\,a : u\,t = \rho u\,t - 1 + \varepsilon t \quad 0 < \rho < 1 \tag{2}$$

or

$$H\,a : u\,t = \rho u\,t - 1 + \varepsilon t \quad -1 < \rho < 0 \tag{3}$$

The procedure is as follows. The interest model for OLS is estimated, and the residuals û t are calculated for $t = 1\ldots$ T. The statistic of Durbin–Watson is calculated as:

$$\mathrm{DW} = \sum Tt = 2(\hat{u}t - \hat{u}t - 1)2 \sum T \tag{4}$$

Thus, the value of the DW statistic is between (0.4). If the DW value is close to zero it indicates that the value of $\hat{\rho}$ is close to 1. If $\hat{\rho} \in (0.1)$, then DW $\simeq \in (2.0)$, so DW values of 2 to 0 will indicate evidence for positive order one autocorrelation. If $\hat{\rho} \in (-1.0)$, then DW $\simeq \in (4.2)$, so DW values from 2 to 4 would indicate evidence toward autocorrelation of order one negative.

Formally, to perform the contrast, one needs to know the distribution of the statistic under the null hypothesis of no autocorrelation, but the distribution of this DW statistic is not the usual one and it is shown that it depends on the matrix X. Durbin and Watson have tabulated some heights, $d\ i$ (lower) and $d\ s$ (upper) of critical values, which allow deciding whether to reject the hypothesis null at a chosen level of significance.

Procedure

Given the sample size T and the number of regressors except the term independent $K' = K - 1$, one should obtain in the table of tabulated critical values the lower bounds $d\ I$ and greater than $d\ s$ for the selected level of significance.

- If DW $\in (d\ s, 4 - d\ s)$, then H 0: $\rho = 0$ is not rejected against either of the two alternatives either Ha: $0 < \rho < 1$ or Ha: $-1 < \rho < 0$. Therefore, there is no evidence of autocorrelation of order one.
- If DW $< d\ I$, $H0$ is rejected: $\rho = 0$ and there is positive one order autocorrelation evidence Ha: $0 < \rho < 1$.
- If DW $> 4 - d\ i$, $H0$ is rejected: $\rho = 0$ versus the autocorrelation alternative of order one negative Ha: $-1 < \rho < 0$.

- If DW $\in (di, ds)$ or if DW $\in (4 - ds, 4 - di)$, the statistic falls into the indeterminacy zones contrast, so the contrast is inconclusive.

Comments

- This contrast can detect other higher-order AR or MA processes, as long as they give rise to residuals with a significant order one autocorrelation. For the same reason, it can detect problems of bad specification of the systematic part of the model. But contrast does not indicate what may be the case.

1.3 Econometric Methods and Data Analysis

- If the autocorrelation of order one is not significant, this contrast may not detect higher-order AR or MA processes.
- In the indeterminacy zones, the contrast is not conclusive. However, one should have noted that the null hypothesis of no autocorrelation is also not accepted, being cautious with the OLS results, especially in the case of inference using the usual t and F statistics.

This region of indeterminacy is greater the lower the number of degrees of freedom is (lower T and higher K).

- The DW statistic is not reliable in models that include as explanatory variables the lagged dependent variable $Yt - 1$, $Yt - 2$, … or another type of stochastic regressor.

2 Literature Review

2.1 Background

It can be said that there is autocorrelation when the error term of an econometric model is correlated with itself over time, such that C o v $(e\ i, e\ j) \neq 0$. This does not mean that the correlation between errors occurs in all periods, but can occur only among some of them. The most common reason is that one or more of the parameters βj are not constant throughout the sample. In the field of economics, it can come from structural changes, special events, etc.

As in the previous case, in the presence of autocorrelation, the OLS estimators continue to be unbiased but do not have a minimum variance, and the generalized minimum square (CGM) estimation method must be used instead. Logically, if the lack of constancy in the value of the parameters is serious, the calculated estimate will be useless.

To address this problem, it is possible to choose to:

- Improve model specification (add, remove, or transform variables) or add terms of type MA to neutralize autocorrelation.

Use consistent estimators for V to $r\ (\beta)$ Var (β). The existence of autocorrelation in the residues is easily identifiable obtaining the functions of autocorrelation (ACF) and partial autocorrelation (ACP) of the least-quadratic errors obtained in the estimation. If these functions correspond to a white noise, the absence of correlation between the residues will be verified. However, the mere visual examination of the previous functions can be confusing and not very objective, so that in the econometric practice, various contrasts are used for autocorrelation, the most often used being that of Durbin–Watson.

2.2 Related Work

Autocorrelation is the presence of linear relationship between disturbances for different observations:

$$\operatorname{cov} X\ (u\ s) = EX(u_t u_s) = 0 \quad \forall t, \quad st = s$$

Autocorrelation is expected when working with series data. The disturbance covariance matrix is:

$$V_X(u) = E_X(uu') = \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \sigma_{23} & \dots & \sigma_{2T} \\ \sigma_{31} & \sigma_{32} & \sigma^2 & \dots & \sigma_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \sigma_{T3} & \dots & \sigma^2 \end{bmatrix} = \Sigma$$

2.3 Consequences of Autocorrelation

- **Properties of the OLS estimators conditioned to *X***
  The OLS estimator is linear, unbiased, but it is NOT efficient, in the sense of variance being minimal, because the Gauss–Markov theorem is not fulfilled. There is another estimator better than the OLS, that is, with less variance [7]: the least squares estimator.
- **Inference with the OLS estimator**
  When assumption $S.5$ is not fulfilled and there is autocorrelation, the true matrix of covariances of the OLS estimator is:

Fig. 1 Positive autocorrelation

Fig. 2 Negative autocorrelation

$$V_X(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

Autocorrelation Detection

- Graphic analysis
  The following graph visualizes the temporal evolution of the disturbances when there are:
- Positive autocorrelation: long streaks of positive and negative shocks [8].
- Negative autocorrelation: disturbances alternate positive and negative sign (Figs. 1 and 2).

3 Proposed Work

The hypothesis of non-autocorrelation (one of the basic hypotheses of the RLM) means that the terms of disturbance are incorrectly interrelated. Moreover, if the disturbance term follows a normal distribution, the hypothesis of non-autocorrelation implies independence between the terms of disturbance; that is, the residue of any observation is not correlated or the residue of another observation. This implies that the covariance between any pair of disturbance terms is zero.

The above is translated into the fact that the elements outside the main diagonal of the variance and covariance matrices of the disturbance term are all equal to zero. Under the assumption of autocorrelation, there will be at least a couple of disturbance terms [9], for which it will not be true that the covariance between these is zero.

In short, it can be concluded that if the error for one observation is positive, then it will be systematically increased in the next observation in a time-based analysis, and hence, we will have a positive error correlation.

By extracting the error factor within the model for an observation, a variable must be incorporated in the model that will capture the autocorrelation error and will nullify its impact on the observation and hence will decrease the error within the model.

A graph of the autocorrelation of a time series by delay is called a correlogram, or an autocorrelation diagram, or autocorrelation function plot (ACF). When executing the example, a 2D graph is created that shows the delay value along the X-axis (squared waste) and the Y-axis [3] correlation between -1 and 1. These indirect correlations are the ones the partial autocorrelation function seeks to eliminate. Without going into mathematics, this is the intuition for partial autocorrelation.

3.1 Ljung–Box Test

The Ljung–Box test refers to a statistical test to determine whether an autocorrelation group taking in a time series is nonzero [1]. Instead of testing randomness at each different delay, this tests generic randomness.

- Null hypothesis: Absent autocorrelation
- Alternative hypothesis: Autocorrelation present.

Given the p-value, we cannot reject the null hypothesis of the absence of autocorrelation.

3.2 Durbin–Watson Statistic

The Python statsmodel incorporates the Durbin–Watson statistic by default. The Durbin–Watson test [10] informs us with a value from 0 to 4, of possible the presence of autocorrelation Ar (1) (first-order autocorrelation, when consecutive residues are correlated), where

- Autocorrelation is null at 2
- Positive autocorrelation ranges from 0 to 2
- Negative autocorrelation ranges from 2 to 4

Values in the range of 1.5–2.5 can be considered relatively normal. Values obtained outside of this range may be a cause for concern. Whereas, values below one or more than 3 need to be addressed on immediate basis.

4 Concluding Remarks

As seen throughout this chapter, the presence of structural changes in series of time-stamped data significantly affects analysis or relationships between independent and dependent variables. In this work, we have discussed the representation of time series with changes structural, of the existing tests for the both simple and multiple changes and to calculate the number of changes. Equally, the implications for series of nonlinear time were considered. Here, it has been discussed how the presence of structural changes affects the development of unit root contrasts, long memory, and prognosis. The predominant conclusion that emerges from this work is that structural changes have a marked effect on time series modeling, which encourages its analysis. A study of the chronological evolution of research on structural changes in time series shows that approximately 83 articles on the subject were published. In the analysis of the literature presented here, a cursory study of some of the important issues in time series analyzes structural changes. However, in this process the wealth and diversity

of the subject analyzed and the degree of growth of investigations in the area was acknowledged. While the research on changes structural has had a marked growth, there are still fields of research that require further study, such as the development of formal procedures for specification and model building for time series no linear with structural changes, where both characteristics are considered simultaneously and do not require a strong prior knowledge of the series; developing a set of tests sufficient to allow identification, estimation and representation of series with the characteristics cited; building models oriented to forecast; and the characterization of different types of structural changes.

Acknowledgements Mentorship of Prof. Marcin Paprzycki, Galgotias University, Greater Noida, India, is acknowledged and appreciated.

References

1. N.S. Balke, Detecting level shifts in time series. J. Bus. Econ. Stat. **11**(1), 81–92 (1993)
2. P. Perron, The great crash, the oil price shock, and the unit root hypothesis. Econometrica J. Econometric Soc., 1361–1401 (1989)
3. R.L. Lumsdaine, D.H. Papell, Multiple trend breaks and the unit-root hypothesis. Rev. Econ. Stat. **79**(2), 212–218 (1997)
4. L.C. Nunes, P. Newbold, C.-M. Kuan, Spurious number of breaks. Econ. Lett. **50**(2), 175–178 (1996)
5. G.C. Chow, Tests of equality between sets of coefficients in two linear regressions. Econometrica J. Econometric Soc., 591–605 (1960)
6. R. Quandt, Tests of the hypothesis that a linear regression obeys two separate regimes. J. Am. Stat. Assoc. **55** (1960)
7. D.W. Andrews, Tests for parameter instability and structural change with unknown change point. Econometrica J. Econometric Soc., 821–856 (1993)
8. A. Banerjee, R.L. Lumsdaine, J.H. Stock, Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence. J. Bus. Econ. Stat. **10**(3), 271–287 (1992)
9. J. Bai, Estimation of a change point in multiple regression models. Rev. Econ. Stat. **79**(4), 551–563 (1997)
10. C.W. Granger, N. Hyung, Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. J. Empir. Financ. **11**(3), 399–421 (2004)

Trends of Artificial Intelligence in Revenue Management of Hotels

Smrutirekha, Priti Ranjan Sahoo, and Ashtha Karki

Abstract The authors provide an in-depth summary of artificial intelligence's role in the revenue management of hotels across the globe. Focusing on artificial intelligence, the chapter draws the reader's attention toward its usage and role in the lives of the customers and service providers in the hotel industry, which ultimately leads to the revenue management system. The chapter focuses on the current trends of artificial intelligence to generate revenue for hotels more innovatively. Some of the finest and latent features of artificial intelligence in the revenue management system have been discussed in the chapter by elucidating few examples of few renowned hotels in the world. The revenue management system of the hotel industry has been discussed briefly in the chapter. The essence of the chapter lies in the role and importance of the ever-changing and emerging trends of artificial intelligence in the hotel industry, particularly for revenue generation. Though artificial intelligence cannot replace the emotional intelligence provided by humans, it is the need of the hour to enhance and sustain in the innovative and digital era. AI rules across manufacturing and production sectors, but it also plays a strategic role in the service industry.

Keywords Artificial intelligence · Hotels · Revenue management

1 Introduction

Artificial intelligence (AI) is spreading like wildfire across manufacturing, processing, and services in this digital era. The AI-based technologies are intelligent enough to replace human activities completely. The speed and accuracy of the process also increase while using AI. AI has a positive impact to a greater extent, but at the same time, it has a low value for the human touch, particularly in the service industry. AI technology used in hotels is not a far-fetched imagination anymore.

Smrutirekha (✉) · P. R. Sahoo · A. Karki
KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India
e-mail: smrutirekha195@gmail.com

P. R. Sahoo
e-mail: prsahoo@ksom.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_29

Several parts of the world's most well-known businesses are already implementing so [22]. In hotels, artificial intelligence (AI) encompasses everything from robots to intelligent supercomputers. The fundamental reason why a revenue management system is imperative for those within the hotel industry, notwithstanding the measure of their lodging, is since it permits complex calculations to be carried out rapidly and permits for the kind of real-time tracking of advertising information that's outlandish to duplicate physically.

It has instead become imperative to sustain in the world of business. The immediate strategic benefits of incorporating AI in hotels for revenue management come from cost reduction, efficiency in its operations, and increased revenue [22]. The gray areas like people management and social responsibility are also a part of the evolving technology of AI in the hotel industry. The ability for a hotel to make smart estimating decisions in a computerized mold makes the business case for investing in an AI-powered revenue management system persuasive. In terms of boosting productivity, it is a compelling argument. It is also effective in avoiding potential revenue losses that can occur when a hotel falls short of maximizing occupancy or, even worse, experiences a decrease in occupancy. With the research from various literature sources on AI-based technologies of the world, this chapter is based on the discussion of the roles and trends of AI in the hotel industry for revenue management purposes, along with an overview of the revenue management systems and tools.

2 Trends of Artificial Intelligence in the Hotel Industry

A thorough grasp of the technology, its properties, and applications is required to facilitate AI adoption and implementation in the real world. The enormous amount of data that our world offers every moment needs to be catered to enhance the services. Through the acquisition and utilization of data, hospitality firms are rapidly investing in AI-powered platforms, which are becoming vital for providing elevated, authentic, and personalized experiences to customers [2]. The current trends of artificial intelligence in the hotel industry are fascinating enough to draw the attention of both the guests and the manager. Few major and critical trends of AI in the hotel industry are given below.

- **Emotional AI**—It can read both vocal and nonverbal clues to comprehend customer behavior. By researching how people react to different materials, products, and services, one can quickly detect numerous forms of human emotions.
- **Augmented Reality and Virtual Reality**—It is related to experiences that AI can provide to the consumers, which can transform the view of the consumers in terms of personalization and customization [20].
- **AI used in Robots**—AI has made robots more innovative and efficient than ever. Robots can perform most human activities, giving an advantage over human labor [9].

- **AI in the Internet of Things**—Smart and wearable devices helps ease the service provider's process. It is an innovative and efficient way of accessing the required data. Analytical decisions can be made using the data without any human intervention [16, 17].
- **Predictive Analytics**—Provides a better understanding of the customer's needs, wants, and demands. Accordingly, the services can be tailored to meet their expectations [5].
- **Ethical AI**—Ethics and fair play are critical aspects of AI in the hotel industry. Business leaders advocate using ethical AI to maintain transparency and work as per the business requirements [6].
- **AI in Cybersecurity**—Cybersecurity has always been a significant concern, now one of the most enhanced prefaces in the digital world.
- **AI in Computer Vision**—Postures, movements, gestures, facial recognition (of the guests/consumers), etc., are captured and analyzed for understanding human intelligence for improvisation of the services and management.

3 Overview of Revenue Management of Hotels

Revenue management can be defined as increasing the revenue of a business by matching the supply and demand by selling the fitting room to the designated customer at the right time at a perfect price through the best available distribution channel, maintaining the best cost-efficiency. Data analytics is used behind the scenes to assess consumer behavior and forecast demand.

The use of revenue management in the hotel industry has expanded dramatically in recent years. It began in the room sections and has since expanded to include meeting and event facilities and food and beverage shops. At the same time, innovative enhancements empower the improvement of unused merchandise, such as mechanized income administration frameworks. On the one hand, there are opportunities to misuse huge information and move to a science-based income administration framework. Risks, on the other side, include greater susceptibility as a result of hacking or data leaks. In any case, the industry is gradually moving toward computerization and will need to adjust over time.

When a customer submits a booking request, the hotel's RM system records it. The latter comprises four structural elements: data and information, hotel revenue centers, revenue management software and tools, and the revenue management team. The specific booking aspects of the given booking request are the operational results of the RM process—booking status, number of rooms, types, categories of rooms, duration of stay, price, cancelation, and revision terms and conditions. The customer's opinions of the fairness of the hotel's RM system and their intentions for future bookings with the same hotel/hotel chain are influenced by the booking information and the operation of the entire RM system [8]. The RM system is constantly affected by external (macro- and micro-) and internal environmental factors in which the hotel operates (the company's goals, financial situation, legislation, competition, changes

Fig. 1 *Source* Adapted from [12]

in demand, destination's image), and revenue managers' decisions must take all of these factors into account.

As stated by Ivanov and Zhechev [14], the revenue management system of hotels has been given in Fig. 1.

3.1 Key Performance Indicators of Revenue Management System in Hotels

The KPIs help track and maintain an effective revenue management system (RMS). Numerous technology solutions are available for their measurement. The major KPIs include.

- GOPPAR—Gross operating profit per available room—considers total available rooms despite the occupancy status.
- RevPAR—Total revenue per available room—takes into account all revenue from rooms, including room service, meals, and other amenities.
- ARPA—Average revenue per account—average amount of revenue generated per customer account.
- EBITDA—Earnings before interest, taxes, depreciation, and amortization—total revenue minus the expenses (which exclude ITDA).

3.2 Important Revenue Management Tools

- ***Property Management System (PMS) Management System*** (***PMS***)—From the front desk (check-in/check-out, room assignment) to finance and revenue management, the PMS oversees all areas of the hotel's operations (managing room rates, billing). Examples of popular PMS—Apaleo, Cloudbeds, and Mindmatrix
- ***Channel Manager—***Helps connect and manage online distribution channels for hotels to sell their inventory to OTAs and agents globally. Examples—Siteminder, Lodgable
- ***Business Intelligence and Data Analytics***—Business intelligence (BI) platforms are technologies that help collect, integrate, analyze, and present hotel data in the reports. Helps gain specific insights into the current scenario in business. Examples—OTA Insight, Power BI, Looker
- ***Upselling tools—***Upselling is one of the most effective strategies to boost revenue and ADR while maintaining the same level of demand. It is one of the most critical pricing strategies used in the revenue management system. Examples—Oaky, Revinate Upsell

4 Major Roles Played by Artificial Intelligence for Revenue Management of Hotels

As studied by Haddad et al. [10], revenue management (RM) is an essential tool for matching supply and demand by segmenting consumers based on their purchase intents and allocating capacity to the various groups in a way that maximizes a company's revenues. In today's digital business industry, AI has a crucial and intricate role to play in carrying out profitable yet sustainable business.

AI-assisted revenue management is all about intelligent estimation. It is all about maximizing room occupancy at the lowest possible cost by combining request figures, competitor rates, and cost sensitivities—all while taking into account a variety of additional factors, including request drivers like regularity, special event dates, and day-of-week differences. In addition to the type of room, the length of stay, and the degree to which a lower cost advancement may weaken income and benefits, in the long run, skillful estimating entails taking into account other factors like the type of room, the length of stay, and the degree to which a lower cost advancement may weaken income and benefits in the long run [24]. The combinatorial difficulties of clever estimating are not to be trifled. The hotel industry is one of its kind where AI is gradually overpowering human activities or legacy systems.

By building up manufactured insights in the hotel industry, inns can make more critical openings to provide excellent guest-friendly administrations extending from coordinating visitor inclinations, recommending books or music, adjacent sports club to complement customers' taste, all the way to consequently alarming inn staff for

personalized supper choices, extraordinary benefits, and complimentary administrations, etc. Through interconnected gadgets, sensors, and machine learning, hotel operations framework can associate superbly with the physical world, engaging visitors with a profoundly personalized client involvement. Hotel rooms can use the existing arrangement of technological developments and virtual collaborators to advance and improve the encounter to the following level.

A few of the most critical roles AI play for revenue management in hotels have been stated below.

- Optimization of booking technologies used to book the rooms online. It enhances the preface between the customer and the hotels during the entire process of booking a room or any service online [1]. The software system of the hotel is optimized using artificial intelligence. This, in turn, improves the cost and finance management of the hotels' products and services. Group booking software is a significant development in the hotel reservation system, which has made the task easier. The use of AI technology is also upgrading cancelation policies.
- Social media presence is one of the most crucial effects of using AI-backed technologies in today's world, where a customer is highly dependent on social media to create an impression and image of the service before buying it [4]. The decision-making part heavily depends on the hotels' level of presence and activity on the social media platforms. This helps the customer assess the hotels' quality of services and facilities.
- AI-backed concierges where the guests are offered services of automated check-in and check-out using bots. Room service facilities, mobile keys, and intelligent rooms are essential facilities AI provides through mobile phones [21]. Through AI technologies, the service providers can create consistent guest experiences and provide them with the best of services, eventually leading to an increase in revenue.
- Automated services within the hotel industry have proved to be of advantage in terms of human labor. The chatbots, widgets, and robots used within the premises of the hotels for maintaining a soft touch servicescape, automated food and beverage services with the option for customization, augmented reality, virtual reality tours, etc., have brought about a revolution in the hotel industry [7].
- The after-sales service is gaining momentum in the hotel industry, where the feedback and review of the guests are of utmost importance for revenue management. This feature incorporates the predictive analysis feature of AI-based technologies. The scope for improvement is highlighted by the feedback, rating, and reviewing system of a hotel's software. The artificial neural network comes into play to assess the demands, needs, and expectations of the consumer, which directs toward pricing and revenue management [18].

A list of the segregated AI functions as studied by Ivanov and Webster [13].

- Website automation—XML links to suppliers, white labels, chatbots
- Virtual reality
- Booking engine

- Metasearch aggregator
- Own aggregator platforms
- Online travel aggregators
- Dynamic packaging software
- Digital concierge
- Booking–as-a-service software
- Mobile applications for feedback
- Self-service kiosks
- Augmented reality
- Virtual guide
- QR codes
- Virtual personal assistant software
- Personalized intelligent solutions
- Document issuance software
- Payment transactions applications
- Management of information software
- Integration of all suppliers CRS data through XML links and White-label
- Inventory management software
- Intelligent retargeting and bid management
- User profiling
- Tracking software
- Data management platforms
- Navigation.

5 Discussion and Conclusion

In most cases, a great RMS will utilize the information and claim computations to conduct a real-time assessment of the status of the system and request accordingly to calculate perfect room costs. As a result, the more significant part of fundamental lodging income administration decisions can be made from a single, centralized dashboard.

The futuristic view of the RMS might direct in the discussed direction. RMS, with time, will become much more powerful and strategic. Social media and data analytics will significantly impact and would technologically drive AI [3]. RMS will be more centralized, which would help in gross resource management. Strong communication and organizational behavior will be seen in the future of AI-based RMS. The functional spaces would incorporate revenue management; its measurement would be innovative in nature, and strategies would be more refined and streamlined for better financial decisions [11]. The hotel industry and the customers will grow more aware of RM restoration procedures, making the customer experience an additional strategic factor for the sector [9].

In the hospitality industry, revenue management is a crucial job. Even though AI is still considered an emerging technology trend, several hotels have already reaped the benefits and significantly influenced their bottom line [19].

While present systems rely primarily on previously acquired data (big data) to extract any implications and recommendations to tourists, predictive analytics solutions, aimed at anticipating customer behavior and projecting industry success, will make significant development in the future [23]. Intelligent systems will estimate, forecast, plan, and control activities across channels and suppliers, maximizing resource efficiency, revenue growth, and customer satisfaction [15]. AIs will be the primary instrument for enhancing operations, redesigning, and reengineering business processes with the ultimate goal of putting the customer first while still maximizing revenues and costs.

Robots as self-contained gadgets would not replace intermediaries' communication and coordination but improve hotel industry information management [26].

Traditional legacy system has always prevailed in the revenue management cycle in the hotel industry. But with the evolving landscape of AI technology, the revenue management system has experienced a hurricane. A comparative analysis between the legacy system and the AI-based RMS clarifies the improvisations achieved in Fig. 2.

Mechanical propels, including frameworks integration, will progress the capability to recognize the esteem of each client, expanding the viability and making a difference to estimate requests precisely, boosting development buy reservations. This permits companies to centralize their income divisions, as one director presently can direct more than one property.

Fig. 2 Comparison between legacy system and AI-based RMS

The last decade's technological improvements have had a massive impact on the hospitality business, tourist behavior, and hotel operations. Most lodgings and accommodations have adjusted this current estimating frame after income administration was introduced to the business. While having a revenue management system for hotel rooms is the norm, the industry is increasingly moving to support its implementation in conference and convention venues and food and beverage outlets.

Artificial intelligence-based technologies are rapidly evolving, presenting modern opportunities such as using vast amounts of data. Still, they also provide concerns, such as the worry of current workers losing their jobs. Social networking is a fantastic way for businesses to connect with customers and get information about them. This can be used for short- and long-term planning and a new way to provide rooms. While entire computerization is not yet possible, a beautifully designed revenue management computer software now supports the revenue chief in analyzing and estimating and strategy-oriented intelligent decision-making. The commerce insights gathered from the detailing capabilities, for illustration, can offer assistance progress deals adequacy, create competitive insights, and give good bits of knowledge into inhabitance patterns, visitor socioeconomics, showcase situating, and channel benefit. A promoting office can utilize the estimates as a director for deciding when to extend special spending to a reasonable request. An operations team can know when to develop (or diminish) staffing based on anticipated inhabitance. In brief, the benefits tend to go well past the office known as revenue management, eventually rising above all parts of the organization.

Hotel visitors are ultimately helped from take-off to goal with personalized suggestions that upgrade their travel and after-arrival choices at the lodging (eating, drinks, breakfasts, on-property exercises, etc.). After gathering loads of client information, hotels can utilize machine learning with business-specific calculations to create expectations of what clients are most likely to select. Such expository capacity of AI can keep the inns spry in recognizing the following set of patterns regarding items or benefits offerings. It can be utilized to assess and acknowledge the traveler or client persona to coordinate them with essential administrations [25]. Using the brilliance of proactive analytics and identity branding, hotels can recognize long-term needs. It would be ideal if the clients can be satisfied by the services they have been searching for some time recently and which they indeed inquired for.

Hotel marketing professionals who anticipate guest emotions and motivations ahead of time are more likely to acquire traction in their business growth than those that use a more traditional strategy. To determine the impact of various services on clients, hotel operators must work out a variety of emotions and develop a comprehensive approach.

Hotel marketing leaders will adapt to creative tactics to strengthen their marketing messages as AI improves the means of collecting client information. Even for newcomers to the hospitality industry, the trend of integrating AI-powered predictive analytics in hotels for marketing purposes can be beneficial.

Artificial intelligence has entered the traditional hospitality landscape with the promise of increasing revenue, improving hotel reputation, and elevating the

customer experience—all of which legacy systems were incapable of doing efficiently. But still, AI technology's implementation in revenue management system (RMS) with more sophisticated pricing discrimination and real-time demand adaptability is in its early phases.

References

1. N. Antonio, A. de Almeida, L. Nunes, Big data in hotel revenue management: exploring cancellation drivers to gain insights into booking cancellation behavior. Cornell Hospitality Q. **60**(4), 298–319 (2019). https://doi.org/10.1177/1938965519851466
2. M. Bounatirou, A. Lim, A case study on the impact of artificial intelligence on a hospitality company. Adv. Ser. Manag. **24**, 179–187 (2020). https://doi.org/10.1108/S1877-636120200000024013
3. E. Brynjolfsson, A. McAfee, Will humans go the way of horses. Foreign Aff. **94**, 8 (2015)
4. E. Çeltek, I. Ilhan, *Big Data Artificial Intelligence, and Their Implications in the Tourism Industry* (2020), pp. 115–130. https://doi.org/10.4018/978-1-7998-1989-9.ch006
5. D.J. Connolly, R.G. Moore, Decision making in the context of hotel global distribution systems: a multiple case study. *Hospitality* (1999)
6. T.H. Davenport, *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. MIT Press (2018)
7. S.J. DeCanio, Robots and humans–complements or substitutes? J. Macroecon. **49**, 280–291 (2016)
8. N.F. El Gayar, M. Saleh, A. Atiya, H. El-Shishiny, A.A.Y. Fayez Zakhary, H.A.A. Mohammed Habib, An integrated framework for advanced hotel revenue management. Int. J. Contemp. Hosp. Manag. **23**(1), 84–98 (2011). https://doi.org/10.1108/09596111111101689
9. L. González-Serrano, P. Talón-Ballestero, Revenue management and E-tourism: the past present and future. Handb. E-Tourism, 1–28 (2020). https://doi.org/10.1007/978-3-030-05324-6_76-1
10. R. Haddad, A. El, Roper, P. Jones, The impact of revenue management decisions on customers' attitudes and behaviours: a case study of a leading UK budget hotel chain. Angew. Chem. Int. Ed. **6**(11), 951–952, 5–24 (2008)
11. S.H. Ivanov, Tourism beyond humans–robots, pets and Teddy bears, in *Paper to Be Presented at the International Scientific Conference Tourism and Innovations* (2018)
12. S. Ivanov, V. Zhechev, Hotel marketing. Varna Zangador (2011)
13. S. Ivanov, C. Webster, Economic fundamentals of the use of robots, artificial intelligence, and service automation in travel, tourism, and hospitality, in *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and Hospitality* (Emerald Publishing Limited 2019)
14. S. Ivanov, V. Zhechev, Hotel revenue management–a critical literature review. Tourism Int. Interdisc. J.**60**(2), 175–197 (2012)
15. M. Ivanova, Robots, Artificial intelligence, and service automation in travel agencies and tourist information centers, in *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and Hospitality* (Emerald Publishing Limited 2019)
16. R.S. Jha, P.R. Sahoo, Internet of Things (Iot)—Enabler for connecting world, in *ICT for Competitive Strategies* (2020), pp. 1–7. https://doi.org/10.1201/9781003052098-1
17. R.S. Jha, P.R. Sahoo, Influence of big data capabilities in knowledge management—MSMEs, in *ICT Systems and Sustainability* (Springer 2021), pp. 513–524
18. J. Kim, S. Wei, H. Ruys, Segmenting the market of West Australian senior tourists using an artificial neural network. Tour. Manage. **24**(1), 25–34 (2003)
19. S.E. Kimes, The future of hotel revenue management. J. Revenue Pricing Manag. **10**(1), 62–72 (2011). https://doi.org/10.1057/rpm.2010.47

20. S.E. Kimes, J. Ho, Revenue management in luxury hotels. J. Revenue Pricing Manag. **17**(4), 291–295 (2018). https://doi.org/10.1057/s41272-017-0113-1
21. V. Liljander, F. Gillberg, J. Gummerus, A. Van Riel, Technology readiness and the evaluation and adoption of self-service technologies. J. Retail. Consum. Serv. **13**(3), 177–191 (2006)
22. K. Nam, C.S. Dutt, P. Chathoth, A. Daghfous, M.S. Khan, The adoption of artificial intelligence and robotics in the hotel industry: prospects and challenges. Electron. Mark. (2020). https://doi.org/10.1007/s12525-020-00442-3
23. A. Parasuraman, C.L. Colby, An updated and streamlined technology readiness index: TRI 2.0. J. Serv. Res.**18**(1), 59–74 (2015)
24. P.R. Sahoo, S.K. Lenka, B.B. Pradhan, Transactional and psychological concerns of all parties involved in opaque hotel room distribution. Int. J. Econ. Res. **14**(8), 1–9 (2017)
25. P.R. Sahoo, B.B. Pradhan, S.K. Lenka, S.K. Patra, Review research on application of information and communication technology in tourism and hospitality industry. Int. J. Appl. Bus. Econ. Res. **15**(11), 311–334 (2017)
26. A.P. Srivastava, R.L. Dhar, Technology leadership and predicting travel agent performance. Tourism Manag. Perspect. **20**, 77–86 (2016)

Sentiment Analysis and Vector Embedding: A Comparative Study

Shila Jawale and S. D. Sawarkar

Abstract Automatic intent/sentiment classification can be done with various machine learning approaches as well as methods. But the success of these techniques majorly depends on the representation of words or documents in vector space. That can be easily consumed by machine toward learning the hidden pattern of text/corpus. To achieve this, various methods have been proposed, and many are commercially accepted as well. In deep learning architecture for intent/sentiment analysis, the vector embedding plays a crucial role. It represents feature extraction. In this paper, various methods for vector embedding are discussed along with their comparison.

Keywords Sentiment analysis · Glove · FastText · BERT · Vector embedding

1 Introduction

The use of the Internet is growing day by day resulting in large amounts of data. The huge data getting occurs in different online applications too. It may be in the form of various types of reviews, opinions generated with multiple public/social media applications such as Amazon and Twitter. [1] To find good insights from this data to make it as information, the processing is a must. To accomplish this process in finding good insights and make use of it for finding the effectiveness of those opinions or intent for further decision-making.

Sentiment analysis as the name reflects to identify the opinion or intent or attitude or consent toward things. The sentiment is generally called opinion mining or review analysis. It is used to identify, study, quantify, obtain, tacit states, and subject-related information [2]. It plays a very important role in varied kinds of decision-making. Sentiment analysis is very well implemented in various kinds of domain such as financial sector [3], e-commerce domain [4], educational domain [5–7], clinical domain [8], disaster management [9], smart city [10], and many more [11–13].

S. Jawale (✉) · S. D. Sawarkar
Datta Meghe College of Engineering, Airoli, Navi Mumbai, Maharashtra, India
e-mail: shila.jawale@dmce.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_30

There are three forms of sentiment analysis: granularity-oriented, task-related, and methodologically oriented. Polarity categorization, level of valence or arousal at specific/beyond polarity, subjectivity/objectivity identification, and feature/aspect-based sentiment analysis are some of the subcategories of task-oriented sentiment analysis. Document, phrase, and word levels, as well as character levels, are all granularity-oriented. Method-based is divided into three categories: supervised learning, unsupervised learning, and semi-supervised learning. Sentiment categorization based on features found in the text. Words, their occurrence, grammatical representation and categorization, the presence of sentiment words, phrases in sentences, language norms, sentiment shifters, and syntactic dependency seem to be instances of these features [14].

To have effective sentiment analysis, many techniques have been applied. Some are based on lexical and some are based on machine learning approaches. Machine learning has its subset as deep learning. Nowadays, it is becoming a popular way to handle massive text data. Due to the larger dimension of text data, deep learning becomes an effective technique to get decision-making accurate with better accuracy.

Deep learning is a multilayer architecture. Each layer is responsible for completing the dedicated task. The different layers such as embedding layer, convolution layer, sequential layer, and bidirectional layer. After defining the model, the very first layer is the embedding layer. This is an important layer in terms of vector representation of the available text. Different approaches till now are available in the literature, and still very few are applicable in the production environment. The performance of deep learning is heavily depending on kinds of word embedding is applied as feature extraction is done by model itself [15].

The paper gets divided into multiple Sect., viz. 2 Embedding method for Sentiment Analysis, Sect. 3. Related Work, Sect. 4. Proposed Model, Sect. 5. Result and discussion followed by Conclusion.

2 Embedding Method for Sentiment Analysis

Various strategies have been used to achieve the automatic sentiment analysis. The computational methods are shown in Fig. 1. The general sentiment analysis process shown in Fig. 2.

There are various methods to handle vector embedding in other words vector conversion to represent each word or entire document with a multiple or single numerical representation. Sometimes, it is just represented in binary or in decimal format. The various vector computational ways in state-of-literature as bag-of-word, TF-IDF, Glove, FastText, EMLO, pre-trained model BERT, and their variations.

Bag-of-words: The bag-of-words (BOW) [16] model is an oldest and very common method used for text representation. The words are represented with a particular number and their occurrences. At best, the BOW is effective for topic-based text classification rather than sentiment analysis. By disrupting word order as well as discarding contextual information, the BOW loses performance (which is

Fig. 1 Computational methods process

Fig. 2 General sentiment analysis process

critical for accurate sentiment assessment). The BOW model struggles with polarity shifts like negation.

Term Frequency: TF-IDF is the product of two statistical terms, as shown in Eq. (1). The first is the term frequency (TF), and the second is the inverse document frequency (IDF) [17]

$$W_{t,d} = \mathrm{tf}_{t,d} - \mathrm{idf}_t \tag{1}$$

where $W_{t,d}$ is the weight for feature t (i.e., a word) in document d, and $\mathrm{tf}_{t,d}$, and idf_t are defined in Eqs. (2) and (3), respectively [4]:

$$\mathrm{tf}_{t,d} = \frac{\text{Frequency of term}}{\text{Total number of term in documents}} \tag{2}$$

$$\text{idf} = \log \frac{N}{\text{df}_t} \tag{3}$$

where $\text{tf}_{t,d}$ is the term frequency, which calculates the number of times term occurs in a document, and $\log(N/\text{df}_t)$ is the inverse document frequency, obtained by dividing the total number of documents, N, by the number of documents containing that term, df_t and using the log of that value.

Word2Vec: [18] Due to the lack of keeping the context in TF-IDF model, another model based on the idea of representing the word with pre-trained vocabulary which was computed based on two model such as Skip-gram model and CBOW.

Skip-Gram Model (SG): The idea of SG, in each estimation step you are taking one word as a center word, and you are going to try and predict words in their context out to some window size, and the model is going to define a probability distribution that is the probability of a word appearing in the context given this center word. And we are going to choose vector representation of words so we can try and maximize that probability distribution (we just have one probability distribution of the context word).

Continuous Bag-of-Words Model (CBOW): In this model, the target word (center word) is predicted based on the surrounding words. The context is represented by multiple words, and to predict the missing target word, the sequence of context words must be given, then try to predict the missing word in the context according to the window size. This can be applied to different sizes of windows (10, 15, 20, 100, etc.). According to Nandakumar and Salehi [19], CBOW works several times faster than SG to train and has slightly better accuracy for frequent words, while SG works well with a small amount of training data, and represents well even rare words or phrases. Figure 3 shows the architecture of CBOW and SD algorithms. Figure 3 shows the model [20].

Fig. 3 CBoW and Skip-gram model

Glove: It is a word representation model. Glove, for global vectors, because the global corpus statistics are captured directly by the model [21]. Glove is the most popular count-based method in word embedding models which work upon dimension reduction in the global co-occurrence matrix of words. The global co-occurrence matrix contains statistical amounts of co-occurring events of every word couple in different contexts. This matrix is then decomposed to a much smaller matrix that contains the embedding features for each word. The process is done by matrix factorization and the minimization of a loss function. The continuous bag-of-words (CBoW) model [17] and the Skip-gram model [18] are both included in Word2Vec, a deep neural network that learns word embeddings from text. The CBoW predicts the target word (e.g., "the girl is a television," where " " signifies the target word) given a collection of context words, whereas the Skip-gram model predicts the related terms given the target word [19, 20].

FastText and ELMO: Character-based model [19] is used for creating vector. The author instead of using words made the model convert embedding the character data. Due to this, out of vocabulary words also got treated very well. The model considers sentence structure and arrangement by considering sub-word units and representing words by a sum of their character n-grams. Begin by presenting the general framework to train word vectors, then present the sub-word model and eventually describe dictionary of character *n*-grams [22].

ELMO: Another character embedding model, the vector for a word is constructed from the character n-grams that compose it. Since character n-grams are shared across words, assuming a closed-world alphabet, these models can generate embeddings for OOV words, as well as words that occur infrequently. EMLO can handle a novel sort of deep contextually relevant word representation that reflects complex aspects of word use (e.g., structure, syntax, and semantics), and how these uses vary across linguistic contexts (i.e., polysemy-words with similar pronunciation). The learnt functions of the internal states of a deep bidirectional language model (Bi-LSTM) that was pre-trained on a large text corpus were represented by the word vectors [23].

Pre-trained word vector Model- BERT: The BERT (bidirectional encoder representations from transformers) model was proposed by author in [21]. This pre-trained BERT model has provided, without any substantial task-specific architecture modifications, provided state-of-the-art performances over various NLP tasks. The BERT pre-trained model [22], which uses Google's vanilla transformer language model [23] as a substructure, has been dubbed the "rediscovery" of the natural language processing (NLP) pipeline because of the higher level of language understanding it provides [24]. "However, the BERT model suffers from fixed input length size limitations, word piece embedding problems, and computational complexities" [25]. Table 1 shows a variety of pre-trained BERT models. The primary BERT models as a function of L (number of encoders) and H (number of hidden units). Because bigger models include a significant number of trainable parameters, smaller models are ideal for environments with limited computational resources: an average-sized model like BERT-Base has roughly 110 million trainable parameters, whereas BERT-Large has over 340 million [26].

Table 1 Various BERT model variations

BERT models	$H = 128$	$H = 256$	$H = 512$	$H = 768$	$H = 1024$
$L = 2$	BERT-tiny	–	–	–	–
$L = 4$	–	BERT-miny	BERT-small	–	–
$L = 8$	–	–	BERT-medium	–	–
$L = 12$	–	–	–	BERT-base	–
$L = 24$	–	–	–	–	BERT-large

3 Related Work

Due to the lack of varied kinds of languages their nature, the context of various words, and different meanings with different contexts, the above-discussed techniques do not give a prominent result in the commercial domain. That is why various other approaches evolved, which might be an improvement in existing methods or adding the domain knowledge and then creating those vector representations of word or combining multiple vectors embedding together etc. Many of the researchers try to combine the different vector models to get the effective word embedding for sentiment analysis and related tasks [27]. For text classification, the author proposes a novel language-independent word encoding method. Using a new method named binary unique number of words "BUNOW," the suggested model translates raw text input to low-level feature dimensions with minimum or no preprocessing steps. Each unique word in a dictionary can have an integer ID that is represented as a k-dimensional vector of its binary equivalent with BUNOW. The output vector of this encoding is fed into a convolutional neural network (CNN) model for classification [28]. The most extensively used approach for choosing features is the TF-IDF algorithm. Even though the TF-IDF method is straightforward to apply, it still suffers from semantic deletion, which means it ignores the semantic information in the text. This study introduces LSA as a solution to the problem. The cosine value was used to check the word similarity. The feature extraction is achieved according to the similarity between the words, which compensates for the TF-IDF insufficiency. Finally, the extracted features are applied into classification algorithms [29]. "Bag-Of-Words model is the most widely used technique for sentiment analysis, it has two major weaknesses: using a manual evaluation for a lexicon in determining the evaluation of words and analyzing sentiments with low accuracy because of neglecting the language grammar effects of the words and ignore semantics of the words. Author propose a new technique with enhancement bag-of-words model for evaluating sentiment polarity and score automatically by using the weight of the words instead of term frequency" [30]. Many researchers' works have focused on refining the concept of embedding. The authors proposed models that learn sentiment-specific word embeddings (SSWE) in [31]. The learned word vectors, as well as the semantic sentiment information, are represented as embeddings. The authors of [32] devised word embedding, which is a combination of sentiment supervision at both document

and word levels. The model was trained and learned a sentiment-related embedding. Li and Gong [33] proposes a significant improvement in semantics-oriented word vectors by combining the word embedding model with standard matrix factorization via a projection level. The feature extraction is done by different vector embedding methods such as FastText and Word2Vec given to the RNN module and character-level embedding given to CNN through this hybrid model designed to get the combined feature for sentiment analysis. The author [15] proposes a feature-based fusion adversarial recurrent neural networks (FARNN-Att) composite model with an attention mechanism. Firstly, extracted the long-term dependence of text using the Bi-LSTM network and put forward a novel contextual feature representation way. Then, combined the prediction results of two features vectors in the full connection layer, which can be captured through the feature connection and attention mechanism. To make model more robust and generalized, a regularization method applied can be treated as of adversarial training. In the LE-LSTM model [34], author combined the word sentiment embedding along with its corresponding word embedding. The final layer output of the word sentiment classifier is considered as the sentiment embedding of each word. The word sentiment classifier and word embedding lookup are updated during the main training process. Aiming at the problem of insufficient sentiment word extraction ability in existing text sentiment analysis methods and OOV (out-of-vocabulary) problem of pre-training word vectors, a neural network model combining multi-head self-attention and character-level embedding is proposed by the author in their paper [35]. The author applied the word embedding technique (Word2Vec) along with different types of stemming such as root-based and stem-based, and no stemming for document classification, and found good results as compared to just root model. They also found when using a stem-based algorithm, skip-gram achieves good results with a vector of smaller dimension, while CBOW requires a larger dimension vector to achieve similar performance [36]. The word embedding of word2vec training provides solely semantic information. Before going on to word embedding fusion, the text classification algorithm is employed to learn original text multi-word embedding in morphological, grammatical, and sentiment expressed information. The enhanced convolution neural network [37] is used for sentiment analysis. The author provides an ensemble technique that merges the information from Glove and word2vec embeddings with structured knowledge from the semantic networks ConceptNet and PPDB. For huge text, this reflects a multilingual vocabulary. On several word similarity assessments, the embeddings it generates reach state-of-the-art performance [38]. An embedding method paragraph vector in document ranking given by the author in [39]. They focused on the effects of combining knowledge-based with knowledge-free document embeddings for the text classification task. The author provides a solution to FastText's shortcomings, such as missing words containing contextual information that was not adequately encoded. They initially suggest a shape near word filtering-based approach for retrieving missing nouns. Then, for successful text classification, a self-training word embedding approach based on a knowledge network was combined with pre-training word embeddings to provide the optimum mixed word vector with rich semantics [40]. To overcome the problem in short text news headlines such as sparse data and weak relationship

between words, the author proposed a fusion vector. The Word2Vec tool is used to train the word vector in the short text corpus, and the average word vector is incorporated by the additive averaging method. Afterword's the gist of the short text corpus is formed by the WTTM model to obtain the topic extended feature vector. Finally, the average word vector and the topic extended feature vector are combined [41]. To address the issues of word embedding such as realize the coexistence of words and phrases in the same vector space, the author proposes an enhanced word embedding (EWE) method. Before completing the word embedding, this method introduces a unique sentence reorganization technology to rewrite all the sentences in the original training corpus. Then, all the original corpus and the reorganized corpus are merged together as the training corpus of the distributed word embedding model, so as to realize the coexistence problem of words and phrases in the same vector space. The glove model is used for this fusion embedding [42, 43].

4 Proposed Model

By looking at the state-of-the-art literature above, we try to address the prominent issue of word embedding along with checking the effectiveness of individual word embedding and a fusion model. This will address the various problems of effective sentiment analysis such as variable vector length, context of sentences, representation of words, and OOV words. The proposed model tries to address these issues by combining the word-level model glove for handling general word embedding, a character-based model such as FastText to resolve the problem of out-of-word vocabulary and to maintain the semantic relationship between the word and phrases the BERT model. The model generates the fused vector further passed different DNN classifiers such as Bi-LSTM, GRU, and CNN model.

Combing different vector representation will create a better embedding as compared to the individual embedding vector. Keeping the vector length constant, we used here padding for BERT representation. It generates variable length vector representation. The modified combined vector then feed to deep neural classifier. The classifier will run on test data to learn the weight of words which determines the polarity of the sentence. The two polarities here we considered as positive and negative of the sentence.

5 Result and Discussion

Here, we considered three-word embedding state-of-the-art methods such as Glove, FastText, and BERT. Along with deep learning architecture such as Bi-LSTM, GRU, CNN, all these DNN architecture are implemented as a shallow network for very few-layer. For the experimental purpose, the dataset used are Amazon Kindle, Yelp, and IMDB all taken from the Kaggle dataset library. The models are trained with

Fig. 4 Performance graph of various embedding such as Glove [21], FastText [28], BERT [22], and custom embedding

supervised approach, and positive and negative polarities have been considered (Fig. 4).

From the various experiments conducted during this study, it has been observed that the performance of deep learning model very much depends on kind of vector embedding applied. Words represented with meaningful numbers (vector). To encode this tacit knowledge for a machine to learn effectively just by applying as it is, available encoding method gives good performance with respect to accuracy; however, due to limitation in this general vocabulary, presence for Amazon kindle dataset does not provide good accuracy, whereas for yelp and IMDB, results are very promising. In future, will going to try other deep learning models to check the performance of vector embedding without presence of domain knowledge.

6 Conclusion

Sentiment analysis is used in broad spectrum of industry. Data is available in abundance for sentiment analysis in various forms or reviews, comments, opinions, and many more ways. To have an effective sentiment analysis encoding, this data is important, and to do these, various techniques are available in literature along with commercial environment. But still there is significant improvement based on the domain and varied nature of available languages. The various vector embedding such as Glove, FastText, and BERT provides a good way to represent word but insignificant in maintaining the context, OOV handling, variable length to represent words. To reap the best from each model, the proposed model creates the combined vector representation. The experimental result shows the significant performance with shallow neural network. In future, we would like to check the interpretability of model by adding attention mechanism look for what model learned.

References

1. Cheng, *Deep Learning for Automated Sentiment Analysis of Social Media* (n.d.). https://doi.org/10.1145/3341161.3344821
2. S. Jawale, *Interpretable Sentiment Analysis Based on Deep Learning: An Overview* (n.d.). https://doi.org/10.1109/PuneCon50868.2020.9362361
3. Esichaikul, *Sentiment Analysis of Thai Financial News* (n.d.). https://doi.org/10.1145/3301761.3301773
4. S. Nimje et al., Prediction on stocks using data mining, in *Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST)* (2020)
5. Cirqueira, *Improving Relationship Management in Universities with Sentiment Analysis and Topic Modeling of Social Media Channels: Learnings from UFPA* (n.d.). https://doi.org/10.1145/3106426.3117761
6. Whitelaw, *Using Appraisal Groups for Sentiment Analysis* (n.d.). https://doi.org/10.1145/1099554.1099714
7. Wong, *Sentiment Analysis across the Courses of a MOOC Specialization* (n.d.). https://doi.org/10.1145/3287324.3293864
8. Smith. *Cross-Discourse Development of Supervised Sentiment Analysis in the Clinical Domain* (n.d.)
9. Buscaldi, *Sentiment Analysis on Microblogs for Natural Disasters Management: A Study on the 2014 Genoa Floodings* (n.d.). 0.1145/2740908.2741727
10. P. Yenkar, *A Survey on Social Media Analytics for Smart City* (n.d.). https://doi.org/10.1109/I-SMAC.2018.8653707
11. Paul, *Compass: Spatio Temporal Sentiment Analysis of US Election What Twitter Says!* (n.d.) https://doi.org/10.1145/3097983.3098053
12. Yun, *Multi-Categorical Social Media Sentiment Analysis of Corporate Events* (n.d.). https://doi.org/10.1145/3154943.3154957
13. Xu, *Fast Learning for Sentiment Analysis on Bullying* (n.d.). https://doi.org/10.1145/2346676.2346686
14. W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. **5**(4), 1093–1113 (2014)
15. M.U. Salur, I. Aydin,A novel hybrid deep learning model for sentiment classification. IEEE Access **8**, 58080–58093 (2020)
16. Y. Zhang, R. Jin, Z.-H. Zhou, Understanding bag-of-words model: a statistical framework. Int. J. Mach. Learn. Cybernet. **1**(1–4), 43–52 (2010)
17. P. Bojanowski et al., Enriching word vectors with subword information. Trans. Assoc. Computat. Linguist. **5**, 135–146 (2017)
18. S. Al-Saqqa, A. Awajan,The use of word2vec model in sentiment analysis: a survey, in *Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control* (2019)
19. N. Nandakumar, B. Salehi, T. Baldwin,A comparative study of embedding models in predicting the compositionality of multiword expressions, in *Proceedings of the Australasian Language Technology Association Workshop* (2018)
20. E. Peters Matthew et al.,Deep contextualized word representations (2018). arXiv preprint arXiv:1802.05365
21. J. Pennington, R. Socher, C.D. Manning,Glove: global vectors for word representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2014)
22. A. Chiorrini et al.,Emotion and sentiment analysis of tweets using BERT. EDBT/ICDT Workshops (2021)
23. T. Mikolov et al.,Efficient estimation of word representations in vector space (2013). arXiv preprint arXiv:1301.3781
24. R. Al-Rfou et al., Character-level language modeling with deeper self-attention, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33(01) (2019)

25. J. Devlin et al.,Bert: pre-training of deep bidirectional transformers for language understanding (2018). arXiv preprint arXiv:1810.04805
26. I. Tenney, D. Das, E. Pavlick,BERT rediscovers the classical NLP pipeline (2019). arXiv preprint arXiv:1905.05950
27. J.M.C. Lima, J.E.B. Maia, A topical word embeddings for text classification, in *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional* (SBC, 2018)
28. R. MohammadiBaghmolaei, A. Ahmadi, Word embedding for emotional analysis: an overview, in *2020 28th Iranian Conference on Electrical Engineering (ICEE)* (IEEE, 2020)
29. B.O. Oscar et al.,Sentiment analysis with word embedding, in *2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST)* (IEEE, 2018)
30. A.A. Helmy, Y.M.K. Omar, R. Hodhod,An innovative word encoding method for text classification using convolutional neural network, in *2018 14th International Computer Engineering Conference (ICENCO)* (IEEE, 2018)
31. Y. Li, B. Shen,Research on sentiment analysis of microblogging based on LSA and TF-IDF, in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)* (IEEE, 2017)
32. D.M. El-Din, Enhancement bag-of-words model for solving the challenges of sentiment analysis. Int. J. Adv. Comput. Sci. Appl. **7**(1) (2016)
33. S. Li, B. Gong,Word embedding and text classification based on deep learning methods, in *MATEC Web of Conferences*, vol. 336 (EDP Sciences, 2021)
34. Y. Ma, H. Fan, C. Zhao, Feature-based fusion adversarial recurrent neural networks for text sentiment classification. IEEE Access **7**, 132542–132551 (2019)
35. X. Fu et al., Lexicon-enhanced LSTM with attention for general sentiment analysis. IEEE Access **6**, 71884–71891 (2018)
36. H. Xia, C. Ding, Y. Liu, Sentiment Analysis model based on self-attention and character-level embedding. IEEE Access **8**, 184614–184620 (2020)
37. H.A. Almuzaini, A.M. Azmi, Impact of stemming and word embedding on deep learning-based Arabic text categorization. IEEE Access **8**, 127913–127928 (2020)
38. X. Sun et al.,Text sentiment polarity classification method based on word embedding, in *Proceedings of the 2018 2nd International Conference on Algorithms, Computing and Systems* (2018)
39. R. Speer, J. Chin,An ensemble method to produce high-quality word embeddings (2016). arXiv preprint arXiv:1604.01692
40. H. Machhour, I. Kassou, Concatenate text embeddings for text classification, in *2017 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)* (IEEE, 2017)
41. H. Wang, H. Kun Guo, Z. Liu, Mixed word embedding method based on knowledge graph augment for text classification, in *2019 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)* (IEEE, 2019)
42. J. Ge, H. Wang, Y. Fang,Short text classification method combining word vector and WTTM, in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Vol. 1 (IEEE, 2020)
43. S. Hu et al.,Enhanced word embedding method in text classification, in *2020 6th International Conference on Big Data and Information Analytics (BigDIA)* (IEEE, 2020)
44. S.P. Shamseera, E.S. Sreekanth, Word vectors in sentiment analysis. Int. J. Curr. Trends Eng. Res. (IJCTER) **2**(5), 594–598 (2016)
45. F.A. Acheampong, F. Adoma, W. Chen, N.-M.Henry, Text-based emotion detection: Advances, challenges, and opportunities. Eng. Rep. **2**(7), e12189 (2020)

3G Cellular Network Fault Prediction Using LSTM-Conv1D Model

N. Geethu and M. Rajesh

Abstract Cellular network plays an important role in daily life by exploring digital world of communication. Cellular network technology continuously evolves in past decades from 1 to 5G and beyond. The evolution results in more network accessibility and data utilization. As the availability of network, mobility, and portability of cellular devices are increasing, the network traffic will also be increasing. Higher the transmission rates, higher will be the fault occurrence possibility. Monitoring network parameters and finding fault in cellular network are key factor in determining consistency of network. Cellular network which is highly dynamic than usual networks needs intelligent way of fault handling as the human over head will be unpredictable and very high. Modeling intelligent network fault identification system can simplify human efforts and improve efficiency with better accuracy. The research is on real-time data of 3G cellular network including various network parameters like uplink threshold and identifies the behavior of data usual or unusual to predict the fault occurrence. The study is on various LSTM techniques such as bidirectional LSTM, vanilla LSTM, and stacked LSTM combined with time distributed Conv1D.

Keywords Fault prediction · Deep learning · Prediction model · Cellular network · Bidirectional LSTM · LSTM-Conv1D

1 Introduction

Emerging cellular network technologies evolve with better connectivity and scalability. The competing cellular business world ensures the quality of service by maintaining stable fault management mechanism. This is one of the key factors to have growth in the market. Mobile applications, online shopping, banking, online

N. Geethu (✉) · M. Rajesh
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: geethunellikal@gmail.com

M. Rajesh
e-mail: m_rajesh@blr.amrita.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_31

classes, video conferencing, etc., are all part of common life. These are all demanding stable network and consistent performance. The trust factor will be mainly based on effective fault identification and even can be applied for fault prevention as next level.

Increasing network connectivity and traffic demands efficient fault management technique. As the cellular networks increase the mobility and availability, the chances of fault occurrence will also be high. Unhandled faults in connectivity result in performance degradation of the network. There should be effective fault management system with efficient way of managing the network traffic without failure.

The analysis of the cellular network real-time dataset with fixed and variable parameters on normal and fault detected cases is essential to develop model to handle faults in network. The human resource utilization on this consumes more time, cost and prone to errors. Formulate prediction model that can diagnose fault in the presence of certain parameters occurrence using neural networks is highly advanced and has future scope in terms of evolving deep learning techniques. In case of limited data, generative adversarial neural networks can be used to enhance the dataset of the deep learning model.

Existing studies explain the scope of research in this area considering the limitations of confidential real-time cellular network data. Here the objective is to model advanced fault management system for predicting faults for 3G cellular network based on fixed and variable network parameters using various deep learning models and identify the best. The focus is on LSTM techniques such as bidirectional LSTM, vanilla LSTM, and stacked LSTM combined with time distributed Conv1D.

The study identifies advanced deep learning technique for fault prediction in terms of analyzing different types of LSTM models combined with time distributed 1D convolution feature extraction on 3G cellular network data.

2 Literature Review

Tan and Pan [1] explain neural network-based fault prediction model with high efficiency. The logs of network transmission are used as dataset. Network log is preprocessed first for this. It uses time window to get the required samples. This is implemented on two levels. Sample features are extracted using convolutional neural networks. The convolutional networks extract the hidden features from logs. The LSTM model is used for predicting the results. Based on this, a hybrid model is developed for prediction of the network fault. Here it uses CNN-based feature extraction technique. Model has high accuracy in real-time data processing. But the CNN-based feature extraction from logs approach, adds more complexity to system.

Ji et al. [2] explain NLP-CNN-based fault prediction model using network log analysis. The logs from various log files are extracted and processed using NLP. The CNN model has embedded layer to get features from logs. The vectors obtained from this will be the input to next layer. 1D convolution is applied to input. The model is

compared with LSTM model. As it is doing convolution for feature extraction of logs on top of embedded layer, the model has more complexity in processing real-time data.

Zheng et al. [3] explain CNN-LSTM prediction model for predicting failure in equipment in industry. The model is basically using industry equipment data for study. The paper involves comparison of LSTM model with hybrid LSTM-CNN model. The dataset contains matrix of values over a period of variables affecting system to predict failure. CNN is used for feature extraction. The convolutional layer extracts the features using filter. The relevant matrix data is extracted and is filtered using kernels. Here it reduces the dimension using pooling. The final fully connected layer reduces features to one dimension. This is fed to LSTM network. The model can be used to predict future fault occurrence for very large dataset effectively. This is effective way to overcome exploding gradient problem compare to other RNN techniques. While modeling cellular network fault prediction, the model needs to be analysed using real-time cellular network data input and related research using advanced LSTM-CNN network models.

Rezaei et al. [4] explain unsupervised machine learning model. For fault diagnosis, the paper applies expectation maximization (EM), an unsupervised method. It uses DBSCAN along with it, which is a method of grouping data based on density, providing random noise. The method is applied in mobile network communication to implement the self-decision-making systems. Here the different algorithms for grouping data are being compared with each other, and the results are being analysed. Each data group is related to a specific condition causing the fault. Also, there will be related guidelines to solve the faults. The machine learning technique of grouping both traffic and signaling is considered as the most relevant sources of fault information. The method is affected by human error during the modeling stage. The efforts will be high and are difficult to manage highly complex system.

The analysis of existing papers on studies conducted in this area mainly gives the statement of developing a most advanced deep learning-based model, considering the evolving highly dynamic cellular network nature. The combined LSTM–Conv1D models should be investigated more with possible varieties, as the advancement in such latest technology can give the most desirable results considering the above analysis.

3 Design and Implementation

3.1 The LSTM-Conv1D Design

Long short-term memory is advanced RNN technic to overcome the short memory. LSTM has three gates, forget gate, input gate, and output gate. The forget gate decides the information to discard from the cell, input gate decides values from the input to update the memory state, and output gate decides what to output based on input and

Fig. 1 Block diagram of basic fault prediction model using LSTM

the memory of the cell [5]. This is mainly used for models which consider previous data states for result generation. The LSTM which has memory to store previous state can help on the dependability of output on previous results. Three types of models are:

- Prediction of next input sequence value.
- Prediction of input sequence category.
- Generation of another output sequence.

This paper focuses on the second type of LSTM application which predicts the category, (here the natural/unusual network parameters) of inputs as in Fig. 1. This is commonly used in applications such as

- Detect and predict the anomalies: For input containing various observation of parameters or sensors, predict if the sequence is having anomaly or is normal [5].
- Categories DNA Sequence: For DNA sequence input having *A*, *C*, *G*, and *T* values, predict whether the sequence is for a coding or non-coding region.
- Analyze Sentiment data: Given a sequence of text such as a review or a tweet, predict whether the sentiment of the text is positive or negative

There are three approaches considered for building the model. The models are all using deep learning techniques for feature extraction and prediction as the traditional ways are found to have less efficiency and less improvement during the analysis on the previous studies itself (even with the regular network data). Time distributed layer is used to wrap convolutional layer to apply to all samples.

a. **Vanilla LSTM and Time distributed Conv1D**

The approach uses 1D convolution for feature extraction. The 1D convolution has kernel move only in one dimension. Unlike the normal 2D CNN for image data, which has two dimensional kernel movement, it can extract features of text, numeric and audio inputs. One of such examples is to extract the sensors data and time series data [6]. It uses input data of two dimensions, for example, change of network parameter values with respect to time. Here the network parameters can be given as input to the Conv1D layer. The Vanilla LSTM is LSTM model with single hidden layer. Figure 2 shows the fault prediction model using Vanilla LSTM and time distributed Conv1D.

b. **Bidirectional LSTM and Time distributed Conv1D**

The bidirectional LSTM as name indicates, the input sequence is analysed in forward and backward directions. Then it combines both results. The model uses 1D convolution for feature extraction and later combined with bidirectional LSTM for predicting the fault for given network parameters. Figure 3 shows the fault prediction model using bidirectional LSTM and time distributed Conv1D.

Fig. 2 Fault prediction model using Vanilla LSTM and time distributed Conv1D

iii. **Stacked Bidirectional LSTM and Time distributed Conv1D**

The term stacked indicates that multiple layers are combined. The stacked bidirectional model used 1D convolution layer and multiple bidirectional LSTM layers. Figure 4 shows the fault prediction model using stacked bidirectional LSTM and time distributed Conv1D.

3.2 Implementation

For all the three model categories, the 12 features are given as input from input layer to time distributed Conv1D with 64 neurons.

Fig. 3 Fault prediction model using bidirectional LSTM and time distributed Conv1D

For vanilla LSTM model, it is performing convolution operation in one direction to extract the relevant features. The flatten layer converts the outcome of this into one-dimensional vector. This is fed to LSTM layer. There is no regularization used as the best result from this type of model is without any regularization [7].

For bidirectional LSTM with time distributed Conv1D also, the input layer is fed output to time distributed Conv1D with 64 neurons, to extract the relevant features. The bidirectional LSTM layer processes the input for prediction in two directions and yields best results among this category of models during the implementation.

The stacked bidirectional model has three-layered bidirectional LSTM stack. The extracted features from the convolution are fed to the first layer of the LSTM stack. The model is three-layered stack.

Fig. 4 Fault prediction model using stacked bidirectional LSTM and time

4 Results and Analysis

The Vanilla LSTM and time distributed Conv1D model are implemented with various parameters such as number of neurons, epochs, test train splits, loss functions, output activation function, optimizer, and the regularization.

The dropout is the regularization mechanism which will remove unwanted weight and reduce overfitting.

This L2 regularization also reduces overfitting by reducing the unimportant weight values. Table 1 indicates the notable models with various parameter values.

Table 2 indicates the accuracy and loss plots of Vanilla LSTM model. It shows that the accuracy is maximum with second model which is using Adam optimizer, MSE loss function, and RELU activation function. The SGD optimizer is giving poor results compared to Adam optimizer. The accuracy and loss graphs in Table 0.2 show there is no bias or variance for this model as the test and validation loss curves are almost syncing with each other.

The bidirectional LSTM-Conv1D models with various parameters and the accuracy and loss plot table are shown in Tables 3 and 4, respectively.

The best accuracy result is without regularization. The model3 without dropout and L2 regularization are showing the best result without any bias or variance. The loss curves and accuracy curves are showing a good sync in this model.

The stacked bidirectional LSTM-Conv1D models with parameter changes are shown in Table 5. The high accuracy is with model2 without any regularization, but it is showing high variance in Table 6 explaining the accuracy and loss plots. Regularization model has less accuracy than the bidirectional LSTM model.

The best model for the fault prediction is using bidirectional LSTM. The confusion matrix is given in Table 7.

The precision value is 0.97, recall is 0.83, and the F1 score is 0.8. This indicates the better performance of the model.

Table 1 Vanilla LSTM models with various parameter values

	Model 1	Model 2	Model 3	Model 4
Accuracy	72.5	83.18	82.1	81.01
Validation loss	0.19	0.12	0.13	0.13
Regularization	–	–	Dropout (0.2)	Dropout and L2
Epochs	50	100	100	100
Optimizer	SGD	Adam	Adam	Adam
Test split	0.1	0.1	0.1	0.1
Loss function	MSE	MSE	MSE	MSE
Output activation function	RELU	RELU	RELU	RELU

Table 2 Accuracy and loss plots of Vanilla LSTM models

	Accuracy	Loss
Model 1		
Model 2		
Model 3		
Model 4		

5 Conclusion and Future Scope

The cellular network fault prediction model provides better results using bidirectional LSTM and time distributed Conv1D architecture. As the cellular network data is highly dynamic and large in amount, the deep learning model will provide an efficient fault detection mechanism compared to traditional mechanisms.

Table 3 Bidirectional LSTM models with various parameter values

	Model 1	Model 2	Model 3	Model 4
Accuracy	83.5	82.6	85.18	81.66
Validation loss	0.12	0.12	0.11	0.13
Regularization	–	Dropout	–	Dropout and L2
Epochs	100	100	200	100
Optimizer	Adam	Adam	Adam	Adam
Test split	0.1	0.1	0.1	0.1
Loss function	MSE	MSE	MSE	MSE
Output activation function	RELU	RELU	RELU	RELU

Table 4 Accuracy and loss plots of bidirectional LSTM models

	Accuracy	Loss
Model 1		
Model 2		
Model 3		
Model 4		

Cellular network fault management using deep learning techniques is a research area of wide scope as it is less under investigation compared to other areas. The limited and incomplete real-time confidential data is a major hurdle. There can be used generative adversarial neural networks (GAN) in further steps to generate data based on available one. The model can be evaluated using other cellular network data from 4G or 5G networks as well.

Table 5 Stacked bidirectional LSTM models with various parameter values

	Model 1	Model 2	Model 3	Model 4
Accuracy	74.6	87.5	81.4	76.04
Validation loss	0.11	0.18	0.13	0.16
Regularization	–	–	Dropout	Dropout and L2
Epochs	50	100	100	100
Optimizer	SGD	Adam	Adam	Adam
Test split	0.1	0.1	0.1	0.1
Loss function	MSE	MSE	MSE	MSE
Output activation function	RELU	RELU	RELU	RELU

Table 6 Accuracy and loss plots of stacked bidirectional LSTM models

	Accuracy	Loss
Model 1		
Model 2		
Model 3		
Model 4		

Table 7 Confusion matrix of prediction model

	Actual positive (1)	Actual negative (0)
Predicted positive (1)	2572	75
Predicted negative (0)	515	520

References

1. Z. Tan, P. Pan, Network fault prediction based on CNN-LSTM hybrid neural network, in *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)* (Haikou, China, 2019), pp. 486–490, https://doi.org/10.1109/CISCE.2019.00113
2. W. Ji, S. Duan, R. Chen, S. Wang, Q. Ling, A CNN-based network failure prediction method with logs, in *2018 Chinese Control Decision Conference (CCDC)* (2018), pp. 4087–4090 https://doi.org/10.1109/CCDC.2018.8407833
3. L. Zheng et al., A fault prediction of equipment based On CNN-LSTM network, in *2019 IEEE International Conference on Energy Internet (ICEI)* (Nanjing, China, 2019), pp. 537–541. https://doi.org/10.1109/ICEI.2019.00101
4. S. Rezaei, H. Radmanesh, P. Alavizadeh, H. Nikoofar, F. Lahouti, Automatic fault detection and diagnosis in cellular networks using operations support systems data, in *NOMS 2016—2016 IEEE/IFIP Network Operations and Management Symposium* (Istanbul, 2016), pp. 468–473. https://doi.org/10.1109/NOMS.2016.7502845
5. https://github.com/Ahmed-Hussein2009/Jason-Brownlee-books/blob/main/8.%20Long%20Short-Term%20Memory%20Networks%20With%20Python%20Develop%20Sequence%20Prediction%20Models%20With%20Deep%20Learning.pdf
6. https://towardsdatascience.com/understanding-1d-and-3d-convolution-neural-network-keras-9d8f76e29610%20WRITTEN%20BY%20Z%C2%B2%20Little%20Data%20Science%20Diary%20Follow%2010%2010
7. https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/

Deep Learning for Part of Speech (PoS) Tagging: Konkani

Annie Rajan, Ambuja Salgaonkar, and Arshad Shaikh

Abstract This is the first time that an experiment using deep learning has been attempted with a Konkani language data set. For this study, over 100,000 PoS tagged Konkani sentences were used. The f-scores for deep learning are 90.73% for training data and 71.43% for test data. These results are better than the ones previously reported for Konkani. We have provided a list of references of PoS tagging for Indian languages specifically using deep learning to place our research in perspective.

Keywords Part of speech tagging · Konkani · Deep learning · Part of speech tagger · Computing methodologies · Machine learning

1 Introduction

Tagging can be performed by employing supervised learning if a manually tagged data set is available; otherwise, the learning is unsupervised [1]. Among the various machine learning approaches, the most widely used for PoS tagging are rule-based and stochastic. The rules in a rule-based tagger are initially handcrafted by manual analysis of the data [2, 3]. PoS tagging methods include hidden Markov model (HMM) [4], support vector machine (SVM) [5], decision tree [6], maximum entropy [7], conditional random field (CRF) [8], and deep learning [9–12].

A. Rajan (✉) · A. Salgaonkar
Department of Computer Science, University of Mumbai, Mumbai, India
e-mail: ann_raj_2000@yahoo.com

A. Salgaonkar
e-mail: ambujas@udcs.mu.ac.in

A. Rajan
DCT's Dhempe College, Panaji, India

A. Shaikh
Learning and Personalization, BYJU'S, Bangalore, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_32

The research results presented in this paper are: With deep learning, the f-score obtained is 90.20% after training and 71.43%, when employed on a set of 108,546 sentences of a Konkani database developed by the Indian Languages Corpora Initiative (ILCI).[1]

As far as we can ascertain, this is the first time that deep learning methods have been implemented for PoS tagging of Konkani language on a data set of such a large size.

2 Literature Survey

Comparable work in various Indian languages is summarized for an appraisal of our PoS tagging experiments in Konkani: Table 1 describes data sets and a performance results when deep learning is employed.

Konkani, the official language of the small but important Indian State of Goa, is making rapid progress in putting NLP applications to use. The Bureau of Indian Standards (BIS) has published annotation standards for Konkani [20]. The tag set has eleven parts of speech: verb, adverb, noun, pronoun, demonstrative, adjective, conjunction, postposition, quantifiers, particles, and residuals.

3 Brief History of Konkani Language and Its Speakers

The central portion of the narrow coastal strip between the Arabian Sea and Western Ghats in peninsular India is referred to as Konkan. The region is about 800 km long and about 45–75 km wide. The Konkani language is spoken along a total of about 600 km of the coast, along the central and southern parts of Konkan, and further south. As such, the geographical extent of the main Konkani speaking area is relatively small. Furthermore, it is not the only language spoken in this region; the major Indian languages like Marathi and Kannada are also spoken alongside. It is spoken by a population of 2.2 million people, spread across neighboring states of Goa, Karnataka, Maharashtra, and Kerala. Konkani attained its full potential around the eleventh century, and by the fifteenth century, it had evolved into a full-fledged literary language [21].

Small migrant communities of Konkani speakers[2] are found in Uganda, Kenya, Pakistan, Portugal, and the Persian Gulf [21, 22]. As a result, the language has been written in a be-wildering multiplicity of scripts: Devanagari, Roman, Kannada, Malayalam, Perso-Arabic. This led to the outward splitting up of one and the same language as it is spoken [23]. For historical as well as technical reasons, resources

[1] http://tdil.meity.gov.in/

[2] https://en.wikipedia.org/wiki/Konkani_language.

Table 1 PoS tagging in Indian languages using deep learning

S. No.	Language	Size of data set (words)		Performance measure (%)
		Training data	Test data	
1	Malayalam [13]	9000 tweets	915 tweets	92.54 (f-score)
2	Kannada [14]	10,500 sentences	1600 sentences	92.00 (f-score)
3	Sanskrit [15]	34,270 words	4321 words	97.86 (accuracy)
4	Bengali [16]	102,933 words	–	93.28 (f-score)
5	Hindi [17]	23,565 words	23,281 words	92.19 (accuracy)
6	Nepali [18]	100,720 words	–	92.00 (f-score)
7	Telugu [19]	67,861 sentences	–	81.73 (accuracy)

for Konkani have remained paltry, and their development, which although began several years ago, is slow. Sunaprant, a Konkani newspaper in Devanagari script, was operational during the year 1987–2015.[3] Bhaangar Bhuin, another Konkani newspaper in Devanagari script, was started in the year 2016 and made its presence in digital form in the year 2017.[4] Over the past two centuries, many newspapers were published in Devanagari, Kannada, and Roman script, but none has survived up to the present day.

4 Data Set

A manually annotated corpus from ILCI in the domains of tourism, health, entertainment, agriculture, and history with a total of 108,546 sentences is used for the experiment. Initially, the PoS tagger is trained on 86,837 sentences (1,006,177), and the test data has 21,709 sentences (125,389 words). The number of tags for the two data sets is 39, while the BIS tag set for Konkani has only 29 tags [20].

5 Experiments

The steps taken for PoS tagging using deep learning are shown in the block diagram of Fig. 1, and each step is explained in sufficient detail.

Embedding: A sequence of words to be tagged is the input to a hidden layer that converts each word of the sequence into a dense vector. It has three parameters, as follows: The input dimension is the size of the vocabulary of the test data. The words are encoded by the integers 0, 1, 2,…, because what is relevant is not the

[3] https://en.wikipedia.org/wiki/Sunaparant.

[4] https://en.wikipedia.org/wiki/Bhaangar_Bhuin.

Fig. 1 Block diagram for the biLSTM experiment

exact string of letters making up each word, but the various contexts in which it occurs. The input length is the length of the sequence of words that is input to the layer. The sequence length is a constant; if an actual sequence is of smaller length, it is to be padded with dummy integers. F-scores were calculated for word sequences of length of 7, 15, 23, 30, 40, and 50 obtained from the ILCL data set. It was found that sequence length of 23 gave the highest score among all, see Table 2. The output dimension is the size of the vector space, i.e., the number of components of the vector, into which the word will be embedded. The model is trained for 10 epochs, where an epoch is when the complete data set is passed through the model. It was found that the accuracy did not improve after 10 epochs, so the model was trained for 10 epochs.

Dropout: To reduce the overfitting of the dense vectors to the training data, i.e., to regularize the overfitted data, some components of the dense vector are randomly reset to 0 and the weightage of the remaining components is increased correspondingly.

Bidirectional: In bidirectional LSTM, the training data is supplied to two LSTM net- works (Fig. 2), one in the forward direction and another in the reverse direction. This technique is employed to learn the prefix and suffix of a given word.

The overall structure of bidirectional LSTM is same as LSTM. In biLSTM, the sequence is passed in both directions, two set of hidden states output form the biLSTM, and these states are combined to get a rich hidden state.

Consider the sentence “sə̪pt̪ə̪puɾiʧẽ d̪ə̪ɾʃən gʱet̪ɭjaɾ mokʃə̪ mel̪t̪a” (सप्तपुरीचें द्र्शन घेतल्यार मोक्ष मेळटा), as an example to predict part of speech for the word “gʱet̪ɭjaɾ” (घेतल्यार) in the sentence mentioned above. In bidirectional LSTM, it will be able to see the past information and also the information further down the road, i.e., forward LSTM: “sə̪pt̪ə̪puɾiʧẽ d̪ə̪ɾʃən gʱet̪ɭjaɾ” and backward LSTM: “gʱet̪ɭjaɾ

Table 2 F-scores using biLSTM method for different sequence lengths of input data

S. No.	Sequence length	Recall (%)	Precision (%)	F-score (%)
1	7	53	58	55
2	15	56	66	61
3	23	60	72	66[a]
4	30	60	71	65
5	40	61	68	64
6	50	50	57	54

[a]Sequence length of 23 gives the best results

Fig. 2 Example of biLSTM-based analysis

mokʃə melʈa". The computed hidden state of both forward and backward LSTM is passed to the next layer. The equations used for this experiment are as per the reference of the paper [24].

Time-distributed dense layer: The last layer of the model is used to assign an appropriate PoS tag to each word. The dense layer generates a sequence of one-hot vectors, each of which contains probabilities of tags to be assigned to a single word. Since the dense layer needs to run on each word of the sequence, we used a time-distributed modifier. The predicted one-hot vectors are converted back to tagged sentences, i.e., sequences of words with PoS tags, after prediction. For training and testing, we used Keras with TensorFlow. (TensorFlow is an open-source software library for machine learning applications like neural networks. Keras is a framework for designing and implementing neural networks, with libraries such as TensorFlow, Theano, or the cognitive toolkit (CNTK).

Fig. 3 PoS tagger for Konkani language

6 POS Tagger

A PoS tagger for Konkani language was developed using the Django framework. The tagger[5] allows the user to enter a text in Devanagari script, or upload a TXT file. The text can be tagged using a choice of deep learning. The tagged text is displayed in the output tab or can be downloaded in TXT format. Other tabs display the following information: the statistics of the output, the tags that occurred in the given input text, and the words that could not be tagged by deep learning, and finally, statistics about the training corpus (Fig. 3).

[5] http://103.86.177.205:5000/

Table 3 Tenfold validation output of training data for deep learning

S. No.	Cross-validation	Recall (%)	Precision (%)	F-score (%)	Accuracy (%)
1	Run 1	90.25	90.70	90.38	99.84
2	Run 2	90.06	90.43	90.16	99.84
3	Run 3	90.40	90.74	90.48	99.85
4	Run 4	90.22	90.59	90.33	99.84
5	Run 5	90.38	90.54	90.37	99.84
6	Run 6	90.55	90.63	90.51	99.84
7	Run 7	90.41	90.65	90.45	99.84
8	Run 8	90.19	90.45	90.23	99.84
9	Run 9	90.33	90.55	90.35	99.84
10	Run 10	90.28	90.61	90.35	99.84

Table 4 Fivefold validation output of test data for deep learning

S. No.	Cross-validation	Recall (%)	Precision (%)	F-score (%)	Accuracy (%)
1	Run 1	72.91	71.41	71.83	98.88
2	Run 2	73.18	71.43	71.94	98.87
3	Run 3	73.31	71.56	72.09	98.87
4	Run 4	73.95	71.74	72.23	98.88
5	Run 5	72.98	71.70	72.00	98.87

7 Results

A tenfold cross-validation has been was conducted by randomly selecting 43,418 sentences each time from the training data set for deep learning (Table 3). A fivefold cross-validation has been carried out by randomly choosing 10,854 sentences each time from the test data set for deep learning (Table 4).

8 Discussion

The number of misclassified words for the deep learning algorithm is 44,731 or 4.56% of the training data set of 1,006,088. Table 5 contains the list of tags and the misclassified word count.

From Table 5, we observe that the five tags with the highest errors are N_NN, RB, JJ, V_VM_VF, V_VM_VNF, which resulted in maximum errors. Further analysis of these words will help us to improve the f-score and accuracy of the tagger.

Table 6 contains examples of incorrectly tagged words by biLSTM, respectively.

Table 5 Word count of misclassified words using biLSTM

S. No.	Tag	Incorrectly tagged	Incorrectly tagged (%)
1	N-NN	7932	6.72
2	RB	4973	6.21
3	JJ	4157	3.52
4	V_VM_VF	3282	2.78
5	V_VM_VNF	2912	2.47
6	N_NNP	2289	1.94
7	DM_DMD	2020	1.71
8	PSP	1762	1.49
9	PR_PRP	1434	1.21
10	N_NST	1431	1.21
11	QT_QTF	1398	1.18
12	CC_CCS	1365	1.16
13	DM_DMR	1064	0.90
14	PR_PRL	790	0.67
15	QT_QTC	693	0.59
16	V_VAUX_VF	675	0.57
17	V_VM_VNG	626	0.53
18	DM_DMQ	583	0.49
19	RD_UNK	564	0.48
20	V_VM_VINF	557	0.47
21	CC_CCD	555	0.47
22	CC_CCS_UT	472	0.40
23	RP_INTF	457	0.39
24	QT_QTO	394	0.33
25	PR_PRQ	380	0.32
26	PR_PRI	375	0.32
27	PR_PRF	301	0.25
28	DM_DMI	275	0.23
29	PR_RPD	214	0.18
30	V_VAUX_VNF	213	0.18
31	RP_NEG	160	0.14
32	RP_INJ	123	0.10
33	V_VNG	95	0.08
34	PR_PRC	75	0.06
35	V_VM	46	0.04
36	RD_ECH	38	0.03
37	RD_RDF	23	0.02
38	RP_CL	19	0.02
39	RD_SYM	9	0.01

Table 6 Examples of words tagged incorrectly by biLSTM

S. No.	Word	Tagged in golden standard	Incorrectly tagged by biLSTM
1	təʃɪ (तऴी)	RB	DM_DMD
2	səgljãɫ (सगलययींत)	QT_QTF	N_NN
3	gəʊɾəw (गौरळ)	JJ	N_NN
4	mədʰjə (मध्य)	N_NNP	N_NST
5	patəɭilli (पयतळललऱ)	V_VM_VNF	V_VM_VF

9 Conclusion

In this paper, we have presented deep learning strategies for POS tagging of Konkani language. This is the first attempt to create tagger that uses deep learning. F-scores of 90.73% were obtained using deep learning, while the error rates were 4.55%. A detailed analysis of this situation will aid in enhancing the f-score of the PoS tagger.

References

1. F.M. Hasan, N. UzZaman, M. Khan, *Comparison of different POS Tagging Techniques (N-Gram, HMM and Brill's tagger) for Bangla* (2007)
2. B.B. Greene, G.M. Rubin, *Automatic Grammatical Tagging of English* Department of Linguistics (Brown University, 1971)
3. S. Klein, R.F. Simmons, A computational approach to grammatical coding of English words. J. ACM (JACM) **10**, 334–347 (1963)
4. L. Rabiner, B. Juang, An introduction to hidden Markov models. IEEE ASSP Mag. **3**, 4–16 (1986)
5. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
6. S.R. Safavian, D. Landgrebe, A survey of decision tree classifier methodology. IEEE Trans. Syst. Man Cybern. **21**, 660–674 (1991)
7. J. Skilling, *Classical MaxEnt Data Analysis, Maximum Entropy and Bayesian Methods* (Kluwer Academic Publishers, 1989)
8. J. Lafferty, A. McCallum, F.C. Pereira, Conditional random fields: probabilistic models for segmenting and labeling sequence data (2001)
9. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput. **9**, 1735–1780 (1997)
10. M.I. Jordan, Attractor dynamics and parallelism in a connectionist sequential machine (1990)
11. K. O'Shea, R. Nash, An introduction to convolutional neural networks, arXiv preprint arXiv:1511.08458 (2015)
12. A. Sherstinsky, Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. D Nonlinear Phenom. **404**, 132306 (2020)
13. S. Kumar, M.A. Kumar, K.P. Soman, Deep learning based part-of-speech tagging for Malayalam Twitter data (Special issue: deep learning techniques for natural language processing). J. Intell. Syst. **28**, 423–435 (2019)

14. K.K. Todi, P. Mishra, D.M. Sharma, Building a kannada pos tagger using machine learning and neural network models, arXiv preprint arXiv:1808.03175 (2018)
15. B. Premjith, K.P. Soman, P. Poornachandran, A deep learning based Part-of-Speech (POS) tagger for Sanskrit language by embedding character level features., in *FIRE*, pp. 56–60
16. M.F. Kabir, K. Abdullah-Al-Mamun, M.N. Huda, Deep learning based parts of speech tagger for Bengali, in *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 26–29
17. Parikh, Part-of-speech tagging using neural network. Proc. ICON (2009)
18. G. Prabha, P.V. Jyothsna, K.K. Shahina, B. Premjith, K.P. Soman, A deep learning approach for part-of-speech tagging in Nepali language, in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1132–1136
19. M. Jagadeesh, M.A. Kumar, K.P. Soman, Deep belief network based part-of-speech tagger for Telugu language, in *Proceedings of the Second International Conference on Computer and Communication Technologies*, pp. 75–84
20. M. Sardesai, J. Pawar, S. Walawalikar, E. Vaz, BIS Annotation standards with reference to Konkani Language, in *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing*, pp. 145–152
21. S. Borkar, *Let us Learn Konkani* (Abhiman Publications, Panaji, 1984)
22. M. Almeida, *A Description of Konkani* (2003)
23. O.J. Gomes, Old Konkani language and literature: the Portuguese role, Konkani Sorospot Prakashan, Goa (1999)
24. H. Sak, A.W. Senior, F. Beaufays, Long short-term memory recurrent neural network architectures for large scale acoustic modeling (2014)

DAMBNFT: Document Authentication Model through Blockchain and Non-fungible Tokens

Uday Khokhariya, Kaushal Shah, Nidhay Pancholi, and Shambhavi Kumar

Abstract Hard copies of documents can easily be forged, resulting in the decrease of credibility of issuing institutions and unfair use of forged documents by certain individuals as well. The model that we propose aims to authenticate documents through the use of blockchain technology, non-fungible tokens, and interplanetary file system. When a document is stored on the blockchain, a non-fungible token is created, which contains the unique address of the issuing institution and the hash of the document itself. The ownership of this token is then transferred to the document holder by the corresponding issuing authority. In this way, when someone wants to verify the authenticity of the document, they can use the address mentioned in the token to trace back the creator. If the document's hash differs from the one stored in the token, we know that the document has been altered. Even when unauthorized users are successful in adding the forged documents to the blockchain, they will not have the same unique signature as that of the authorized institution. The proposed model allows autonomous authentication of documents using public blockchain technology.

Keywords Non-fungible token (NFT) · Blockchain technology · Interplanetary file system · Smart contracts · Document forgery

1 Introduction

The issue of hardcopy document forgery is still prevalent. In South Africa, a series of cases have been reported where academic documents are forged [1]. There have been multiple cases of document forgery in India too, where documents related to land and academics have been forged [2]. The digitized records can be preserved for a longer time [3]. Hence, it is essential to combine the digitization property along with the validation criteria to create an efficient document authentication system. The system we propose works using the concepts of blockchain technology, non-fungible tokens (NFTs), and smart contracts. Blockchain is a shared ledger that has the property of

U. Khokhariya · K. Shah (✉) · N. Pancholi · S. Kumar
Pandit Deendayal Energy University (PDEU), Gandhinagar, India
e-mail: shah.kaushal.a@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_33

being averse to malicious changes. All the participants of the network have access to the records, but no participant has the ability to alter the contents of the blockchain. Therefore, once a document is added to the blockchain, it is infeasible to alter [4]. Smart contracts are pieces of code, or in simpler words, a certain set of rules which are executed whenever a document is added to the blockchain.

The system also uses non-fungible tokens (NFTs), which is an irreplaceable and immutable token stored on the blockchain [5]. These tokens are used to represent ownership and also for unique identification of an entity. The blockchain is a distributed system, and all the contents of the blockchain are stored on every node of the network. Therefore, storing the complete documents on the blockchain becomes impractical as a lot of storage space will be needed. As all the documents have to be stored on each node of the network, the method also causes a lot of redundancy. To solve that problem, our proposed system utilizes an interplanetary file system (IPFS) as the off-chain storage network [6]. In this system, documents are stored at certain locations in the network, and a content identifier (CID) which can be used to locate the document, is stored on the blockchain.

2 Background

The development of blockchain technology has paved the way for managing documents in a decentralized manner. If we consider an example, certificates are being stored on ledgers, and these ledgers constitute a blockchain [7]. This illustrates that the blockchain technology can be effectively used for managing documents. The evolution of technology would create another challenge of protecting the documents which are present in the ledgers.

Blockchain was first proposed in 1991. After 2014, another domain of blockchain was explored, and "blockchain 2.0" was born [8]. Eventually, several other implementations of blockchain such as the Ethereum blockchain have developed scope for various implementations [9]. Blockchain is a technology that is presently undergoing extensive research [10–13]. A blockchain is made up of several blocks or ledgers that are linked together. These ledgers are strict, and each one is linked by a cryptographic hash key which ensures that the transaction's validity is preserved [12].

The non-fungible token (NFT) provides a unique authority over a digital asset which is made on the concept of Ethereum blockchain. These are present on public ledgers and are unalterable [13]. NFT has been in discussion since 2018. The NFT market gained its popularity with the development of cryptoart. It is considered to be the most secured and easy method being used in transfer of digital assets [14]. The security of copyright has become a major challenge. Therefore, it is important to secure the rights of the owner. This is where NFTs come in picture because they ensure the ownership authority and are unalterable [15]. The digital assets such as land titles and college degrees can be converted to digital assets, stored in ledgers and are secured using NFTs [2, 3, 7].

Smart contracts are basically predefined code which is stored in blockchain. They get executed when the prior specified conditions are fulfilled [16]. The concept of smart contract was first implemented for storing bond agreements. These agreements are directly signed between the two parties which are involved. Smart contracts result in a successful transaction, providing access to encrypted content, changing of names on business titles, etc. [17]. This reduces the need of trusted third party. These scripted codes are stored on ledgers of blockchain. The script is assigned an address to store it on the ledger of blockchain. To execute the scripted code, one has to call for the assigned address [18]. As described in [19]; smart contracts provide a fast, secured, and transparent transfer of digital assets.

3 Related Work

Blockchain has created a lot of opportunities and has provided us with a scope of development of decentralized and secured applications [20, 21]. The ledgers are minted with NFTs, and each NFT has one defined unique address. The NFTs are unalterable as discussed in [22]. As mentioned in [23], one such breakthrough is evident as Blockcerts. Blockcerts are blockchain-based credentials with a decentralized and trustless verification mechanism. Users have direct access to permanent, transparent, and sustainable evaluation and management tools for learner credentials. Personal encrypted credentials allow users to customize lifetime learning paths and tailor education to their own beliefs and requirements. The original answer to Internet credentialing was badges. Mozilla's open digital badges have established a de facto worldwide standard, and the specifications are still available for free. They are accessible through e-portfolios and social media. Even if badges are genuine, they become worthless if issuers stop hosting them [23]. TrustChain is a three-layered trust management architecture that uses a consortium blockchain to log supply chain contracts and dynamically award trust and reputation scores based on these interactions. TrustChain is the first worldwide jewelry industry partnership to use blockchain technology to trace diamonds and precious metals from mine to completed jewelry and finally into the hands of the consumer [24]. The TrustChain architecture is based on the concept of agents conducting transactions with one another. The exchange of data, the purchase or sale of products, and the movement of funds are all examples of transactions in the real world [25].

4 The Proposed Model

Digital documents need an authentication system to track the creation timestamp, content changes, ownership, publishers, and validity [26]. Blockchain helps in tracing all these steps in the form of an immutable ledger. It prevents tampering of documents and helps verify the document issuer using digital signature methodology

[27]. Our model is divided into two spaces to minimize blockchain nodes' storage and computational requirements, namely on-chain and off-chain storage.

4.1 Participants Overview

- **Blockchain Network (BN)**: It is the leading network of the system where all NFTs will be stored. Smart contracts deployed on BN will be carrying out all the processes in the form of transactions. Further, the transaction log would also be stored on the BN, which gives credibility to every transaction happening on the network (Fig. 1).
- **Interplanetary File System (IPFS)**: It is a decentralized peer-to-peer framework which stores files in the form of hashes [6]. It works on the principle of content-based addressing and would act as off-chain storage in our model. IPFS network provides availability, reliability, and security of stored data [28]. To ensure synchronization of generated hashes, both blockchain and IPFS network would be required to use the same hashing algorithm to identify any document uniquely in the system.
- **Document Authority (DA):** The concept of DA is derived from [29]. DAs are responsible for creating, maintaining, and updating the required documents on BN and IPFS. They sign the documents and carry out the issuing process. Since all the documents are stored on a public blockchain network, it allows any institution or an individual to be a part of the system and register itself as DA.
- **Document Holder (DH)**: Individuals having ownership of particular documents are regarded as DHs. DH needs to hold a secured copy of the document with

Fig. 1 Outline of the participants of the proposed system

him/her to safeguard the document from getting lost in case the IPFS network gets offline or most participants log out from the network. A DA can also act as DH when the institution itself requires secured storage of documents.

- **Document Reviewer (DR)**: When a DH claims the ownership of his/her owned document, DR verifies the claim by checking the authenticity of the claim. Since all the documents lie as NFTs in the public domain, DR does not need permission from the network to perform validation checks.
- **Public Keys Record-keeper (PKR)**: Whenever an institution or individual requests to become a DA, its corresponding address and public key are maintained by PKR. PKR is also responsible for verifying the identity of registered DAs to keep the network safe from identity imitation threats. It would act as a central ledger for all DAs that are a part of the network to ensure transparency and accountability of their signed documents.

4.2 Design Structure

- **Incentive Providence**: Documents issued by DAs occupy storage space on the IPFS network. The storage contributors to this network need to be incentivized to encourage active participation and keep the required number of copies available in the system. DA provides incentives on a short-term basis as agreed upon by the DH. After the expiration time of the agreement, DH needs to remunerate the IPFS network for as long as documents are required to be a part of the system. This process allows more participants to be a part of the system and allows IPFS participants to discard inactive documents after they no longer need the storage requirements. The distribution of incentives will be managed by the IPFS network, where they will be proportionally allocated according to the active time and storage space provided by individual participants (Fig. 2).
- **Membership Registration**: Whenever an entity wants to issue documents on the blockchain, they would need to register themselves with PKR. Users would need to confirm the claimed identity in order to join the network. PKR would be responsible for verifying the user's identity to secure the network from identity impersonation threats. PKR will generate a unique address for each requesting partaker. Public key and address are stored on the central ledger of PKR, whereas private key remains secured with the user. After joining the network, the user is known as DA.
- **NFT Creation**: DA carries out the process of converting required documents to NFT. Hash is used as data for creating the NFT while the original document would be replicated and distributed within the IPFS nodes. Generated hash of each document remains the same on IPFS and BN. When NFT gets successfully created on the blockchain and stored on the IPFS network, its ownership is transferred to its corresponding DH. NFT creation and ownership transfer take place through smart contracts which ensure all transactions get stored on the blockchain for traceability.

Fig. 2 Sequence diagram of the proposed document authentication system

- **Claim Validation**: When DH claims an NFT on the blockchain, DR carries out the validation process and checks the legitimacy of the claim by verifying the digital signature. Authentication requirements are fetched from PKR. Once the signature has been verified, it can also be sent to the DA for testimony. When a claim gets authenticated, the required document is fetched from the IPFS network using the hash stored on the NFT.

5 Discussion and Future Work

As mentioned in [30], the documents related to land entitlements are not properly maintained across India. The documents are not verified and result in forgery. Similarly, making of duplicate certificates is another major issue. With the advent of COVID-19, a lot of data had to be collected and analyzed. Once the vaccination of people started, India had shifted to digital storage of records. These digital assets were present in fragments. Another major challenge faced was to remove this data fragmentation. The government of Gujarat developed "Anveshan," a connector database to overcome data fragmentation, track data alteration, and to track dynamic data [31].

The document authentication model discussed here aims to prevent forgery of documents and identify the institution that issued the document through the use of blockchain and non-fungible tokens (NFTs). As the document issuer is being recognized through the NFT, forged documents can be easily identified based on the authenticity of the digital signature. The model also provides an added advantage in the case of documents that expire after a specific period. The initial transaction fees required for converting the document to NFT and storing it on the interplanetary file system (IPFS) are paid by DA on a short-term basis. However, if DH wants to renew the document, he/she needs to reward the IPFS nodes; otherwise, the document

would be removed from the IPFS network. This characteristic of document removal from the IPFS allows the storage space to be freed up for the new documents to be added.

6 Conclusions

It becomes challenging to identify the authenticity of a physical copy of any document. Many methods have been implemented which focus on finding anomalies in the document, but none is always successful. The model we propose uses blockchain and non-fungible tokens to identify each document uniquely on the network. Whenever a document is issued by its corresponding authority, it is added on the blockchain in the form of a token which contains the hash of that particular document. Interplanetary file system acts as the off-chain storage where the document can be retrieved using its specific hash. The digital signature of the stored token ensures the trust, reliability, and authenticity of the issuing institution. Any person can join the network as an issuer and add documents to the blockchain, but the digital signature of the verified issuing institution cannot be replicated, and hence, the document could be easily identified as being forged. The model achieves its purpose of eliminating document forgery along with being transparent, decentralized, and immutable.

References

1. N. Dlamini, S. Mthethwa, G. Barbour, Mitigating the challenge of hardcopy document forgery, in *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)* (IEEE, 2018), pp. 1–6
2. D.K.S. Puri, Study of a copied forgery. J. Secur. Adm. **3**(2), 79–87 (1980)
3. M. Baldi, F. Chiaraluce, MigelanKodra, L. Spalazzi, Security analysis of a blockchain-based protocol for the certification of academic credentials. arXiv:1910.04622 (2019)
4. P. Snow, B. Deery, J. Lu, D. Johnston, P. Kirby, Factom white paper. Retrieved from Factom: https://www.factom.com/devs/docs/guide/factom-white-paper-1-0 (2014)
5. U.W. Chohan, Non-fungible tokens: blockchains, scarcity, and value. In *Critical Blockchain Research Initiative (CBRI) Working Papers* (2021)
6. M. Naz, M, A. F.A. Al-zahrani, R. Khalid, N. Javaid, A.M. Qamar, M.K. Afzal, M. Shafiq, A secure data sharing platform using blockchain and interplanetary file system. Sustainability**11**(24), 7054 (2019)
7. O.S. Saleh, O. Ghazali, M.E. Rana, Blockchain based framework for educational certificates verification, in *Studies, Planning and Follow-up Directorate. Ministry of Higher Education and Scientific Research, Baghdad, Iraq. School* (2020)
8. M. Ulieru, Blockchain 2.0 and beyond: adhocracies, in *Banking Beyond Banks and Money* (Springer, Cham, 2016), pp. 297–303
9. A. Bogner, M. Chanson, A. Meeuw, A decentralised sharing app running a smart contract on the ethereum blockchain, in *Proceedings of the 6th International Conference on the Internet of Things* (2016), pp. 177–178

10. C. Lepore, M. Ceria, A. Visconti, U.P. Rao, K.A. Shah, L. Zanolini, A survey on blockchain consensus with a performance comparison of PoW, PoS and pure PoS. Mathematics**8**(10), 1782 (2020)
11. V. Patel, K.S. FenilKhatiwala, Y. Choksi. A review on blockchain technology: components, issues and challenges, in *ICDSMLA 2019* (Springer, Singapore, 2020), pp. 1257–1262
12. D. Shrier, W. Wu, A. Pentland, Blockchain & infrastructure (identity, data security). Massachusetts Inst. Technol.-Connect. Sci. **1**(3), 1–19 (2016)
13. L. Kugler, Non-fungible tokens and the future of art. Commun. ACM **64**(9), 19–20 (2021)
14. A. Guadamuz, The treachery of Images: non-fungible tokens and copyright. Available at SSRN 3905452 (2021)
15. R. Raman, B. Edwin Raj, The world of NFTs (non-fungible tokens): the future of blockchain and asset ownership, in *Enabling Blockchain Technology for Secure Networking and Communications* (IGI Global, 2021), pp. 89–108
16. Z. Zheng, H.-N.D. ShaoanXie, W. Chen, X. Chen, J. Weng, M. Imran, An overview on smart contracts: Challenges, advances and platforms. Future Gener. Compu. Syst. **105**, 475–491 (2020)
17. M. Bartoletti, L. Pompianu, An empirical analysis of smart contracts: platforms, applications, and design patterns, in *International Conference on Financial Cryptography and Data Security* (Springer, Cham, 2017), pp. 494–509
18. K. Christidis, M. Devetsikiotis, Blockchains and smart contracts for the internet of things. IEEE Access **4**, 2292–2303 (2016)
19. Z. Allam, On smart contracts and organisational performance: a review of smart contracts through the blockchain technology. Rev. Econ. Bus. Stud. **11**(2), 137–156 (2018)
20. G. Rathee, A. Sharma, R. Kumar, R. Iqbal, A secure communicating things network framework for industrial IoT using blockchain technology. Ad Hoc Netw. **94**, 101933 (2019)
21. P. Pant, R. Bathla, S.K. Khatri., A model to implement and secure online documentation using blockchain, in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)* (IEEE, 2019), pp. 174–178
22. J. Santos, K.H. Duffy, A decentralized approach to blockcerts credential revocation, in *A White Paper from Rebooting the Web of Trust V* (2018)
23. M. Jirgensons, J. Kapenieks, Blockchain and the future of digital learning credential assessment and management. J. Teacher Educ. Sustain. **20**(1), 145–156 (2018)
24. L.E. Cartier, S.H. Ali, M.S. Krzemnicki, Blockchain, chain of custody and trace elements: an overview of tracking and traceability opportunities in the gem industry. J. Gemmol. **36**(3) (2018)
25. P. Otte, M. de Vos, J.A. Pouwelse, TrustChain: a sybil-resistant scalable blockchain. Future Gener. Comput. Syst **107** (2017). https://doi.org/10.1016/j.future.2017.08.048
26. A. Ahmad, S.B. Maynard, G. Shanks, A case analysis of information systems and security incident responses. Int. J. Inf. Manage. **35**(6), 717–723 (2015)
27. H. Stančić, Model for preservation of trustworthiness of the digitally signed, timestamped and/or sealed digital records (TRUSTER preservation model), *InterPARES Trust Project* (2018), pp. 1–42
28. S. Krishnan, V.E. Balas, E. Golden Julie, H.R. Yesudhas, S. Balaji, R. Kumar (eds.), *Handbook of Research on Blockchain Technology* (Academic Press, 2020)
29. S. Yao, J. Chen, K. He, R. Du, T. Zhu, X. Chen, PBCert: privacy-preserving blockchain-based certificate status validation toward mass storage management. IEEE Access**7**, 6117–6128 (2018)
30. N. Gupta, M.L. Das, S. Nandi, LandLedger: blockchain-powered Land property administration system, in *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)* (IEEE, 2019), pp. 1–6
31. J.C. Liu, S. Banerjee, A. Bhagi, A. Sarkar, S.D. BhriguKapuria, V. Sethuraman, S. Patil, S. Banerjee, Blockchain technology for immunisation documentation in India: findings from a simulation pilot. Lancet Glob. Health **9**, S22 (2021)

A Novel Hybrid Translator for Gujarati to Interlingual English MTS for Personage Idioms

Jatin C. Modh and Jatinderkumar R. Saini

Abstract Gujarat is one of the states in the western part of India, and the Gujarati language is the official language of Gujarat. The Gujarati language is more than 700 years old and is spoken by more than 55 million people around the world. A machine translation system (MTS) is needed for the communication between people knowing different languages. Idioms are used in almost all the languages. An idiom is a phrase or an expression whose meaning does not necessarily relate to the literal meaning of its individual words. The idiom generally means something different than what is directly conveyed by its individual words. Translation of idioms for any language, as with Gujarati idioms, into any other language is a challenging task. All existing MTS including Google Translate and Microsoft Bing fail to translate Gujarati idioms suitably. In the current paper, a particular category of Gujarati idioms, called the personage idioms, has been treated for detection from the input text and translation into the English language. We have deployed a dictionary-based algorithm and context-based search, respectively, for idioms having one and multiple meanings. This is a first of its kind work in the world. From a broad technical view point, this is an application of Gujarati to interlingual translation for the MTS sub-domain of natural language processing (NLP). The interlingual language is considered to be English for the present research work.

Keywords Gujarati · Interlingual · Machine translation system (MTS) · Natural language processing (NLP) · Personage idioms

J. C. Modh
Gujarat Technological University, Ahmedabad, India

J. R. Saini (✉)
Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India
e-mail: saini_expert@yahoo.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_34

1 Introduction

The Gujarati language is the official language in the state of Gujarat, India, and it is also the official language in the union territory of Dadra and Nagar Haveli, Daman, and Diu. The Gujarati language is more than 700 years old, and it is the sixth most widely spoken language in India as per 2011 data. Gujarati is the 26th most widely spoken language in the world based on the number of local speakers. Gujarati is the fastest growing Indian language in Canada, America, and many other countries [1].

In the Gujarati language, idioms play a very important role. An idiom is a term whose meaning is different from the literal meaning of the words it is made of. An idiom is a phrase whose collective meaning is different than its literal meaning. Idioms are part of Gujarati culture. Idioms are used frequently by the Gujarati people to express feelings. Elders use idioms in day-to-day interaction, and this way idioms are passed to the new generation.

In the twenty-first century, people are migrating from one country to other countries for study, business, and earnings. Gujarati people are business-oriented, and they are ready to migrate to foreign countries. For communication between different language people, a machine translation system is required. English is the universal language and widely accepted language in the region of Gujarat for the purpose of migration. Many machine translation systems are available in the market. But, all are facing issues related to the translation of idioms.

1.1 Gujarati Personage Idioms

Idioms are found in the Gujarati language like any other language. Idioms are terms that are not to be taken exactly. Personage idiom represents a dramatic, fictional, or historical character in a play or other work. It represents a person or human individual [2]. For example, બાબરોભૂત'babaro bhut' is a Gujarati bigram personage idiom, and its actual meaning in Gujarati is દેખાવમાં બિહામણી વ્યક્તિ'dekhavamam bihamani vyakti'; its meaning in English is derived as 'scary guy in appearance'. Another example પોપાંબાઈ'popambai' is Gujarati unigram personage idiom, and its actual meaning in Gujarati is શિથિલ રાજ્યકર્તા'shithila rajyakarta'; its meaning in English is derived as 'lazy ruler'. Gujarati personage idioms are misunderstood by other language readers as they do not know the culture, history, and popular characters of Gujarat and Gujarati language [3]. The focus of this paper is on such Gujarati personage idioms.

The rest of the paper is organized as follows: Sect. 2 represents the literature review related to translation of Gujarati text and Gujarati idioms translation; Sect. 3 covers the methodology including data collection and the steps of proposed model; Results and analysis are discussed in Sect. 4; finally, conclusion, limitations, and future work are described in Sect. 5.

2 Related Literature Review

The scope of this paper is the translation of Gujarati text containing Gujarati personage idioms into English or any other language. Technically, using the parlance of the domain of natural language processing (NLP), this type of translation can be called translation from Gujarati to interlingual language.

Google Translate [4] is a multilingual machine translation system developed by Google Company. It supports 109 different languages for machine translation. Google Translate uses the Google Neural Machine Translation system. Google Translate also translates the Gujarati language into English or any other language [5]. It is a very popular translator but not able to translate Gujarati idioms correctly.

Microsoft Translator [6] is a multilingual machine translation system developed by Microsoft. It supports 90 different languages for machine translation. Microsoft Translator uses Microsoft cognitive services integrated with other Microsoft products. Microsoft Translator also translates Gujarati language text into English or any other language [7]. It also fails to deal with the translation of Gujarati idioms.

Goyal and Sharma [8] described a neural machine translation system for the Gujarati–English news. They tested multilingual neural machine translation models. They tested and advocated that for low-resource Gujarati–English language pairs; multilingual neural MT models are more effective than individually trained MT models.

Saini and Modth [9] implemented a dictionary-based machine translation system GIdTra for translating only Gujarati bigram idioms into English. They classified bigram idioms on the basis of their possible meanings in the English language.

Modh and Saini [10] discussed various machine translation approaches for the Gujarati language. The researchers have implemented a context-based machine translation system (MTS) for translating Gujarati bigram and trigram idioms into the English language [11]. They used IndoWordNet for finding context and improving machine translation of Gujarati idioms [12]. They detected all possible forms of Gujarati idioms from the input text using diacritic and suffix-based rules and by dynamic phrase generation technique [13].

Sen et al. [14] depicted a constrained news translation system for Gujarati–English language pairs. They used original parallel corpus acquired from back-translation of monolingual data and trained transformer-based neural machine translation system. They did not use any external data and advocated to improve the BLEU score.

Agrawal et al. [15] introduced a multilingual parallel idiom dataset of 2208 idioms for English and other seven Indian languages including the Gujarati language. They illustrated sentiment analysis and machine translation and advocated significant improvement in handling idioms without using the idiom dataset. They concluded that phrase-based SMT is better than NMT to handle idioms. They used English idiomatic sentences as a source and converted them into Indian languages.

Based on this literature review and study on machine translation of Gujarati language, all researchers are facing problems in dealing with the translation of idioms.

All researchers had issues in handling idiomatic text of any language. No researchers have analyzed specifically Gujarati personage idioms in detail.

3 Methodology

In the Gujarati language, 3200 distinct idioms are available [13]. Personage idioms are other exceptional idioms used for the character in a literary work or a history, a play, a novel, etc. [3]. Personage idioms are used for important or strange or any distinguished person.

3.1 Data Collection

Overall, 88 Gujarati personage idioms were collected from different sources [16, 17]. Personage Gujarati idioms are stored in the idiom database. Tables 1 and 2 show the two different snapshots of idiom database containing personage Gujarati idioms. Idiom database contains fields like id, idiom, gmeaning (Gujarati meaning), emeaning (English meaning), gcwords (Gujarati context words), and popularity field. Id denotes sequence number; idiom column stores personage idiom; gmeaning stores the actual Gujarati meaning of particular personage idiom; emeaning stores the derived English meaning of particular personage idiom. Field gcwords stores the Gujarati context words to identify the correct gmeaning from the database. Field popularity stores the integer number. Both the fields Gcwords and popularity are needed only when the Gujarati personage idiom has more than one meaning [12]. Table 1 shows the snapshot of database containing single-meaning Gujarati personage idioms, whereas Table 2 shows the snapshot of database containing Gujarati personage idioms having more than one meaning. In Table 1, gcwords and popularity fields are not shown as they are having NULL and 0 values, respectively, but they are used for the actual execution of proposed model.

3.2 Proposed Model

MySQL was used to create and maintain the idiom database for storing Gujarati personage idioms, while PHP was used as a scripting language. XAMP cross-platform was used for the local Webserver to implement the proposed model, while Python 3.7 was used for executing the Python code. Notepad + + was used as an editor for writing Python code and PHP code. Googletrans free and unlimited Python library were used for the implementation of Google Translate API [18, 19]. The algorithmic representation of the steps followed toward systematic methodology is presented below:

Table 1 Snapshot of database contaıننg sınگle-meanıng Gujaratı personage idıoms

Id (1)	Idiom (2)	Transliteration of (2) (3)	Gmeaning (4)	Transliteration of (4) (5)	Emeaning (6)
1	ઊંટવૈદ	untavaida	લાયકાત વગરનો વૈદ	layakat vagarano vaida	an unqualified doctor
2	કડકો	kadako	ખિસ્સે તદ્દન ખાલી	khisse taddana khali	having no money
3	કુબ્જા	kubja	બેડોળ સ્ત્રી	bedol stri	mis-shapen woman
4	કોડાફાડ	kodaphad	આખાબોલી વ્યક્તિ	akhaboli vyakti	outspoken person
5	ગુંદરિયો	gundariyo	ચિટકી રહેનાર વ્યક્તિ	chitaki rahenar vyakti	clingy person
6	ઝોડ	jhoda	વળગાડ જેવી ત્રાસજનક વ્યક્તિ	valagad jevi trasajanak vyakti	annoying person
7	ઢ	dha	મંદબુદ્ધિ	mandabuddhi	incompetent person
8	પ્રજ્ઞાચક્ષુ	prajnachakshu	અંધજન	andhajan	blind person
9	લલ્લુ	lallu	મૂર્ખ વ્યક્તિ	murkha vyakti	stupid person
10	લુખ્ખો	lukhkho	ખોટી ગુંડાગીરી કરનાર	khoti gundagiri karanar	grabing money by bossiness
11	હંસલો	hansalo	આત્મા	atma	soul
12	રાધા	radha	પ્રેમિકા	premika	sweetheart
13	અક્કલનો બારદાન	akkalano baradan	મૂર્ખ માણસ	murkha manas	foolish man
14	ધોબીનો કૂતરો	dhobino kutaro	ઠામઠેકાણા વગરની વ્યક્તિ	thamathekana vagarani vyakti	person with no whereabouts
15	બારમો ચંદ્રમા	baramo chandrama	દુશ્મનાવટ	dusmanavata	animosity
16	બીકણ બિલાડી	bikana biladi	ખૂબ બીકણ વ્યક્તિ	khub bikan vyakti	very timid person
17	મંથરા	manthara	કાવતરાખોર અને કાન ભંભેરણી કરનાર સ્ત્રી	kavatarakhor ane kanabhambherani karanar stri	conspiratorial woman
18	તાડકા	tadaka	અતિશય ઊંચી સ્ત્રી	atishaya unchi stri	extremely tall woman
19	છાપેલું કાટલું	chapelum katalum	નામચીન	namachin	notorious person
20	ચમચો	chamacho	ખુશામત કરનાર વ્યક્તિ	khushamat karanar vyakti	flatterer
21	કુંભકર્ણ	kumbhakarna	ઊંઘણશી વ્યક્તિ	unghanashi vyakti	drowsy person
22	નારદ	narad	બંને પક્ષોને લડાવી મારનાર	banne pakshone ladavi maranar	one who makes persons quarrel
23	શકુની	shakuni	મુત્સદી અને કપટી વ્યક્તિ	mutsadi ane kapati vyakti	diplomat and insidious person
24	દિલ્હીનો શાહુકાર	dilhino shahukar	મોટો ઠગ	moto thaga	rogue person

(continued)

Table 1 (continued)

25	ભૂખડી બારસ	bhukhadi baras	કંગાળ અને ખાઉંખાઉની દાનતવાળું	kangal ane khaumkhauni danat valum	poor and voracious
26	બાબરો ભૂત	babaro bhut	દેખાવમાં બિહામણી વ્યક્તિ	dekhavamam bihamani vyakti	scary guy in appearance
27	બલિનો બકરો	balino bakaro	કોઈના કૂડા કરતૂતનો ભોગ બનતી વ્યક્તિ	koina kuda karatutano bhoga banati vyakti	scapegoat
28	અંગૂઠા છાપ	angutha chhap	અભણ કે નિરક્ષર	abhana ke nirakshar	illiterate person
29	જૂનો જોગી	juno jogi	અનુભવી	anubhavi	experienced person
30	સહદેવ જોશી	sahadev joshi	પૂછ્યા વિના ઉત્તર ન આપે તેવી વ્યક્તિ	puchhya vina uttara na ape tevi vyakti	a person who does not answer without asking
31	દશમો ગ્રહ	dasamo grah	કનડગત કરનાર જમાઈ	kanadagat karanar jamai	harassment by son-in-law
32	પોપાંબાઈ	popambai	શિથિલ રાજ્યકર્તા	shithil rajyakarta	lazy ruler
33	ચારસો વીસ	charaso vis	ઠગ	thaga	rogue
34	પાંચમી કતારિયો	panchami katariyo	દેશદ્રોહી વ્યક્તિ	deshadrohi vyakti	traitor person

- Step1: Overall, 88 Gujarati personage idioms were collected.
- Step2: As a preprocessing activity, extraneous spaces and characters were removed from the collected idioms.
- Step3: Idiom database for Gujarati personage idioms was created with the fields shown in Table 2.
- Step4: Accept input Gujarati text from the user containing Gujarati personage idiom(s).
- Step5: The proposed algorithm searches for personage idioms from the input text by comparing them with the idiom field of the database.
- Step6: Proposed algorithm replaces the Gujarati personage idiom (idiom) with the actual Gujarati meaning text (gmeaning) using the database. Also, note that if a personage idiom has more than one meaning, then Gujarati context words (gcwords field) and popularity field are used to decide the suitable meaning of a specific personage idiom [12]. Other input Gujarati text remains the same. This intermediate output contains literal Gujarati text only.
- Step7: The proposed algorithm translates intermediate Gujarati literal text into English language using Google Translate API.
- Step8: Display input Gujarati text, intermediate Gujarati text output, and final English translation.

Table 2 Snapshot of database contaınıng multiple-meanıng Gujaratı personage idıoms

Id (1)	Idiom (2)	Transliteration of (2) (3)	Gmeaning (4)	Transliteration of (4) (5)	Emeaning (6)	Gcwords (7)	Popularity (8)
1	નંગ	nang	ભાગ અથવા ટુકડો	bhag athava tukado	piece	અંશ અદદ અનુરાગ અભિપ્રાય અરથ અર્ણવ અર્થ અવક્રય આધ આશય ઇશક ઇશ્ક ઉચિત ઉદ્દેશ એટલે કંદર કિંમત કિલો કીમત કેટલા કોતર ખંડ ખરીદ ખો ગહ્વર ગુફા ગુહા ગોપુર ગ્રાહક ચિત્તવૃત્તિ જવાહિર જાયજ ટુકડો ડઝન તાત્પર્ય તોલ તોલમાપ થાય દમોડ દર દરવાજો દરિ દામ દુવાર દ્વાર ધોરણ નંગ નગીન નગીનો નેડો પણ પૃથ્વીગૃહ પૈસા પ્યાર પ્રણય પ્રમાણ પ્રમાણે પ્રાઇસ પ્રીતિ પ્રેમ બખોલ બદલો બાકું બારણું બિલ ભાગ ભાવ ભાવતાલ ભાવાર્થ ભોણ મણ મણિ મત મતલબ મનોભાવ મનોભાવના મનોવિકાર મનોવૃત્તિ મનોવેગ મફત માનસિક ભાવ માને માપ મુજબ મુનાસિબ મૂલ મૂલ્ય મોંઘુ મોલ મોલભાવ યથાર્થ યોગ્ય રતન રત્ન રાગ રૂપિયા રેટ લગન વજન વળતર વાજબી વિવર વેચાણ વેપારી વ્યાજબી શાક શાકભાજી સંગત સંબદ્ધ સનેડો સસ્તું સાર સોદો સ્નેહ હેત હેતુ	2
2	નંગ	nang	મૂર્ખ કે ખંધી વ્યક્તિ	murkha ke khandhi vyakti	a word indicating stupidity	અક્કલ અણઘડ અધમ અનિર્મલ અનિષ્ટ અનુગ અનુચર અનુચિત અપભ્રંશ અપભ્રંશિત અપવિત્ર અપુનિત અપ્રશંસનીય અપ્રિય અભદ્ર અભિચર અભિસર અયોગ અયોગ્ય અવદ અવિનયી અવિમલ અવિશુધ્ધ અવિહિત અશિષ્ટ અશુચિ અશુદ્ધ અશુધ્ધ અશ્લાઘ્ય અશ્લીલ અસંસ્કારી અસભ્ય અસાધુ અસ્વચ્છ આચારભ્રષ્ટ આજ્ઞાંકિત આદમ આદમી આધાર આધીન ઇન્સાન ઉદ્ધત કપટી કેવો ખરાબ ખરાબ રીતે ખલ ગંદુ ગંદું ગંધારૂ ગાંડા ગ્રામ્ય ચાકર છે જંગલી જન જેવું જેવો ટેકો તદ્દન તાબેદાર દાસ દુષ્ટ દૂષિત દોષ્યુક્ત ધૂર્ત નઠારૂં નર નરસું નિંદનીય નિંદાત્મક નિંદ્ય નીચ નોકર પરાધીન પરિચારક પાપી પામર બગડેલું બદ બદતર બિલકુલ બીભત્સ બુદ્ધિ બૂરું બેઅદબ બેકાર ભરોસો ભૂંડું ભ્રષ્ટ મનુષ્ય મલિન માણસ માનવ માનવી માનુષ મૂર્ખ મેલુ મ્લાન લઠ્ઠ વાહિયાત વિકારી વિકૃત વિચારશક્તિ વ્યક્તિ શખ્સ શઠ સાવ સેવક હલકટ	1
3	નંગ	nang	રત્ન	ratna	a crystallized jewel	અંગુરી અંગુલિ અંગુશ્ત અગુ અનુભાવ અપરાધ અપલક્ષણ અરુણ અર્ણવ અવગુણ અસર આંગળી આંગળું આગસ આમદ આમદાની આમિષ આય આવક આશ્રવ ઇચ્છા ઇનકમ ઉંગલ ઉંગલી ઉપલબ્ધિ ઉલઝન એબ કમાણી કલ્યાણ કામયાબ કુટેવ કુલક્ષણ કૃતકૃત્ય કૃતાર્થ કેતુ ખરાબી ગુત્થી ગુનો ગુરૂ ગ્રહ ચંદ્ર ચાંદી છાપ જય જવાહિર જાતક જીત જીર્ણવજ્ર જુર્મ જ્યોતિષ તાસીર દુર્ગુણ દોષ ધારણ નંગ નગીન નગીનો નફો નીલમ નુકસાન પંકજરાગ પદ્મરાગ પદ્મરાગ મણિ પદ્મરાજ પન્ના પહેરવું પાપ પોખરાજ પ્રતાપ પ્રભાવ પ્રશ્ન પ્રાપ્તિ પ્લેનેટ ફતેહ ફાયદો બગાડ બરકત બુધ મંગલ મંગળ મણિ મળતર મસલો માણિક માણેક મોતી યશ યાકૂત રંગ રક્તોપલ રતન રત્ન રાતો મણી રાશિ રાહુ લબ્ધિ લાભ લાલ મણી વિજય વિજયશ્રી વીંટી વૃદ્ધિ વૈક્રાંતમણિ શનિ શુક	3

(continued)

Table 2 (continued)

4	બલા	bala	આ ફત	aphat	trou ble	શોણરત્ન સંસર્ગ સફળ સફળતા સમસ્યા સમૃદ્ધિ સમૃદ્ધિજીત સાર્થક સિદ્ધ સિદ્ધિ સુખ સૂર્યમણિ સોનું સ્પર્શ હિત હીરો હેતુસિદ્ધિ અંતરાય અક અઘ અડચણ અયોગ અરિષ્ટ અવગ્રહ અવરોધ અવસન્નતા અવસન્નત્વ અશર્મ અસુખ અસુવિધા આડખીલી આદીનવ આપત્તિ આપદ આપદા આફત આભીલ આરામ ઉદ્વેગ કઠોર કષ્ટ કાચો ખરડો કુદરતી સ્થિતિ ક્લેશ ખરબચડું ખાડાટેકરાવાળું ખોડખામીવાળું ગભરામણ ગર્દેશ ચેન ડોળિયું તકલીફ તસ્દી તોફાની થાક ઉતારવો દુ દુખ દોચન નડતર નિરાંત પરેશાની પીડા બાધા ભીડ મવાલી મુશ્કેલી મુસીબત રાહત લાગણીવિહોણું વિક્ષેપ વિક્ષેપણ વિઘ્ન વિપત્તિ વિપદ વિપદા વિલોપ વિશ્રાન્તિ વિશ્રામ વૃજિન વ્યથા વ્યાકુલતા વ્યાકુલપણું વ્યાકુળતા વ્યાઘાત શામત સંકટ સંતાપ સુફ્રન સુખ હરકત હેરાનગત	1
5	બલા	bala	વણ આ મં ત્રિ તમ હેમા ન	vana amantri t mahem an	unin vited gues t	અચાનક અતિથિ અન્ન અન્નગ્રહણ અભ્યાગત અશન આગંતુક આગત આમંત્રણ આવકાર આવાહન આહાર આહ્વાન ઈજન કંકોતરી કંકોત્રી કહેણ કુટુંબ કુટુંબકબીલો કુળ ખટલો ખબર ખાવું ખોરાક ઘર ઘરખટલો જમણ જમવાનું જમવું જાણ જાહેરાત તલબ તેડું નિમંત્રણ નિમંત્રણપત્ર નોતરું નોતરૂં પરિવાર પરોણો ફેમિલી ભક્ષ્ય પદાર્થ ભોજન ભોજનકાર્ય મહેમાન મિજમાન રસોઈ રહેઠાણ રાંધણું. વગર વસ્તાર શિકાર સંદેશો સમાચાર સુવિધા સૂચના	2

4 Results and Analysis

For the output authentication, three linguistics were consulted and asked to perform human intelligence task (HIT). As part of HIT, they were clearly explained the motivation behind the research as well as the requirement of precision of their translation. All the experts were having a qualification of at least post-graduation in Gujarati. Two experts were males, while one was Female. For example, input text is રામુ પાંચમી કતારિયો છે'Ramu panchami katariyo chhe'; the intermediate output is is રામુ દેશદ્રોહી વ્યક્તિ છે'Ramu desadrohi vyakti chhe', and the final English translation by the proposed algorithm is 'Ramu is a traitor person'. Table 3 shows the translation of Gujarati language text containing Gujarati personage idioms into the English language by the proposed model. The instances presented in Table 3 make use of imaginary and fictitious names Raju and Ramu to exemplify the input text used for testing the proposed model. We submit that like any other NLP test data used by other researchers, these names are used by us just for improving the comprehension of the presentation. In Table 3, input text and intermediate output both are in Gujarati language, but transliteration is shown for understanding only.

The same Gujarati text is given to Google Translate and Microsoft Bing Translator for the comparison of the proposed model results. If Gujarati input text રામુ પાંચમી કતારિયો છે'ramu panchami katariyo chhe' is given to Google Translate and Microsoft Bing Translator, Google Translate translates it into English as 'Ramu is fifth rower' incorrectly; Microsoft Bing Translator also translates it into English as 'Ramu is the fifth Qatari' incorrectly; Google Translate and Microsoft

Table 3 Translation of Gujarati personage idioms into English language by proposed model

Sr. No	Input text	Intermediate output	Final English translation output
1	રાજુ ઊંટવૈદ છે. raju untavaida che	રાજુ લાયકાત વગરનો વૈદ છે. raju layakat vagarano vaida chhe	Raju is an unqualified doctor.
2	રામુ ચારસો વીસ છે. ramu charaso vis chhe	રામુ ઠગ છે. ramu thag chhe	Ramu is a thug.
3	રાજુ લુખ્ખો છે. raju lukhkho chhe	રાજુ ખોટી ગુંડાગીરી કરનાર છે. raju khoti gundagiri karanar chhe	Raju is a false bully.
4	રામુ પાંચમી કતારિયો છે. ramu panchami katariyo chhe	રામુ દેશદ્રોહી વ્યક્તિ છે. ramu deshdrohi vyakti chhe	Ramu is a traitor person.
5	મનુષ્યના મરણ પછી હંસલો જતો રહે છે. manushyana maran pachhi hansalo jato rahe chhe	મનુષ્યના મરણ પછી આત્મા જતો રહે છે. manushyana maran pachhi atma jato rahe chhe	The soul is going after the death of a human being.
6	રામુ બીકણ બિલાડી છે. ramu bikan biladi chhe	રામુ ખૂબ બીકણ વ્યક્તિ છે. ramu khub bikan vyakti chhe	Ramu is a very timid person.
7	રામુ તો રાજુનો ચમચો છે. ramu to rajuno chamacho chhe	રામુ તો રાજુનો ખુશામત કરનાર વ્યક્તિ છે. ramu to rajuno khushamat karanar vyakti chhe	Ramu is a man who flatters Raju.
8	રામુ અભ્યાસમાં સાવ ઢ છે. ramu abhyasamam sav dha chhe	રામુ અભ્યાસમાં સાવ મંદબુદ્ધિ છે. ramu abhyasamam sav mandabuddhi chhe	Ramu is very dull in his studies.
9	રાજુ તો સાવ નંગ છે, તેના પર ભરોસો ના રાખી શકાય. raju to sav nang chhe, tena par bharoso na rakhi shakay	રાજુ તો સાવ મૂર્ખ કે ખંધી વ્યક્તિ છે, તેના પર ભરોસો ના રાખી શકાય. raju to sav murkh ke khandhi vyakti chhe, tena par bharoso na rakhi shakaya	Raju is a very stupid person, he cannot be trusted.

Bing could not recognize the Gujarati personage idioms, so both could not able to translate it correctly, whereas the proposed model first detects Gujarati personage idiom પાંચમી કતારિયો'panchami katariyo' from the input and replaces it with Gujarati actual meaning દેશદ્રોહી વ્યક્તિ'desadrohi vyakti', so intermediate output is રામુ દેશદ્રોહી વ્યક્તિ છે'Ramu desadrohi vyakti chhe' and then using Google Translate API to translate the intermediate Gujarati result into the English language correctly as 'Ramu is a traitor person'. Table 4 shows the comparison of the English translation of Gujarati text containing Gujarati personage idiom(s) by Google Translate, Microsoft Bing Translator, and the proposed model.

Some interpretations, results, and ambiguities faced during the experimentation.

1. Proposed algorithm provides most nearby English translation of Gujarati personage idioms; whereas Google Translate and Microsoft Bing Translator do not provide correct translation of Gujarati personage idioms. For example, for Gujarati sentence રાજુ ઊંટવૈદ છે'raju untavaida chhe', Google Translate translates in English as 'Raju is a camel doctor' that is absolutely wrong one;

Table 4 Snapshot of comparisons of proposed model with translation by Google Translate and Microsoft Bing

Sr. No	Input text	English translation		English Translation (Proposed Algorithm)	English Translation by HIT
		By Google Translate	*By Microsoft Bing Translator*		
1	રાજુ ઊંટવૈદ છે. raju untavaida che	Raju is a camel doctor.	Raju is a camel vaid.	Raju is an unqualified doctor.	Raju is an unqualified doctor.
2	રામુ ચારસો વીસ છે. ramu charaso vis chhe	Ramu is four hundred and twenty.	Ramu is four hundred and twenty.	Ramu is a thug.	Ramu is a thug.
3	રાજુ લુખ્ખો છે. raju lukhkho chhe	Raju is lurking.	Raju is a lukhkho.	Raju is a bully.	Raju is a bully.
4	રામુ પાંચમી કતારિયો છે. ramu panchami katariyo chhe	Ramu is the fifth rower.	Ramu is the fifth Qatari.	Ramu is a traitor person.	Ramu is a traitor person.
5	મનુષ્યના મરણ પછી હંસલો જતો રહે છે. manushyana maran pachhi hansalo jato rahe chhe	Laughter is gone after the death of a human being.	After the death of a man, the goosegoes go away.	The soul is going after the death of a human being.	The soul is leaving after the death of a human being.
6	રામુ બીકણ બિલાડી છે. ramu bikan biladi chhe	Ramu is a timid cat.	Ramu is a beaded cat.	Ramu is a very timid person.	Ramu is a very timid person.
7	રામુ તો રાજુનો ચમચો છે. ramu to rajuno chamacho chhe	Ramu is Raju's spoon.	Ramu is the spoon of Ramu.	Ramu is a man who flatters Raju.	Ramu is a man who flatters Raju.
8	રામુ અભ્યાસમાં સાવ ઢ છે. ramu abhyasamam sav dha chhe	Ramu is in the study.	Ramu is very good at studies.	Ramu is very dull in his studies.	Ramu is very dull in his studies.
9	રાજુ તો સાવ નંગ છે, તેના પર ભરોસો ના રાખી શકાય. raju to sav nang chhe, tena par bharoso na rakhi sakay	Raju is completely naked, he cannot be trusted.	Raju is very naked, he cannot be trusted.	Raju is a very stupid person, he cannot be trusted.	Raju is a very stupid person, he cannot be trusted.

Microsoft Bing Translator translates it in English as ‘Raju is a camel vaid’ that is also wrong one, whereas proposed algorithm translates as ‘Raju is an unqualified doctor’ that is very similar to correct translation ‘Raju is an unqualified doctor’.

2. Common Gujarati idioms are n-grams where $n = 2$ to 9 [11], whereas our research shows that Gujarati personage idioms are either 1-g or 2-g only. The number of unigram or 1-g Gujarati personage idioms is 58, and the number of bigram or 2-g personage idioms is 30. Hence, the total number of Gujarati personage idioms is 88.
3. Most of the Gujarati personage idioms are found to have single meaning except નંગ‘nang’ and બલા‘bala’; નંગidiom has three different literal meanings depending on the surrounding context words as (1) મૂર્ખ કે ખંધી વ્યક્તિ‘murkh

ke khandhi vyakti' a stupid person (2) ભાગ અથવા ટુકડો 'bhag athava tukado' piece (3) રત્ન 'ratna' a crystallized jewel, whereas personage idiom બલાhas two different literal meanings depending on the surrounding context words as (1) વણઆમંત્રિત મહેમાન 'vana amantrit maheman' uninvited guest (2) આફત 'aphat' trouble.

4. Literal meaning of Gujarati personage idioms are shown in Gujarati and English both the languages in the idiom database. It provides useful personality insight about the particular personage. For example, for Gujarati personage idiom શકુની 'shakuni', actual meaning in Gujarati is મુત્સદી અને કપટી વ્યક્તિ 'mutsadi ane kapati vyakti', and its English meaning is derived as 'diplomat and insidious person', so proposed model helps other language users about the correct meaning as well as popular character meaning.
5. An idiom database was created with 88 distinct personage idioms and tested with different Gujarati sentences containing personage idiom(s). The proposed algorithm can able to find out personage idiom within Gujarati text by searching it in the idiom database if it exists in the idiom database.

5 Conclusion, Limitations, and Future Work

All the existing MTSs, including Google Translate and Microsoft Bing Translator, face the problem of accuracy in translating the Gujarati idioms into English or any other language. The proposed algorithm detects all the Gujarati personage idioms present in the input Gujarati text and translates them, as an intermediate step, into Gujarati itself but with semantically correct meaning. This Gujarati intermediate output contains no Gujarati personage idioms. Finally, intermediate Gujarati output is translated into the interlingual language. We have considered English as an interlingual language for the present work by using Google Translate API.

We deployed a hybrid approach comprising of a dictionary-based algorithm and context-based search for the translation of personage Gujarati idioms. These two approaches have been used, respectively, for idioms with literal meanings and the idioms with multiple idioms. To the best of our study of related literature, this is a first of its kind implementation for accurate translation of personage idioms of any language in world. Based on the results obtained from the proposed algorithm, it is advocated that the proposed algorithm is promising and can be extended for all the Gujarati idioms available in the Gujarati language. The proposed algorithm is specifically useful for translating Gujarati personage idioms into the English language accurately. Furthermore, all Gujarati personage idioms can be translated into any other language by implementing the proposed model. The limitation of the proposed model is in differentiating between the Gujarati personage idiom and the same word used as a noun. For the future work, we have already started to work on a special case where the same idiom is used as a noun as well as in the idiomatic sense simultaneously.

The proposed model opens the pathway for translating Gujarati idioms into any other language. We believe that the proposed model is free from the geographic boundaries on the translator, and the proposed work will definitely drive other researchers in the area of MTS in specific and NLP in general.

References

1. Wikipedia, Gujarati language. https://en.wikipedia.org/wiki/Gujarati_language . Accessed 6 Aug 2021
2. Merriam Webster Dictionary, Definition of personage. https://www.merriam-webster.com/dictionary/personage. Accessed 11 Aug 2021
3. Your Dictionary, Personage meaning. Available online https://www.yourdictionary.com/personage. Accessed 11 Aug 2021
4. Google Translate, Google translate. Google Corporation Ltd. Available online https://translate.google.co.in/. Accessed 11 Aug 2021
5. Wikipedia, Google translate. Available online https://en.wikipedia.org/wiki/Google_Translate. Accessed 11 Aug 2021
6. Bing Microsoft Translator, Microsoft bing. Microsoft Corporation Ltd. Available online https://www.bing.com/translator. Accessed 11 Aug 2021
7. Wikipedia, Microsoft translator. Available online https://en.wikipedia.org/wiki/Microsoft_Translator. Accessed 11 Aug 2021
8. V. Goyal, D. Sharma, The IIIT-H Gujarati-English machine translation system for WMT19, in *Association for Computational Linguistics; Proceedings of the Fourth Conference on Machine Translation (WMT), vol 2: Shared Task Papers (Day 1)* (2019), pp. 191–195
9. J.R. Saini, J.C. Modh, GIdTra: a dictionary-based MTS for translating Gujarati Bigram Idioms to English, in *Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC) 22–24 December 2016* (2016), pp. 192–196, Available online https://ieeexplore.ieee.org/document/7913143
10. J.C. Modh, J.R. Saini, A study of machine translation approaches for Gujarati language. Int. J. Adv. Res. Comput. Sci. **9**(1), 285–288 (2018). Available online: https://ijarcs.info/index.php/Ijarcs/article/download/5266/4497
11. J.C. Modh, J.R. Saini, Context based MTS for translating Gujarati Trigram and Bigram Idioms to English, in *2020 International Conference for Emerging Technology (INCET), Belgaum, India* (2020), pp. 1–6. https://doi.org/10.1109/INCET49848.2020.9154112. Available online https://ieeexplore.ieee.org/document/9154112/
12. J.C. Modh, J.R. Saini, Using IndoWordNet for contextually ımproved machinetranslation of Gujarati Idioms, Int. J. Adv. Comput. Sci. App. (IJACSA) **12**(1) (2021). https://doi.org/10.14569/IJACSA.2021.0120128
13. J.C. Modh, J.R. Saini, Dynamic phrase generation for detection of Idioms of Gujarati language using diacritics and suffix-based rules. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **12**(7) (2021). https://doi.org/10.14569/IJACSA.2021.0120728
14. S. Sen, K.K. Gupta, A. Ekbal, P. Bhattacharya, IITP-MT system for Gujarati-English news translation task at WMT 2019, in *Association for Computational Linguistics; Proceedings of the Fourth Conference on Machine Translation (WMT), Volume 2: Shared Task Papers (Day 1)* (2019), pp. 407–411
15. R. Agrawal, V.C. Kumar, V. Muralidharan, D. Sharma, No more beating about the bush: a step towards Idiom handling for Indian language NLP, in *LREC 2018, Eleventh International Conference on Language Resources and Evaluation.* Available online http://www.lrec-conf.org/proceedings/lrec2018/pdf/1052.pdf. Accessed 12 Aug 2021
16. GujaratiLexicon, Gujaratilexicon.com, Available online http://www.letslearngujarati.com/about-us. Accessed 11 Aug 2021

17. Rudhiprayog Ane Kahevat Sangrah, Published by Director of languages, Gujarat State, Gandhinagar (2010)
18. Googletrans 3.0.0, Free google translate API for Python. Translates totally free of charge. Available online https://pypi.org/project/googletrans/. Accessed 10 August, 2021
19. Googletrans 3.0.0 Documentation, Googletrans: free and unlimited google translate API for Python. Available online https://py-googletrans.readthedocs.io/en/latest/. Accessed 10 Aug 2021

Alleviation of Voltage Quality-Related Issues: A Case Study of Bahir Dar Textile Share Distribution System

Dessalegn Bitew Aeggegn, Yalew Werkie Gebru, Takele Ferede Agajie, Ayodeji Olalekan Salau, and Adedeji Tomide Akindadelo

Abstract This paper presents a unified power quality conditioner (UPQC)-based compensation approach for voltage quality improvement in Bahir Dar textile share company distribution system. The existing system was modeled and simulated to evaluate its performance. The base case waveform shows that power quality is deteriorating in the factory. The voltage sag and swell at the base case are 65% and 140%, respectively, with voltage unbalance and harmonic content for a duration of 0.1 s. After using the proposed fuzzy logic controller (FLC)-based UPQC, both voltage sag and swell are alleviated by 100% and the waveform is maintained and kept at 1.0 p.u.

Keywords Fuzzy logic controller · Voltage sag · Voltage quality · Harmonics

1 Introduction

Power quality (PQ) and reliability of the supply are vital for the proper running of industrial processes that have critical, sensitive, and nonlinear loads. The major concerns of the industries are short duration voltage variations such as voltage sag, voltage swell, and short interruptions. Due to the advancement of semiconductor technology, these factories use a wide range of sensitive and nonlinear equipment and loads during production phases. If this short duration power quality (PQ) issues lasts for a long period, they may cause production loss, power outage, and other costs [1–5].

D. B. Aeggegn (✉) · Y. W. Gebru · T. F. Agajie
Department of Electrical and Computer Engineering, Debre Markos University, Debre Markos, Ethiopia
e-mail: dessalegnbitew29@gmail.com

A. O. Salau
Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti, Nigeria

A. T. Akindadelo
Department of Basic Sciences, Babcock University, Ilishan Remo, Nigeria

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_35

The root causes of voltage quality-related issues are majorly uneven distribution of single-phase loads among the three-phase supply in distribution grids, technical and non-technical faults, large loads and motor drives, and the utility grid [6–10].

The rest of this paper is structured as follows. Section 2 presents the proposed method. The experimental results and discussion are presented in Sect. 3, and Sect. 4 concludes the paper.

2 Methodology

2.1 System Design

In this study, Bahir Dar textile share company distribution system was considered as the case study. To alleviate power quality problems such as voltage sag, voltage swell, voltage unbalance, and harmonic distortion, a fuzzy logic-based UPQC is proposed. System modeling and simulations were performed using MATLAB/Simulink. The system was simulated in scenarios of without and with UPQC at steady-state conditions. The overall diagram is consisting of all existing features, and UPQC components (like series transformers, controllers, VSI converters, filters, and others complimentary features) are shown in Fig. 1, while all the system parameters used are presented in Table 1.

The load variety was made using a load survey during the assessment. Finding the load parameters was not required, due to the available clear load data such as power factor, real and reactive power values from the company. These values are presented in Table 2.

2.1.1 Sizing of the System

The system was sized based on the load survey and assessment. The sizing of both series and shunt APFs is sized in parallel for maximum demand. This supports the reactive power and partial compensation of active power. The system has a supply grid/feeder of 15 kV and MDB of 400/230 V and the total load of 4546.10 kW with a power factor of 0.65. Table 2 presents the summary of the load assessment.

Figure 2 shows the series APF subsystem block diagram that is used to improve voltage quality in Bahir Dar textile share company distribution system. Figure 3 shows the shunt APF which regulates the voltage by improving the voltage unbalance.

Fig. 1 Proposed system network with the use of the proposed UPQC

Table 1 System parameters

S/No	System model	Parameter description
1	Source	Three phase, 15 kV rms (L-L), 15 kV base, $R = 1.5\ \Omega$, $L = 2.4\ \mu$H
2	VSI converter	IGBT based, 3 arm, 6 pulse, 4 kHz
3	PLL	Three phase, discrete, $K_p = 20$, $K_i = 50$, Sampling time = 50 μs
4	LC filter	$L = 3$mH, $C = 1\ \mu$F
5	DC link	$C = 400\ \mu$F, $V_{DCref} = 700$ V
6	PI controller	$K_i = 12$, $K_p = 0.5$
7	Transformer	One three phase: 7.5 MVA, 15/0.4 kV, 50 Hz;
		Three single-phase: 2.5 MVA, 0.4/0.4 kV, 50 Hz
8	Load	Linear: three phase, 400 V, 50 Hz, 2.49 MW, 1.34 MVAr
		Single phase, 230 V, 50 Hz, 0.276 MW, 0.14 MVAr
		Nonlinear: Three phase, 4000 V, 50 Hz, 0.592 MW, 0.32 MVAr
		Single phase, 230 V, 50 Hz, 0.592 MW, 0.32 MVAr

Table 2 Load summary from load assessment

		Capacity		
S/No.	Type of load	MVA	MW	MVAr
1	Linear three phase	3.83	2.49	1.34
2	Nonlinear three phase	0.91	0.59	0.32
3	Linear single phase	0.42	0.276	0.148
4	Nonlinear single phase	0.91	0.59	0.32
5	Reserved for future	1	0.6	0.35
6	Total	6.421	4.546	2.478

Fig. 2 Simulink model of the series APF

Fig. 3 Simulink model of the shunt APF

3 Results and Discussion

3.1 Steady-State Simulation Result of Existing System Without UPQC

The power flow was simulated using the latest version of PSAT 2.1.10 toolbox added to MATLAB toolbox directory. The existing system PSAT single line diagram is shown in Fig. 4. Power flow simulation was performed using PSAT to determine the load voltage profile, power loss, load real, and reactive power flow. The results are presented in Table 3.

Fig. 4 PSAT single line diagram of the existing system

Table 3 Network components and simulation results

Buses	10	Number of iterations	4
Transformers	13	Maximum P mismatch (p.u.)	– 0.002
Generators/Grid	1	Maximum Q mismatch (p.u.)	0.008
Loads	9	Power rate [SMVA]	10

Table 4 Power flow results (all values in p.u.)

Bus	Voltage	Angle	P_{gen}	Q_{gen}	P_{load}	Q_{load}
Bus_1	1.0 (if…)	0	0.403	0.224	0	
Bus_2	0.97179	−0.07061	0	0	0.0208	0.0068
Bus_3	0.96554	−0.03511	0	0	0.0208	0.0068
Bus_4	0.94554	−0.03511	0	0	0.0208	0.0068
Bus_5	0.93854	−0.03511	0	0	0.0208	0.0068
Bus_6	0.9243	−0.01746	0	0	0.0622	0.0017
Bus_7	0.8596	−0.11437	0	0	0.1418	0.1075
Bus_8	0.8929	−0.07061	0	0	0.0208	0.0068
Bus_9	0.9256	−0.03499	0	0	0.0622	0.0017
Bus_10	0.9156	−0.11437	0	0	0.1418	0.1075

As shown in Table 4, only buses 7 and 10 are below the required voltage according to the IEEE standard, because the p.u voltage value is less than 0.95. Hence, this indicates that the PQ issue is caused by the sensitive, nonlinear loads and improper grounding and non-uniform distribution of single-phase loads.

Power loss is experienced in some sections, even when the overall voltage status is good. This is because of the adaptive under voltage, the overloading of transformers (i.e., not using of parallel operation of transformers) and non-uniform distribution of single-phase loads in the company's distribution system. Therefore, the line flows (P_{flow} and Q_{flow}) can be negative from the source and load bus and vice versa. This is shown in Table 5.

The existing system PSAT power flow result is almost the same with the MATLAB/Simulink scope display reading of voltage, real, reactive power parameters discussed in Sect. 3.2.

Table 5 Line flow (all values are in p.u.)

From bus	To bus	Line	P_{flow}	Q_{flow}	P_{Loss}	Q_{Loss}
Bus_8	Bus_2	1	−0.02208	−0.068	8e−04	0.0020
Bus_9	Bus_2	2	−0.055	−0.002	1e−04	0.0019
Bus_10	Bus_2	3	−0.14018	−0.1075	0.0083	0.0165
Bus_2	Bus_2	4	−0.02208	−0.068	8e−04	0.0157
Bus_7	Bus_2	5	−0.1418	−0.1075	0.00083	0.0165
Bus_3	Bus_2	6	−0.014	−0.0034	2e−04	0.00039
Bus_3	Bus_2	7	−0.014	−0.0034	2e−04	0.00039
Bus_4	Bus_2	8	−0.014	−0.034	2e−04	0.00039
Bus_4	Bus_2	9	−0.014	−0.034	2e−04	0.00039
Bus_5	Bus_2	10	−0.014	−0.034	2e−04	0.00039
Bus_5	Bus_2	11	−0.014	−0.034	2e−04	0.00039
Bus_6	Bus_2	12	−0.028	−8e−03	0	5e−04
Bus_6	Bus_2	13	−0.028	−8e−03	0	5e−04

3.2 *MATLAB/Simulink Simulation Result of the Existing System*

3.2.1 Existing System Steady-State Load Side Simulation Result

The existing system load side steady-state simulation result is shown in Fig. 5, and the results show the extent to which both the load voltage and current has been distorted.

Fig. 5 Load side (base case) system simulation result

This is generally due to PQ issues in the system. Figure 5 shows both the voltage and current waveforms. It is clear that harmonics are detected in the system with real power of about 4.5 MW and the reactive power of about 2.3 MVAR. This shows that the PSAT simulation result is similar to the MATLAB/Simulink result.

A. Simulation of Voltage Sag and Voltage Swell

A fault was applied to the existing system, and simulations were performed. A result of 0.65 p.u sag was achieved as shown in Fig. 6.

Voltage swell was also simulated using a step signal and breaker for tripping OFF heavy loads for a duration of 0.1 s. The resultant waveform is shown in Fig. 7. The instantaneous tripping OFF loads of the system help to examine the capability of the

Fig. 6 Voltage sag simulation result without using UPQC

Fig. 7 Voltage swell simulation result without using UPQC

base case system to withstand voltage swell. As seen in Fig. 7, the base case systems voltage swell reaches up to 1.2 p.u for a duration of 0.1 s. The application of UPQC is used to reduce this effect.

3.3 Dynamic Analysis and Simulation Result with UPQC

The waveforms of the system after the insertion of UPQC are shown in Fig. 8. The results show that both the grid and load voltages are perfect at 1.0 p.u, and the load current magnitude is purely sinusoidal at 0.4 p.u. This therefore shows that UPQC has improved the systems PQ.

Voltage unbalance alleviation was achieved with the use of UPQC. The waveform in Fig. 9a shows the extent the system was affected by non-uniform distribution of single-phase loads among the three-phase supply. The results show that all the three-phase voltages are under the standard limit and unbalanced. But after the insertion of UPQC as shown in Fig. 9b, the load voltage becomes balanced at 1.0 p.u.

Fig. 8 Waveforms of compensated grid and load side voltage and current

Fig. 9 Load voltage unbalance **a** without and **b** with

4 Conclusion

This paper presented a fuzzy logic controller (FLC)-based UPQC for the alleviation or regulation of voltage sag/swell and voltage unbalance. The proposed UPQC model mitigated all voltage quality-related issues in the investigated distribution system. The voltage sag and swell at the base case were 65% and 140%, respectively, with voltage unbalance, harmonic content, and other issues for a duration of 0.1 s. After

the introduction of the FLC-based UPQC, both voltage sag and swell are mitigated to 100% and the variations are maintained and kept at 1.0 p.u. The overall results show that the UPQC is effective in regulating the voltage quality-related issues with the ability to inject partial real power aimed at minimizing power losses.

References

1. T. Kang, S. Choi, A.S. Morsy, P.N. Enjeti, Series voltage regulator for a distribution transformer to compensate voltage sag/swell. IEEE Trans. Ind. Electron. **64**(6), 4501–4510 (2017)
2. D.B. Aeggegn, A.O. Salau, Y. Gebru, Load flow and contingency analysis for transmission line outage. Arch. Electr. Eng. **69**(3), 581–594 (2020). https://doi.org/10.24425/aee.2020.133919
3. S.S. Rao, P.S.R. Krishna, S. Babu, Mitigation of voltage sag, swell and THD using dynamic voltage restorer with photovoltaic system, in *International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)* (2017), pp. 1–7
4. M. Addisu, A.O. Salau, H. Takele, Fuzzy logic based optimal placement of voltage regulators and capacitors for distribution systems efficiency improvement. Heliyon **7**(8), e07848 (2021). https://doi.org/10.1016/j.heliyon.2021.e07848
5. R.K. Patjoshi, V.R. Kolluru, K. Mahapatra, Power quality enhancement using fuzzy sliding mode based pulse width modulation control strategy for unified power quality conditioner. Int. J. Electr. Power Energy Syst. **84**, 153–167 (2017)
6. M. Prasad, A.K. Akella, S. Das, S. Kuma, Voltage Sag swell and harmonics mitigation by solar photovoltaic fed ZSI based UPQC. Int. J. Renew. Energy Res. **7**(4), 1646–1655 (2017)
7. T.F. Agajie, A.O. Salau, E.A. Hailu, Y.A. Awoke, Power loss mitigation and voltage profile improvement with distributed generation using grid-based multi-objective harmony search algorithm. J. Electr. Electron. Eng. **13**(2), 5–10 (2020)
8. A.O. Salau, J. Nweke, U. Ogbuefi, Effective implementation of mitigation measures against voltage collapse in distribution power systems, in *Przegląd Elektrotechniczny* (2021), pp. 65–68. https://doi.org/10.15199/48.2021.10.13
9. A.O. Salau, Y.W. Gebru, D. Bitew, Optimal network reconfiguration for power loss minimization and voltage profile enhancement in distribution systems. Heliyon **6**(6), e04233. https://doi.org/10.1016/j.heliyon.2020.e04233
10. J.N. Nweke, A.O. Salau, U.C. Ogbuefi, "Bus voltage sensitivity index based approach against voltage collapse in distribution systems," in *International Conference on Decision Aid Sciences and Application (DASA)* (2021), pp. 1062–1066. https://doi.org/10.1109/DASA53625.2021.9682359.

Customer Perception, Expectation, and Experience Toward Services Provided by Public Sector Mutual Funds

R. Malavika and A. Suresh

Abstract In recent years, mutual funds have gained popularity as a means of ensuring the financial security. Mutual funds have benefited families in capitalizing on India's wealth while also contributing to the country's economic record. As knowledge and understanding of mutual funds grow, most people are reaping the benefits of investing in them. This study is focused on the various customer perceptions, expectations, and experiences toward services provided by public sector mutual funds. This study mainly analyzed the main features and major factors affecting the public sector mutual fund's customers' experience, expectation, and their perception. A quantitative study was undertaken through a structured questionnaire to analyze major factors that attracted customers toward public sector mutual funds, and based on the analysis, it was found that customers are satisfied because they have got better responsiveness from public sector mutual funds to customer complaints. As part of the study, many customers' opinions and reviews have been collected. This project will help public sector mutual funds to develop new strategies to create a more customer-friendly approach for increasing the sales and improving the customer satisfaction and their experience with public sector mutual funds.

Keywords Customer expectation · Customer perception · Customer experience · Mutual funds

1 Introduction

A mutual fund is a trust that pools the reserve funds of several investors with a common budgetary goal according to the website. The money is subsequently invested in various sorts of protections by the reserve director, based on the plan's goal. These could range from debentures to foreign exchange trading products. The

R. Malavika (✉) · A. Suresh
Department of Management, Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, Kerala, India
e-mail: ramachandranmalavika@gmail.com

A. Suresh
e-mail: suresh@am.amrita.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_36

most cost-effective investment for the average individual is a mutual fund, which allows them to invest in a better, carefully controlled portfolio with minimum effort. In a mutual fund, the asset director, who is otherwise called the portfolio administrator, exchanges the asset's fundamental protections, acknowledging capital increases or misfortunes, and gathers the profit or interest pay. The venture proceeds are then given to the individual financial backers. The worth of a portion of the common asset, known as the net resource esteem per share, net asset value (NAV) is determined everyday dependent on the all-out worth of the asset isolated by the quantity of offers at present given and remarkable. A trust that accumulates the reserve funds of numerous financial backers with a shared monetary purpose is called a common asset. This sum of money is being put into a project with the goal of reaching a specified goal. The joint reserve obligation is common, and the asset, for example, has a place among all financial backers. The money is subsequently invested in capital market assets such as offers, debentures, and various protections. A mutual fund is a form of financial instrument that pools money from several different participants in order to achieve a shared objective. The funds are then invested in a variety of assets.

A mutual fund is a type of investment vehicle that gives small investors access to a diverse selection of assets, securities, and other forms of protection. Diversification decreases risk because all stocks cannot move in the same direction in the same proportion around the same time. Investors are given units determined by the amount of money they have placed in mutual funds. Unit holders are people who invest in mutual funds. When an investor buys units in a mutual fund, he becomes a part owner of the fund's assets in the same proportion as the sum he contributes to the corpus (the total amount of the fund). A mutual fund investor is also known as a unit holder or a mutual fund shareholder. The growth of mutual funds is of three stages: The Unit Trust of India, that had an ultimate resource of Rs. 6700 crores at the end of 1988, was the lone player throughout the first stage, which took place from 1964 and 1987. The next stage occurred from 1987 and 1993, during which time eight funds were established (6 by banks and one each by LIC and GIC). By the end of 1994, the resources available under management had grown to 61,028 crores, with 167 programs in place. The third stage started with the passage of private and unfamiliar areas in the mutual fund industry in 1993. Kothari Pioneer Mutual Fund was the principal fund to be set up by the private sector in a relationship with an unfamiliar fund. As toward the finish of the monetary year 2000 (31st walk), 32 funds were working with Rs. 1,13,005 crores as complete resources under administration. As of the end of August 2000, there were 33 funds under management, with 391 plans and capital totaling Rs 1,02,849 crores. In 1993, the Securities and Exchange Board of India (SEBI) issued a comprehensive guideline that defined the architecture of Mutual Fund and Asset Management Companies. In 1993 and 1994, a few private segments mutual funds were distributed. From that point on, the number of private participants has rapidly increased. In India, there are 34 mutual fund companies that manage a total of 1,02,000 crores. An investor can create many folios.

1.1 Objectives

- To identify the customer perception and experience about the services provided by public sector mutual funds.
- To know about the preferences of investors toward mutual funds with special reference to public sector mutual funds as well as in other mutual fund companies.
- To evaluate the factors affecting the preference of various schemes.
- To determine the level of investor satisfaction with public sector mutual funds.

1.2 Scope of the Study

The study has got a very significant scope to add the advancement of public sector mutual funds by determining the customers' interests and expectations in terms of portfolio, mode of investment, choice, and rate of return. The research can be used to help the organization make better decisions and prepare its marketing campaign. The analysis will be useful in determining the respondents' general knowledge of mutual funds.

2 Literature Review

Following studies done by researchers on mutual fund have been referred for this study.

For most of the households, mutual funds are a good source of returns. It is especially beneficial to those who have reached retirement age. Ordinary investors are still limited to traditional investments such as gold and fixed deposits. This is due to a shortage of knowledge about how mutual funds operate. In reality, many people who invest in mutual funds do not understand how they function or how to administer them. As a result, organizations that offer mutual funds must provide prospective investors with comprehensive information about mutual funds [1].

India's mutual fund business has a big untapped market. More consumers are relying on low-cost, low-risk professional management of their assets. As more funds enter the business, these companies' strategic marketing strategies are becoming increasingly important for their survival. The success of a mutual fund is contingent on a thorough knowledge of the psychology of small investors. Saving products with varying risk–return combinations compete with MF businesses. Investors have become more vigilant and selective [2].

A mutual fund is the best investment for the ordinary individual since it allows them to invest in a diversified, professionally managed range of assets at a cheap cost. Investors buy units in a mutual fund scheme with a specific investment objective and strategy in mind. The income generated by these assets, as well as any capital gains realized by the plan, is distributed to the scheme's units in proportion to their ownership [3].

Customers of mutual funds are willing to invest in mutual funds if they can guarantee a link to the equity market, given that investing in mutual funds yields higher returns than investing in gold, according to the findings of the study. The primary motivation for developing a mutual fund is that it is tied to the stock market and can guarantee a particular rate of return. Consumers of mutual funds regard them as a superior investment alternative than other investment options available, and they choose to invest in them over gold. Mutual fund investment has been reported to be treated as a channel for investors and is capable of attracting clients [4].

When it comes to making investments, most investors relied on historical mutual fund performance. Prospectuses/newsletters have been the most popular source of information for investors. The total return on mutual fund schemes has been used by the majority of investors to measure their performance. Investors in public sector mutual funds expressed dissatisfaction with the services given, while private sector mutual fund investors cited a lack of awareness as a key issue [5].

The purpose of this study is to determine the level of mutual fund knowledge among investors. Investor perceptions and opinions on topics such as mutual fund types, satisfaction levels, and the function of financial advisors and brokers have all been investigated. Investors' views on elements that entice them to invest in mutual funds, sources of information, shortcomings in mutual fund managers' services, issues facing the Indian mutual fund business, and so on [6].

The benefits of mutual fund investment may be divided into many categories. This includes the safety of money invested in mutual funds, the fund's/favorable scheme's credit rating from recognized credit organizations, full disclosure of all relevant information, and frequent updates. The financial advantages offered by funds/schemes in the form of wealth development, liquidity, return on investment (ROI), early bird rewards, fringe perks, and charge discounts are the next category of financial benefits [7].

3 Research Methodology

3.1 Methodology Adopted

Sampling procedure: The sampling procedure used in this study is judgmental sampling, also called purposive or authoritative sampling.

Sample Size: The sample size is 250 which consists of public sector mutual funds investors in Kerala.

Sampling Unit: These comprise customers of public sector mutual funds in Kerala.

Sampling Area: The area of the research was concentrated mainly in Kerala, India.

3.2 Data Collection

Data collection was done using questionnaires. The questionnaire was framed in an easily understandable way, so the respondents may not have any difficulty in answering them. A total of 18 questions make up the questionnaire. There is also 1 suggestion question to get the opinion and suggestions of the people regarding public sector mutual funds.

4 Analysis and Interpretation

4.1 Customer Perception

- ***Time period that you would like to stay invested***

Interpretation

The above chart shows that 14% of the customers like to stay invested for time period up to 6 months, 33.8% of the customers like to stay invested for time period 6–12 months, 34.8% of the customers like to stay invested for time period from 1 to 5 years, 17.6% of the customers like to stay invested for time period greater than 5 years.

- ***Features that attracted customers to public sector mutual funds***

Interpretation

The goal of the study was to discover the characteristics that drew customers to public sector mutual funds and kept them pleased with their investment. Among the 250 respondents, 58.8% strongly agree, 23.6% agree, 10.8% have neutral opinion, 5.2% disagree, 2% strongly disagree that the low risk is a feature which is better in mutual funds. Among the 250 respondents, 23.2% strongly agree, 60.8% agree, 9.6% have neutral opinion, 4.4% disagree, 2% strongly disagree that the safety of their investment is a feature which is better in mutual funds. Among the 250 respondents, 3.6% strongly agree, 35.6% agree, 29.4% have neutral opinion, 4.0% disagree, 4% strongly disagree that the ease of their investment procedure is a feature which is better in mutual funds. Among the 250 respondents, 24% strongly agree, 48.8% agree, 14.8% have neutral opinion, 11.0% disagree, 0.8% strongly disagree that the returns from their investment as a feature which is better in mutual funds. Among the 250 respondents, 29.6% strongly agree, 42.8% agree, 23.2% have neutral opinion, 2.8% disagree, 1.6% strongly disagree that they can achieve their long-term goals with their investment procedure as a feature which is better in public sector mutual funds.

- ***Type of mutual fund that is preferred***

Interpretation

From the above analysis, 42.8% customers have interested in equity funds, 18% invest in debt fund, 16% invest in income fund,11.6% of customers invest in money market fund, 3.6 invest in ELSS, 4.8 invest in SIP, 1.6% invest in balanced fund and 1.6% in other schemes.

Customer Expectations

- ***Improvements are needed in the service provided by public sector mutual funds***

		Frequency	Percent	Valid percent	Cumulative percent
Valid	Yes	91	36.4	36.4	36.4
	No	159	63.6	63.6	100.0
	Total	250	100.0	100.0	
Total		250	100.0		

Interpretation

From the above chart, 36% of the customers suggest that they need some improvements and 64% of the customers do not prefer any improvements or suggestions for the services of public sector mutual funds. They are satisfied with the current services.

- ***Responsiveness of public sector mutual funds toward mutual fund***

		Frequency	Percent	Valid percent	Cumulative percent
Valid	Strongly agree	96	38.4	38.4	38.4
	Agree	111	44.4	44.4	82.8
	Neither agree nor disagree	37	14.8	14.8	97.6
	Disagree	3	1.2	1.2	98.8
	Strongly disagree	3	1.2	1.2	100.0
	Total	250	100.0	100.0	
Total		250	100.0		

Interpretation

The above graph shows that 38% strongly agree that the public sector mutual funds are highly responsive toward the complaints of the customers, 44.4% agree that the public sector mutual funds are highly responsive, 14.2% neither agree neither agree nor disagree they have a neutral statement toward the responsiveness of the public sector mutual funds and 1.2% disagree and strongly disagree that is the public sector mutual funds is highly responsive toward the complaints of the customers.

4.2 Customer Experience

- ***Investment time period of customers***

Interpretation

According to the graph above, 14.1% of consumers are investing for less than one year, 26.8% for one to three years, 29.2% for three to five years, 20.8% for five to ten years, and 8.8% for more than ten years.

- ***Investment amount of respondents***

Interpretation

The above chart shows that 14.8% of the investors invest up to 25,0000 rupees, 30.0% of the investors invest from 25,0000–50,000 rupees, 27.2% of the investors invest from 50,000–75,0000 rupees, 22.4% of the investors invest from 75,0000–1,00,000 rupees, 5.6% of the investors invest above 1,00,000 rupees.

- ***Level of satisfaction with the services provided***

		Frequency	Percent	Valid percent	Cumulative percent
Valid	Very satisfied	111	44.4	44.4	44.4
	Satisfied	97	38.8	38.8	83.2
	Neutral	28	11.2	11.2	94.4
	Dissatisfied	10	4.0	4.0	98.4
	Very dissatisfied	4	1.6	1.6	100.0
	Total	250	100.0	100.0	
Total		250	100.0		

Interpretation

The above chart shows that 44.4% of customers are highly satisfied with the services provided by public sector mutual funds, 38.8% of customers are satisfied with the

services provided by public sector mutual funds, 11.2% are neutral with their opinion, 4.2% are dissatisfied, and 1.6% of customers are very dissatisfied with the services provided by public sector mutual funds.

- ***Factors prompted you to invest in public sector mutual funds***

For factors that influence characteristics that attracted people to public sector mutual funds, a principal component analysis with varimax rotation yielded the following findings.

KMO and Bartlett's test		
Kaiser–Meyer–Olkin measure of sampling adequacy		0.540
Bartlett's test of sphericity	Approx. Chi-square	9.198
	Df	6
	Sig	0.013

Interpretation
Out of the 250 samples collected, the Kaiser–Meyer–Olkin measure of sampling adequacy is 0.5 so the sample is adequate and has a p value less than level of significance so that the value is significant. There are two components: complaints and services obtained from the factor analysis. 63% of customers agree that they consider low risk as a factor that prompted them to stay invested in public sector mutual funds, 54% customers agree that they consider better service as a factor, 43% agree for solving complaints, and 52% customers agree that they consider financial discipline as a factor that prompted them to stay invested in public sector mutual funds.

5 Findings

- Most of the male customers from age group 4–55 and above 55 invest an amount of 25,000–50,000 and female customers of age group 45–55, and there are less customers at age group below 25 invest above 1,00,000.
- Most customers (31%) become mainly aware about public sector mutual funds through newspapers, journals, articles. 22% were influenced by television, 17.6 were influenced by public sector mutual funds employees and 14.8% were influenced by Internet, friends, or relative toward public sector mutual funds.
- Government employees have more investment from the amount of 25,000–50,000, and homemakers have the least investment of up to 25,000 in public sector mutual funds.
- Investment time period is maximum for 5–10 years with an amount of 50,000–75,000, and 25,000–50,000 is the minimum for a time period of less than 1 year.
- The main feature that attracts customers to public sector mutual funds is reduction in risk and cost from investment time period of 5–10 lakhs, safety principle from

investment time period of 5–10 lakhs, easiness to access from investment time period of 5–10 lakhs, returns on investment from investment period from 1 to 5 lakhs, long-term goals from investment period 5–10 lakhs.

- The customers mostly prefer to stay invested (12%) from 3 to 1 year for an amount of 25–50 thousand. And the least preference of investment (5.2%) is for an amount of 75,0000–1,00,000 from a time period of 3–5 years and 25,000–50,000 for less than 1 year.
- Government employees have mainly attracted by the features of safety principle, easiness to access, achieving long-term goals, reduction in risk and transaction cost as a customer of public sector mutual funds
- Safety principle as the main feature attracted by male and female customers to public sector mutual funds and ease of access as the least attractive feature by male and female customers to public sector mutual funds.
- Low risk was recognized as a major factor that influenced consumers to invest in public sector mutual funds among respondents aged 45–55, while addressing complaints was indicated as a minor factor that influenced customers aged 55 and above to participate in public sector mutual funds.
- The majority of male investors cite low risk as a key element in getting consumers to buy in public sector mutual funds, while female investors cite low risk and financial discipline as key reasons in getting people to engage in public sector mutual funds.
- Customers become aware about public sector mutual funds through newspaper, journal, articles, and public sector mutual funds employee as the mode of information by female respondents mainly in equity and income fund and male customers become aware about public sector mutual funds through public sector mutual funds employees as the mode of information through equity fund by female respondents.
- The greatest investment time period was from 3 to 5 years of investment amount 50,000–75,000 and need to stay updated to 6–12 months. The lowest investment time period was more than 10 years for investment amounts above 1,00,000 and needed to stay updated to 1 to 5 years.
- From the analysis, it is identified that customers are satisfied because they have got better responsiveness from public sector mutual funds to customer complaints.

6 Conclusion

The primary objectives were to determine consumer perceptions and experiences with public sector mutual funds as well as to analyze customer experiences with public sector mutual funds. The secondary goal is to learn about investor preferences for mutual funds, with a focus on public sector mutual funds and other mutual fund companies, as well as to investigate the factors that influence investor preferences for various schemes, as well as to learn about the extent of investor satisfaction with public sector mutual funds and to scrutinize investor satisfaction with service

quality provided by public sector mutual funds. Most of the customer's perception and expectation about the services provided by public sector mutual funds have been met with their experience from the services provided by public sector mutual funds, and also customers were highly satisfied with the services provided through this study and to find out the customer perception and experience about the services provided by public sector mutual funds and the experience obtained from the public sector mutual funds by the customers are satisfied with the services provided. The customers have satisfaction; hence, they have better responsiveness from public sector mutual funds to customer's complaints. Most customers prefer public sector mutual funds because they are satisfied that their perception and expectation have been met with their experience. Customers were extremely satisfied with the services, and they are highly interested to invest in other schemes. So it can be concluded that when there is proper awareness and knowledge about mutual funds then the customer will be able to understand all aspects and also make them understand the benefit. The customers have been influenced by the features that prompted public sector mutual funds and investment factors that prompted throughout their investment and also from service provided by the employees have made a great empire to the customers.

References

1. A. Naik, S.G. Pramod, A study on investors' perception towards mutual funds with due reference to 'SBI Mutual Funds'
2. D.K.P. Shah, A study of consumer activities towards SBI mutual funds company with reference to Nadiad City (2020)
3. S. Saha, M. Dey, Analysis of factors affecting investors' perception of mutual fund investment. IUP J. Manage. Res. **10**(2) (2011)
4. R. Sharma, Investment perception and selection behaviour towards mutual fund. Int. J. Techno-Manage. Res. 1 (2013)
5. N.J. Bhutada, D. Pingale, An exploratory study on perception of customers and consumer behavior towards mutual fund
6. C. Madhavi, Performance and evaluation of public sector mutual funds in India
7. K. Rajesh, N. Goel, An empirical study on investor's perception towards mutual funds. Int. J. Res. Manage. Bus. Stud. 1 (2014)

DigiDrive: Making Driving School Management Effective

Manasi Khanvilkar, Atreya Rastradhipati, and Wricha Mishra

Abstract The Indian driving school system consists of schools ranging from small-scale setups to huge franchises with branches spread nationwide. Irrespective of the scale of the school, it needs to have a management system to handle its student and instructors data to ensure seamless communication. In today's world, technology has transformed how industries handle their data management processes. However, from this study, it was observed that the Indian driving schools follow conventional and manual methods for managing their data which generate possibilities for errors. Furthermore, it also leads to several communication problems among the users. This study explores the data management methods, and the problems faced in driving schools through contextual inquiry, semi-structured interviews, and competitive analysis. Based on the problems identified, design intervention is proposed consisting of an in-vehicle infotainment system linked to a mobile and Web application to provide a seamless management experience.

Keywords Data management methods · IVI systems · Driving schools

1 Introduction

Driving is a critical maneuver that provides an individual with a sense of responsibility, freedom, and confidence. Although, according to World Road Statistics 2018, India ranks 1st in the number of road accident deaths across the 199 countries followed by China and US. As per the WHO Global Report on Road Safety 2018, India accounts for almost 11% of the accident-related deaths in the world. Vehicles driven by untrained and unqualified drivers are a serious traffic hazard and can cause accidents, death, and injuries. Driving without valid license/learner's license accounts for about 15% of accidents [1]. One safety measure to prevent driving accidents is effective driving education of the drivers. Effective training of the drivers, related to the increasing tolerance for roadside stimuli and other road environment-related

M. Khanvilkar (✉) · A. Rastradhipati · W. Mishra
MIT Institute of Design, MIT-ADT University, Pune, MH 412201, India
e-mail: manasikhanvilkar0906@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_37

aspects, helps in creating a positive attitude and safe response among the drivers, e.g., for reacting safely toward the roadside stimuli which in turn reduces the chances of aggressive driving and increases road safety [2]. Driving schools provide driving education to prepare a new driver to obtain a learner's permit or driver's license. Due to increasing awareness around driving education, there has been an increase in the number of driving schools and driving learners. As a result, the driving schools need a management system to manage all the students' and instructors' data. A management system is a set of policies, processes, and procedures used by an organization to ensure that it can fulfill the tasks required to achieve its objectives [3]. The Indian driving school system consists of small-scale setups to huge franchises with branches spread nationwide. There have been several studies concerning the domain of driving schools, e.g., use and effectiveness of driving simulators [4–6]; driving license testing procedures [7]; formation of driving skills [8], etc. However, the previous work deals with the learning methods in driving schools, and there is a lack of research on the data management methods in the domain of driving schools in India. This paper aims to understand different data management problems faced by the driving schools and propose a design intervention to minimize them.

2 Methods

Since the study targeted the driving schools domain as a whole and not a specific user segment, all three stakeholders - driving school managers, driving school instructors, and students were considered. A driving school manager is responsible for the administrative and organizational tasks in a driving school. Driving school instructors are responsible for teaching the students. The students are the users who pursue driving education from the driving schools. Figure 1 shows the methods followed during this study.

A contextual inquiry was performed across 12 driving schools in Pune. Contextual inquiry is a widely used process, to gather field data from users with the aim of understanding who the users are and how they work in their day-to-day basis [9]. The visited driving schools included small-scale setups as well as large franchises. Selection of the driving schools was done through convenience sampling. During the contextual inquiry, 20–30-min-long semi-structured interviews were conducted with 12 driving school managers and 10 driving instructors from those driving schools. The

Fig. 1 Methods used in the study

interview questions were based on attendance management, tracking student progress and fee payments, license procedure, use of technology in teaching, selection of road routes while learning, and their session scheduling process. To understand the system from the students' perspective, 15 students who had attended a driving school were interviewed. The interview questions were based on satisfaction with the teaching methods, communication with the driving school, etc.

In order to design an intervention that aimed to solve the identified problems, a stakeholder's analysis was conducted to identify the role and influence of each stakeholder group in the driving school. Stakeholder analysis aims to evaluate and understand stakeholders from the perspective of an organization, or to determine their relevance to a project or policy [10].

After identifying the problems, a competitor's analysis was conducted to understand the existing solutions in the market. Competitor analysis is a method for identifying the strengths and weaknesses of competing products or services before starting work on prototypes [11].

Since usability testing holds great importance in a product development life cycle, the developed designs were tested with 5 users, who were in the process of learning to drive from a driving school, or had gone through the process in the near past. In this case, pragmatic usability rating by experts (PURE) method was used. PURE is a usability-evaluation method in which usability experts assign one or more quantitative ratings to a design based on a set of criteria and then combine all these ratings into a final score and easy-to-understand visual representation [12].

3 Results

3.1 Contextual Inquiry

A flow diagram (Fig. 2) was created based on the insights from the contextual inquiry to understand the interactions that take place in a driving school. It appears that the manager is an important stakeholder in the system as they carry out the critical tasks of communicating between the students, instructors, and any external entities in the system.

Fig. 2 Flow diagram for driving school managers

3.1.1 Driving School Managers and Instructors

Attendance Recording and Student Data Management Methods
Most schools used attendance cards (75% schools), register books (20% schools), excel sheets (5% schools), and other conventional methods to record student attendance (Fig. 3). The handling of student data (documents, forms, etc.) was done manually by maintaining physical records. On enquiring about the issues faced due to these methods, problems like losing data, difficulty in accessing records, space consumption due to physical records, etc., were mentioned.

Teaching Methods in Driving Schools
Figure 4 shows the teaching methods provided by the driving schools. All the schools offer physical training (driving a real car), which is the most crucial part of the learning process. Whereas, methods like online/offline study material (books, videos on driving, traffic rules, etc.), car physical study (explaining internal functioning of a car), and driving simulators are offered in selective driving schools.

Overview of User Groups in Driving Schools
There were some institutes with a large number of cars, students, and limited managers and instructors. This makes managing the data and sessions of all the students, maintenance of cars, appointment of instructors to students, providing sufficient student details to the instructors highly inconvenient for the managers. It also leads to several miscommunication problems. In case of emergencies, the managers face difficulties to track their vehicles and the respective instructors.

Fig. 3 Attendance recording and data management methods

Fig. 4 Teaching methods in driving schools

Fig. 5 Multiple students in one batch

Fig. 6 Route familiarity

3.1.2 Driving School Students

Physical Training in Batches
From the student interviews, it was observed that 75% of students were provided physical training in batches along with other students, and these batches consisted of 2 to 4 students (Fig. 5).

Locating Driving Routes
Most driving schools teach driving on main roads. Since the traffic, pedestrians, and road conditions cannot be estimated beforehand, the students get stuck on the roads which lead to a reduction of the overall session time, thereby affecting student satisfaction. It was observed that around 40% of students practiced on routes that were not familiar to them (Fig. 6).

3.2 Stakeholder's Analysis

These are the insights from the stakeholder' analysis (Fig. 7).

3.2.1 Driving School Managers

The managers have the highest amount of responsibility among all the stakeholders, since they oversee the administrative and organizational tasks of the entire driving school. Hence, they become critical stakeholders in the intervention.

Fig. 7 Stakeholder's analysis

3.2.2 Driving School Instructors

Even though the instructors do not have responsibilities as critical as the manager, they are important stakeholders in a driving school. They have the primary responsibility of a driving school, imparting driving education. Hence, the intervention should equip the instructors to carry out their primary responsibility with ease.

3.2.3 Students

The students are at the receiving end of the driving school services. As a result, their interaction with the intervention would only include information relevant to them such as progress, performance, fees, etc. Their learning experience with the driving school will influence the overall performance and reputation of the driving schools. Hence, the intervention should keep the students well-informed.

3.3 Competitor's Analysis

After understanding the issues, a competitor's map and a competitor's analysis were conducted to understand the existing solutions in the market. From the competitor's map (Fig. 8), several available solutions for tasks like attendance management,

Fig. 8 Competitor's map

theory training, license registration, etc., were discovered. The competitor's analysis compared the existing popular driving school management platforms. Table 1 shows the key highlights of the competitor's analysis.

Table 1 Competitor's analysis key highlights

Features	Drive Scout	TeachWorks	DSManager	Microsoft Excel
Creating a custom learning plan	Yes	No	No	No
Class scheduling	Yes	Yes	Yes	No
Personal student account	No	Yes	No	No
Manage car schedule	Yes	No	Yes	No
Access for instructors	No	Yes	No	No

4 Intervention

The paper [13], which studies the persistence of the use of paper in e-governance, suggests that, in making the transition from a paper-based system to a new electronic system aimed at replacing or minimizing paper, internal actors responsible for implementing and working with for the new system are asked to abandon well-rehearsed routines and longstanding systems, and transfer their faith to a new and unfamiliar system, establishing new routines or integrate new systems into existing routines.

Hence, the intervention we are proposing to tackle the issues mentioned in the results is a system that involves a user interface suited for each user group. Since the data are updated across all three interfaces, it reduces miscommunication among the users. The three interfaces are as follows:

4.1 In-Vehicle Infotainment (IVI) System

The IVI system will be accessed by the instructors during their physical training sessions (Fig. 9). It aims to limit any interaction before and after the driving sessions to reduce the cognitive load on the users during the driving session. The interface was designed with adherence to the Android Auto guidelines [14]. Some of the important features of this interface are

- **Daily Schedule and Session Summary**—The instructor's schedule will be displayed along with the student details like name, contact number, learning

Fig. 9 IVI system screens

module, etc., after logging in. Once a physical training session is completed, the summary of the session is displayed and is updated on the mobile and Web applications.

- **Real-Time Location Sharing**—The real-time location of the vehicle is shared with the manager on the Web application along with the details of the instructor in the vehicle. Similarly, if another vehicle from the same driving school is detected in the vicinity, it will be displayed on the IVI system.
- **Learning Prompts**—This feature intends to make the learning experience more engaging. It provides real-life driving scenarios on the screen based on the route. The instructor can read out and teach the right approach to deal with such situations. E.g., if the map detects a school nearby, it will show the prompt "what should you do when you see a traffic sign of a school nearby?" The instructor can choose to get voice-based prompts to avoid looking at the screen.
- **Calling Feature**—The interface provides the instructor an option to call the students. There is also an emergency calling feature, which connects the instructor with the driving school contact.

4.2 Web-Based Application

The managers interact with the system using the Web application. They have access to all the students' data, instructors' data, schedules, appointments, etc. Saving the documents online will mitigate the need to maintain physical records. They can process payments, track vehicles, send reminders to the students, track student progress, auto-schedule sessions, etc. The Web app will provide a one-stop solution to the manager for managing all the tasks in the school, thereby reducing the need for multiple applications for managing each aspect. The manager will also have the ability to track the real-time location of all the vehicles through this platform.

4.3 Mobile Application

The students will be accessing this system through the mobile application. The application will provide them with all the details regarding their learning sessions, license appointments, instructor contacts, etc. They will also get access to theory learning content like videos, articles, etc. They can contact the instructor and the driving school through the app. As their driving sessions are completed, the summary of those sessions is updated on the app. This way, the students can observe their own progress. The ubiquitous nature of mobile phones makes this an ideal interface for students since the students come from varying age groups, socio-economic backgrounds, etc.

Fig. 10 PURE method results

5 User Testing and Results

The mobile application was tested with 5 users, who were in the process of learning to drive, or had gone through the process in the near past. They were asked to use the designed application and were observed while performing the tasks. Observations were converted into a numeric scale and put in a graphical form where it could be easily visualized. User's verbal feedback was also noted at each step, thus creating a balance between qualitative and quantitative study. Figure 10 shows the results from the PURE method testing.

After the user testing, it was found that users were having difficulty in using the application when they were provided with more complicated visualizations like calendar or reports, where as they thought visibility of different features was good and made learnability of the application fairly easy. Some users pointed out concerns regarding the use of various gradients across the application, which would hinder the visibility of the content. This gives us an insight that crucial screens of the application where PURE analysis suggests users are finding difficulty needs to be reworked and further tested. Figure 11 shows the mobile application screens after the iterations based on the feedback.

6 Discussion

Despite the abundance of research in the domain of driving schools, there is a lack of research that deals with the data management methods in driving schools. This paper explored the different data management methods and problems faced by the driving schools and proposed a design intervention to minimize them. Through contextual inquiry and semi-structured interviews, it was revealed that the driving schools conform to conventional data management methods such as attendance cards,

Fig. 11 Mobile application screens

registers, excel sheets, and storing physical documents. This leads to several data management problems like losing data, difficulty in accessing records quickly, space consumption due to physical documents, miscommunication among the stakeholders, etc. The competitor's analysis reflected insights about existing solutions for these problems and the scope for an intervention.

The proposed design intervention consists of in-vehicle infotainment (IVI) system linked to a mobile and Web application. The intervention was built in consideration to the observation from the paper [13] that designers should aim to make the transition between e-governance systems easy for both internal actors and end-users, and recognize that new systems will co-exist with the existing ecosystem long past the planned transition period. As a result, the intervention builds on the existing data management methods of the driving schools, by transitioning them to a digital platform. The intervention is designed with regards to the needs of all the stakeholders - driving school managers, driving instructors, and students. The IVI system focuses on minimizing the interactions and cognitive load during the driving sessions. The Web application and mobile application facilitate seamless communication between the students and the driving school.

The future scope of this study would encompass testing the proposed intervention across varying school sizes in India based on their infrastructure, the number of users, financial capacity, etc., and iterating based on the feedback. The intervention can expand its interaction beyond the existing interfaces like IVI systems, mobile,

and Web applications, incorporating more advanced technologies like voice user interfaces, augmented reality, etc.

7 Limitations

The solution could be tested only with one set of stakeholders, i.e., students. In order to achieve a holistic perspective about the intervention, the solution needs be tested with all the stakeholders—driving school managers, instructors, and students. Due to the remote nature of work, the prototype was tested remotely. Testing the interface solutions in the real context, i.e., IVI system in a car, Web-based application in the driving school, etc., will provide deeper usability insights.

References

1. Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi. *Road Accidents in India—2019.* (New Delhi, 2020)
2. A.S. Neelima Chakrabarty, Driver training: an effective tool for improving road safety in India. J. Eng. Technol. **2**(2) (2012)
3. Management System, Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Management_system (n.d.)
4. C.F. Joost, S.M. de Winter, Advancing simulation-based driver training: lessons learned and future perspectives, in *MobileHCI '08: Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services* (2008), pp. 459–464
5. F.B. Pierro Hirsch, Transfer of skills learned on a driving simulator to on-road driving behavior. Transport. Res. Rec.: J. Transport. Res. Board **2660**, 1–6 (2017)
6. P.A. Gunhild Birgitte Sætren, Simulator training in driver education—potential gains and challenges, in *Safety and Reliability–Safe Societies in a Changing World*, (2018), pp. 2045–2051
7. I.M. Akshay Uttama Nambi, ALT: towards automating driver license testing using smartphones, in *SenSys '19: Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, (2019), pp. 29–42
8. V. Dmitry, K.E. Gribanov, Formation of driving skills in driving schools. **15**, 123–126 (2019)
9. K.H. Hugh Beyer, *Contextual Design: Designing Customer-Centered Systems* (Elsevier Science & Technology, 1997)
10. R.F. Brugha, Z. V., Stakeholder analysis: a review. Health Policy Plan. **15**(3), 239–246 (2000)
11. *Competitor Analysis.* (n.d.). Retrieved October 2021, from Usability Body of Knowledge: http://www.usabilitybok.org/competitor-analysis
12. C. Rohrer, *Quantifying and Comparing Ease of Use Without Breaking the Bank*. Retrieved from Nielsen Norman Group: https://www.nngroup.com/articles/pure-method/ (2017)
13. P.C. Megh Marathe, Officers never type: examining the persistence of paper in e-governance, in *CHI '20: CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–13
14. *Android Auto.* (n.d.). Retrieved from Google Design for Driving: https://developers.google.com/cars/design/android-auto

Revamping an E-Application for User Experience: A Case Study of eSanjeevaniOPD App

Remya Vivek Menon and G. Rejikumar

Abstract E-governance initiatives are likely to succeed only if the applications created for those purposes offer an excellent user experience (UX). For UX to improve, many components associated with the application should meet expectations and make user interactions easier, intuitive, and relaxing. This study aimed to verify the user experience developed by the 'eSanjeevaniOPD' app and suggest ways to improve it. The sequential incident technique (SIT) revealed user concerns in every stage of the user journey and, therefore, improvement opportunities. Accordingly, a few attributes were chosen to improve UX by providing them in the best possible manner. The Taguchi experiment with ten selected attributes by capturing user perceptions identified an optimum combination of these attributes for maximum UX.

Keywords eSanjeevaniOPD · User experience · Taguchi experiment · App design

1 Introduction

E-governance involves the application of information technology for efficient, effective, transparent, flexible, and stakeholder-friendly implementation of various public policy initiatives for positive outcomes [1, 2]. Technology contributes to achieving several goals related to E-governance such as enhancing the quality of service delivery to citizens, improving interactions with government and stakeholders, accelerating information access to citizens, and eradicating social inequalities [3]. For example, E-governance initiatives in health care are vital to upholding health equity by removing barriers to healthcare access and ensuring safe management of individual health records, society health data, and healthcare information systems [4].

R. V. Menon (✉)
Department of Management, Amrita Vishwa Vidyapeetham,
Amritapuri, Kollam, Kerala 690525, India
e-mail: remyavmenon@am.amrita.edu

G. Rejikumar
Department of Management, Amrita Vishwa Vidyapeetham, Kochi 682046, Kerala, India
e-mail: g_rejikumar@asb.kochi.amrita.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_38

UN Sustainable Development Goal 3 emphasizes the need for everybody to live a healthy life and highlights the well-being of everyone on the planet [5]. To ensure well-being as specified in SDG3, eliminating barriers to healthcare access [6] and achieving health equity [7, 8] are an essential prerequisite. The integration of information and communication technologies (ICT) into health care provides better governance of the healthcare systems in the country and helps to (1) improve the overall quality and efficiency; (2) reduce healthcare expenses; (3) improve healthcare access to a wider population; (4) minimize health disparities [9].

1.1 'eSanjeevaniOPD' for E-governance in Indian Health Care

To vitalize the digital health ecosystem and minimize healthcare access disparities and achieve the National Digital Health Mission, the government of India introduced the 'eSanjeevaniOPD', National Telemedicine Service, to citizens for outpatient health services, in November 2019. The 'eSanjeevaniOPD' is a doctor-to-patient telemedicine system to provide health advice and prescriptions through digital support without visiting the hospital. Telemedicine provides medical services remotely using telecommunications networks [10]. It helps to address the unequal distribution and shortages in the healthcare system by connecting rural patients to medical specialists and doctors in metropolitan areas. The 'eSanjeevaniOPD' platform has completed 120 lakh consultations as of September 2021 and now serves around 90,000 patients daily across the country, indicating broad acceptance from patients as well as doctors. The facility offered immense support during the COVID-19 pandemic when the public could not physically visit hospitals due to travel restrictions. Despite this, patients have criticisms regarding the quality of the user experience from the application. These include the shortage of doctors, especially specialists, accessibility challenges, language limitations, long waiting lines, inadequate two-way communication possibilities, prior booking options, to list a few [11].

1.2 Motivation Behind the Study

ISO defines user experience (UX) as 'the perceptions and reactions people have when using or anticipating to use a product, system, or service' [12]. UX includes both utilitarian and hedonic aspects attached to each stage of user interaction with the technology or application [13]. Optimizing UX involves building an attractive and efficient user interface (UI) that facilitates interaction with an application or website [14]. An effective UI minimizes the user's effort to achieve the maximum desired results from the services [15]. Literature suggests that approaches like 'user-centered design' [16], 'evidence-based design [17], etc. are for enhancing UX. In user-centered

design, the designers give extensive attention to user requirements in each service design stage. Similarly, in evidence-based design, designers use reliable, quantitative, and qualitative inputs from users for finalizing designs and revamping designs for better UX. Hence, exploring and understanding user experiences while interacting with web applications is useful for improving UX. Given the above observations, we attempt to qualitatively understand user experience with the 'eSanjeevaniOPD' using the sequential incident technique (SIT) in this study. Additionally, we attempt to identify certain design modifications and test their impact on UX using the Taguchi approach of robust design. The current study has many practical implications because much research about user perceptions and their experience with telemedicine services offered under 'eSanjeevaniOPD' is not available. We expect that these insights will help technologists identify significant issues in the web design and modify it for better UX and thus helpful in popularizing the E-governance initiative aimed at improving healthcare access in India.

2 Methods and Materials

The first stage in this research involves explorations to understand the general perceptions of users about the 'eSanjeevaniOPD' application. We have used a sequential incident technique (SIT) at this stage to achieve this objective.

2.1 Sequential Incident Technique (SIT)

An SIT uses an exploratory, qualitative interview of identified stakeholders to recollect and report critical incidents that caused pleasant or painful experiences during the service journey [18]. Then, using a visual representation of the customer path diagram, SIT identifies incidents affecting UX in service processes. Figure 1 presents the patient journey map prepared for eliciting patient responses with the 'eSanjeevaniOPD' app.

A modified version of SIT, sequence-oriented problem identification (SOPI), that captures only adverse incidents [19], is more useful when the objective is to identify areas in need of intervention. Therefore, the participants are motivated to recollect the experiences across different episodes in the customer service journey. The participants for SIT were ten patients who have used 'eSanjeevaniOPD' multiple times for various health consultations. The selection of participants was purely judgmental. Since this study aimed at identifying a few modifications contributing to better UX, we asked the participants to recall the negative incidents experienced in each stage of their service journey described in Fig. 1.

Fig. 1 Patient service journey in eSanjeevaniOPD app

2.2 Results of SIT

The SIT results indicated that, in general, participants seemed to prefer the application and were unanimously satisfied with its concept and how it helped them to get medical consultations during the COVID-19 period. Additionally, they pointed out that despite several glitches regarding connectivity, availability of doctors, and general service quality issues, the application still can improve healthcare access considerably. However, the participants are informed about many negative incidents that have affected their UX in each stage of the patient journey.

In the second phase, we have attempted to develop multiple ideas to improve UX based on inputs from the SIP, in consultation with a few experts. For many modification ideas related to service coverage, on-boarding procedure, prior appointment provisions, etc, experts viewed that available alternative options have merits and demerits. For example, an expanded array of services is possible if service coverage is national, though language barriers in consultations can affect effectiveness. Similarly, the option to choose a doctor can increase user freedom but can create waiting-line management issues. There were similar uncertainties about the ideal level concerning many important attributes contributing to UX, and hence, a multi-criteria decision-making experiment is ideal. We have used the Taguchi experiment to decide the relative importance of the above attributes and their most preferred levels that optimize UX. Table 1 provides the attributes and their chosen levels for the Taguchi experiment shortlisted based on expert consultations.

Table 1 The attributes and their levels considered for experimenting UX

Attributes	Options (levels)
Service coverage (SCO)	State, national
Prior appointment booking (PAB)	Yes, no
On-boarding (ONB)	Own password, only OTP, token-based
Option to choose doctor (OCD)	Yes, no
Login (LOG)	Specifically to a selected doctor, without doctor specificity
Doctor profile (PRO)	Name and qualification, name, age, qualification, and experience; name, age, qualification, experience, and contact details
Gender-based appointment (GEN)	Yes, no
Repeat consultation (REP)	Same doctor, different doctor, option to choose
Facility to rate doctor (RAT)	Yes, no
Prescription (PRE)	Only generic name, only brand name, both the brand and generic name

2.3 Taguchi Experiment

A Taguchi experiment [20] attempts to identify the best levels of design attributes that minimize quality variations due to noise factors (uncontrollable) associated with any product/service production. In 'eSanjeevaniOPD', noises emerge due to connectivity issues, heterogeneity in doctor counseling quality, patient characteristics, etc. Hence, to minimize the adverse effect on UX due to noises, the design features of the application should be user-centric. Traditionally, mean-based calculations without estimating variations evaluated the quality of performance of a product or service, and therefore, assessment of optimum performance suffers [21]. The measure named 'signal to noise ratio' (S/N) was obtained by conducting a Taguchi experiment that considers both mean and variance to evaluate the quality. Taguchi proposed three ways to calculate S/N ratios depending on the target performance objective [22]. They are 'Larger the better', 'Smaller the better', and 'Nominal the better'. We have used the 'Larger the better' criterion since the performance objective was to maximize UX as per participants' perceptions in the experiment.

The Taguchi experiment offers insights to improve the system design, parameter design, and tolerance design [23] and helps create a robust design capable of handling noises. The system's design informs the relative importance of each attribute in the design, whereas the parameter design helps identify the best working level of the attribute that makes the system robust, i.e., maximum resistance to noises. Lastly, the tolerance design offers insights into the extent of variations attached to attributes to decide tolerance limits. The Taguchi experiment provides the main effects of each attribute in the design on UX perceptions calculated from participant responses. The Taguchi experiment uses an orthogonal array (OA), a representative sample

from the total possible combinations of all attributes and their levels, for estimating metrics helpful to decide the optimum design [24]. The different stages in a Taguchi experiment are selecting attributes and their levels, the decision of orthogonal array, conducting the experiment and calculating response variable, evaluations using S/N ratios, and deciding optimal settings.

This experiment involved six factors in two levels and four factors in three levels. Hence, the number of combinations is $2^6 * 3^4 = 5184$. An orthogonal array (OA) is a subset of all these combinations [25] and is therefore used as a sample of design combinations for the experiment to choose the best combination. The 'Design of Experiments' (DOE) tab of the tool Minitab v17 includes the Taguchi experiment and has a provision for 'Create Taguchi Design' for generating the OA. Since this study used a multiple-level design of ten factors, of which six are in two levels, and four are with three levels, the tool suggested an L36 OA with 36 combinations (runs) for the experiment. The response variable for each run was calculated based on the importance ranking given by selected evaluators to attributes and levels ($n = 12$). The evaluators were users of 'eSanjeevaniOPD' application. In finalizing the evaluators, we conducted an initial round of discussion, explained the purpose of the experiment to them, and assessed their expertise in offering rankings/ratings. The methodology adopted for calculating the response variable was as follows.

First, the evaluators ranked all attributes (factors) from rank 1 to 10 based on their importance perceptions of these attributes in developing UX. Second, they ranked the levels attached to each attribute based on a preference from rank1 to rank 2 or 3 depending on the number of levels. The weightage scheme for factors and levels was; rank 1 factor = 10, rank 2 factor = 9, so on to rank 10 factor = 1; rank 1 level = 3; rank 2 level = 2, rank 3 level = 1 for three-level factors; and rank 1 level = 2 and rank 2 level = 1 for two-level factors. Thus, the weighted factor level score is factor rank weightage * level rank weightage. Additionally, the response score of the run for each evaluator will be the sum of the weighted factor level score of each attribute level attached to a run in the experiment. The response scores calculated above for each evaluator for the 36 runs formed the response matrix ($36 * 12$) used in the Taguchi experiment. In this study, we adopted the 'Larger the better' criterion for calculating the S/N ratio, since higher values of response score explain better UX. The Taguchi experiment using Minitab v17 yields the OA and S/N ratios.

The residual plots obtained as part of the output appeared as a straight line to imply that residuals are normally distributed, and hence, results are valid. Based on the Minitab output, '*R*-sq (adj)' reported was 52.6, confirming that the model sufficiently explains the variation in the response.

3 Results

The Taguchi experiment identified the relative importance of the factors and their most preferred level. The 'Delta' measure representing the difference between the highest and lowest average performance score (S/N, or Standard deviation) of a

Table 2 Response table for factor and their levels

Level	SCO	PAB	LOG	OCD	GEN	RAT	ONB	PRO	REP	PRE
1	36.24	37.33*	36.76	36.77	36.52	37.09*	36.99*	37	37.18	37.1*
2	37.51*	36.42	37*	36.99*	37.24*	36.66	36.92	36.57	35.87	36.92
3							36.72	37.06*	37.58*	36.61
Delta (S/N)	1.27	0.91	0.24	0.22	0.72	0.43	0.27	0.48	1.71	0.49
Rank	2	3	9	10	4	7	8	6	1	5
Delta (std. dev)	1.77	2.27	0.59	1.64	2.43	0.48	1.76	5.03	2.8	4.8
Rank (std. dev)	6	5	9	8	4	10	7	1	3	2

Fig. 2 Main effects plot for S/N ratio

factor indicates its importance in creating UX. The higher the delta values indicates the greater importance of the factor. The S/N ratio is more accurate than the mean for identifying the most important factor, and its preferred level is less affected by noises. Table 2 provides the details of the importance of factors based on the S/N ratio and standard deviation. Figure 2 illustrates the main effects plot of the S/N ratio to understand the effect of each factor on UX. The main effect exists when different levels of a factor affect UX differently.

Table 3 Contribution of factors and optimum levels for UX

Source	Seq SS	p-value	Contribution (%)	Best level for UX
Service coverage (SCO)	14.489	<0.05	23.88	National
Prior appointment (PAB)	7.506	<0.05	12.37	Yes
Login (LOG)	0.512	<0.1	0.84	Without doctor specificity (general)
Doctor selection (OCD)	0.423	<0.1	0.70	No
Gender-based (GEN)	4.630	<0.05	7.63	No
Doctor rating (RAT)	1.696	<0.05	2.80	Yes
On-boarding (ONB)	0.479	<0.1	0.79	Own password
Doctor profile (PRO)	1.686	<0.05	2.78	Name, age, qualification, experience, and contact details (detailed)
Repeat visit (REP)	19.221	<0.05	31.68	Option to choose
Prescription (PRE)	1.504	<0.05	2.48	Generic medicine
Residual error	8.521	–	14.05	
Total	60.667			

The ANOVA table in the Taguchi output helps to determine the statistical significance of the effect of factors on responses. The results found that all the factors are a significant association with UX at least to 0.1 level. Additionally, the error percentage contribution was only 14.05, less than the permissible limit of 15% [25], confirming role of all factors in creating the UX. The maximum contribution to UX was due to the selection of the doctor for repeat consultation (31.68%), followed by service coverage (23.88%) and prior appointment feature (12.37%). In this experiment, the optimal combination was not available in the sample of 36 runs. The optimal combination offering maximum UX is identified from the last column of Table 3. We have validated the experiment by predicting the S/N ratio of the optimal combination using the 'Predict Taguchi Results' option in Minitab. The results showed an S/N ratio of 42.16 much higher than any of the 36 runs to confirm the experiment's validity.

4 Conclusions

E-governance initiatives will be successful only if the applications created for such purposes develop enriching UX. In improving UX, many attributes associated with the application should perform according to user expectations and make user interactions more simple, intuitive, and relaxing. This study observed that the 'eSanjeevaniOPD' app developed for better governance of healthcare access in India has many drawbacks affecting UX. Every stage in the user journey contains areas of concern and thus the potential for improvement. We found that technological challenges, doctor quality, and service quality of doctors during the consultation are the major noises affecting UX. As part of this study, we have shortlisted a few attributes identified to contribute to UX and conducted a Taguchi experiment to understand their relative importance and choose the optimum way to incorporate them in the app design. Similarly, password-protected on-boarding to the app is most preferred for UX. We observed that users prefer to avail telemedicine services from doctors nationally without the limitations of the state of domicile. A prior appointment facility can significantly enhance UX. The users do not prefer logging into doctor-specific consultation rooms, patient gender-based doctor allotment, or facility to choose the doctor in the first instance. Users prefer to know a detailed profile of the doctor with contact details, the ability to offer ratings to doctors, and a facility to choose a doctor for repeat consultation. In addition, users wish to see generic names of the medicines in the prescriptions.

Based on the study's findings, we suggest that a five-dimensional approach is ideal for improving UX and improving E-governance in health care. These dimensions are (1) technology innovation, (2) service quality of medical professionals, (3) IT infrastructure, (4) ethics, and (5) policy support. Technological innovations focused on UX should provide a more participatory, creative, and real-time experience to users. Also, a favorable change in the attitude of medical professionals is essential for popularizing telemedicine. A serious avoidance from senior doctors is visible. There should be sufficient training programs and technology updates required for empowering doctors to be more patient-friendly during online consultations. IT infrastructure-related challenges adversely affect UX and penetration of telemedicine. More private participation may be invited for infrastructural augmentations. Concerns of privacy and security among users prevent them from adopting telemedicine. A clear legal framework offering trust to users about handling issues related to privacy breaches, security, etc. can significantly improve UX. Successful implementation of E-governance in India would require the government, medical institutions, corporate ICT players, and non-profit organizations to collaborate, remove major roadblocks, and remove major bottlenecks. Hence, a clear policy framework is an overarching requirement for the fast diffusion of telemedicine initiatives.

There are a few limitations for this study. First, the majority of the attributes considered were doctor-specific. Second, the calculation of the response variable was based on rankings and ratings offered by evaluators, and a more quantifiable outcome would have reduced the subjectivity element in the experiment. Third, even

though the SIT revealed that users prefer to get many post-consultation services such as online pharmacy, health plans, tie-up for health check-ups, hospital referrals, such aspects were not included in the experiment. Future studies involving more attributes and experiments focusing on each stage of the service journey can help in revamping the telemedicine services for better UX.

References

1. M. Jain, N. Abidi, A. Bandyopadhayay, Int. J. Technol. Manage. Sustain. Dev. E-procurement espousal and assessment framework: A case-based study of Indian automobile companies **17**(1), 87–109 (2018)
2. J.S. Ojo, J. Public Admin. Policy Res. E-governance: An imperative for sustainable grass root development in Nigeria **6**(4), 77–89 (2014)
3. Y.M. Asi, C. Williams, Equality through innovation: Promoting women in the workplace in low–and middle–income countries with health information technology. Journal of Social Issues **76**(3), 721–743 (2020)
4. I.C. Señor, J.L.F. Alemán, A. Toval, Computer. Personal health records: new means to safely handle health data? **45**(11), 27–33 (2012)
5. L.M. Fonseca, J.P. Domingues, A.M. Dima, Sustainability. Mapping the sustainable development goals relationships **12**(8), 3359 (2020)
6. J.E. Carrillo, V.A. Carrillo, H.R. Perez, D. Salas-Lopez, A. Natale-Pereira, A.T. Byron, J. Health Care Poor Underserv. Defining and targeting health care access barriers **22**(2), 562–575 (2011)
7. J. Sundewall, B.C. Forsberg, Lancet. Understanding health spending for SDG 3 **396**(10252), 650–651 (2020)
8. M. Kludacz-Alessandri, R. Walczak, L. Hawrysz, P. Korneta, J. Clin. Med. The quality of medical care in the conditions of the COVID-19 pandemic, with particular emphasis on the access to primary healthcare and the effectiveness of treatment in Poland **10**(16), 3502 (2021)
9. I. Gole, T. Sharma, S.B. Misra, J. Manage. Public Policy. Role of ICT in healthcare sector: an empirical study of Pune city **8**(2), 23–32 (2017)
10. S.K. Mishra, L. Kapoor, I.P. Singh, Telemed. e-Health. Telemedicine in India: current scenario and the future **15**(6), 568–575 (2009)
11. N. Bajpai, M. Wadhwa, *National Teleconsultation Service in India*. eSanjeevani OPD (2021)
12. T. Lindgren, M. Bergquist, S. Pink, M. Berg, V. Fors, Experiencing expectations: extending the concept of UX anticipation, in *Scandinavian Conference on Information Systems* (Springer, 2018), pp. 1–13
13. J. Häkkilä Designing for smart clothes and wearables user experience design perspective, in *Smart Textiles* (Springer 2017), pp. 259–278
14. A. Pitale, A. Bhumgara, A.: Human computer interaction strategies-designing the user interface, in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (IEEE, 2019), pp. 752–758
15. R. Oppermann, User-interface design, in *Handbook on Information Technologies for Education and Training* (Springer, Heidelberg, 2002), pp. 233–248
16. R.W. Veryzer, B. Borja de Mozota, The impact of user-oriented design on new product development: an examination of fundamental relationships. J. Product Innov. Manage. **22**(2), 128–143 (2005)
17. D.K. Hamilton, Healthcare Des. The four levels of evidence-based practice **3**(4), 18–26 (2003)
18. B. Stauss, B. Weinlich, Eur. J. Market. Process-oriented measurement of service quality: applying the sequential incident technique **31**(5), 33–55 (1997)

19. G. Botschen, L. Bstieler, A.G. Woodside, J. Euromark. Sequence-oriented problem identification within service encounters **5**(2), 19–52 (1996)
20. G. Taguchi, *System of Experimental Design*, vols. 1 and 2. UNIPUB/Krauss International, White Plains, New York (1987)
21. G. Rejikumar, A.A. Ajitha, M.S. Nair, TQM J. Healthcare service quality: a methodology for servicescape redesign using Taguchi approach **31**(4), 600–619 (2019)
22. G. Taguchi, S. Chowdhury, Y. Wu, S. Taguchi, H. Yano, *Taguchi's Quality Engineering Handbook* (Wiley, 2005)
23. S.K. Karna, R. Sahai, Int. J. Eng. Math. Sci. An overview on Taguchi method **1**(1), 1–7 (2012)
24. G. Taguchi, S. Konishi, *Orthogonal Arrays and Linear Graphs: Tools for Quality Engineering* (ASI Press, Michigan, 1987)
25. M. Tanco, E. Viles, L. Ilzarbe, M.J. Alvarez, Appl. Stochast. Models Bus. Ind. Implementation of design of experiments projects in industry **25**(4), 478–505 (2009)

Secured Data Transmission in Low Power WSN with LoRA

Arabinda Rath, S. Q. Baig, Bisakha Biswal, and Gayatri Devi

Abstract Power consumption factor is very important in WSN and other IOT setup. As the nodes remain in distant locations they need to run powered by battery. When we try to send private data through it we can encrypt it. But a single key for a long period may be less secure. Hence, we need to change the key. But in a low power low processing power setup the conventional key generation algorithm may not be quite feasible as it may need a lot of processing and power. Therefore, we study low power alternative for dynamic key generation and implement it in a LoRa-based WSN.

Keywords Encryption · Decryption · AES · Steganography · Dynamic key · Algorithm · WSN · LoRa · Arduino

1 Introduction

Wireless sensor network (WSN) is a group of specialized electronic devices, which is capable of sensing, computation and wireless communication for monitoring and recording the physical conditions of the environment. The collected informations are organized at the central location. Cryptography is important because it provides for secure communication techniques that allow only the sender and the intended recipient of a message to view its contents even if in the presence of malicious third parties. Businesses, government and many individual use it to protect their personal information.

In this paper while using IOT and wireless sensor network the energy and processing constraint has to be considered. Keeping that into mind we have created a novel algorithm that can generate new keys for AES encryption system making it more secure.

A. Rath (✉) · S. Q. Baig · B. Biswal · G. Devi
Department of Computer Science and Engineering, ABIT, Cuttack, Odisha, India
e-mail: arabindarath73@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_39

2 Advanced Encryption Standard (AES)

The symmetric AES encryption algorithm is widely used because AES is much faster than triple DES [1, 2]. The features of AES are as follows

- Symmetric key symmetric block cipher
- 128-bit data, 128/192/256-bit keys
- Provide full specification and design details
- Stronger and faster than Triple DES
- Software implementable in C, C#, Java, etc.

AES is an iterative process based on 'substitution–permutation network'. It consists of a sequence of linked operations, where some involve replacing inputs by specific outputs using substitution and others involve shuffling bits around by using permutations.

All the computations of AES is based on bytes instead of bits. So, AES considers the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in 4×4 matrix.

2.1 AES Security Analysis

In AES, the length of the plain text of size 128-bit, 192-bit and 256-bit supporting three different key lengths of 128-bit, 192-bit and 256-bit, respectively, are used in this paper. The size of the input key must be same as size of the plain text.

The length of the key varies which creates the difference. The strongest level of encryption is provided by the 256-bit key which is the longest. This would ensure that a hacker would have to try 2256 different combinations for 256-bit key to make sure the right one is included. So we require 128- or 256-bit encryption for sensitive data.

AES consists of multiple rounds for processing different key bits such as 10 round to process 128-bit keys, 12 rounds to process 192-bit keys and 14 rounds to process 256-bit keys. This shows that 256-bit key is the most secure form of AES implementation which concludes that the more the number of the round is, the more complex is the encryption. Large key size and more rounds lead to more consumption of energy and space.

3 Key Management

Key management means to manage the cryptographic keys in a cryptosystem. It deals with the generation, exchange, storage, use, crypto-shredding (destruction)

and replacement of keys. It includes cryptographic protocol design, key servers, user procedures and other relevant protocols.

3.1 Key Management Cost

There are various algorithms to establish a key between two nodes. But they can be very costly in terms of processing and energy cost. For low power IOT devices [3] it can be a problem. Hence, we have developed a key establishment technique which relies on having a secret formula in both the devices and passing some random characters and generate the key using the formula. This can be an alternative to the conventional methods for low energy devices that work on battery.

4 Method

We transmit sensitive data in a secured way through Low Power LoRa [4, 5]. For this we use AES algorithm, we alter the keys whenever we need just by pressing a push button. We have a Sender Node and a Receiver Node. The Sender node Encrypts sensitive information and the receiver node decrypts it. There is a default key in the program which starts working as soon as we power up the device. The receiver node has a push button connected to it. Which if pressed initiated the key changing process. We program both nodes accordingly. We have used Arduino microcontroller for this.

4.1 Sender Node

We need to initialize software serial so that we can use it for transmission and receiving process while the default serial can be used for displaying the information. We also define baud rate.

Then we include AES Library so that we can perform AES encryption and decryption process. Then we initialize the AES key with a default value. It is a byte array of size 16. Each byte has 8 bits hence 2^8 different numbers can be stored in it. Which are 0 to 255. We can also store hexadecimal values as shown in the program.

Then we define a function that prints the keys in the default serial, so we can view it in serial monitor. This can be useful when we want to view the new key after we have altered it. The Arduino program starts from void setup() we begin the default serial and software serial. And wait until serial is available. Then we display booting to mark the beginning of the program. Then we define and initialize various constants which we will use in our program. We free the stack for avoiding the error. To demonstrate transmission of secured data Every 5 s we send a generate a string and we encrypt and transmit it to the receiver node.

In Arduino void loop() keeps executing continuously. We need to check if there is any key change request continuously hence we need to put that within void loop(). We read the serial character by character until we come across new line. In that case we conclude that the message is complete. The string is sent by the receiver node which is key change request. We use our custom formula to generate the new key using the string received. Our custom formula performs some different arithmetic operations for each byte and then adds some random value of our choice. In the end we find the remainder of it after division with 256 because it can have value between 0 and 255.

4.2 Receiver Node

The receiver node receives the sensitive information which is encrypted and it decrypts the message and finally gets the message in plaintext which is the original message or the sensitive information.

Like the Sender node we initialize software serial so that we can use it for transmission and receiving process while the default serial can be used for displaying the information. We also define baud rate. Then we include AES Library so that we can perform AES encryption and decryption process. Then we initialize the AES key with a default value which should be the same value as sender node. Then we define a function that prints the keys in the default serial, so we can view it in serial monitor.

After this we define the decrypt function it takes the encrypted message and key as argument and returns the plain text original message.

Similar to sender node we initialize serial free the stack and declare and initialize some variables, although we have different variables in receiver node because the operation is different.

In the loop section of the receiver node we check if the push button is pressed or not. We delay for a little time while the button is pressed so that a single press does trigger the program to run several times. When the button is pressed we generate random numbers and save it in a string we send that string to the receiver node as a key changing request. Then we calculate the key value using the same formulae used in sender node and hence getting the same result. Finally, it decrypts the encrypted message received. If the button has never been pressed the decryption takes place with the default key.

Fig. 1 Sender sends first secured data

Fig. 2 Receiver receives the data

5 Simulation Result

5.1 Testing the Implementation

When we press the push button the key changes and encryption and decryption take place with the new key (Figs. 1, 2, 3, 4, 5 and 6).

6 Conclusion

While using IOT and wireless sensor network the energy and processing constraint has to be considered. Keeping that into mind we have created a novel algorithm that can generate new keys for AES encryption system making it more secure.

```
COM3
Send
03:48:13.387 ->
03:48:13.387 -> Booting...
03:48:21.432 -> 1 : Secured Data : 1
03:48:21.432 -> AES IV:
03:48:21.465 -> 16,48,144,16,48,19,48,48,23,48,48,48,48,48,48,48
03:48:21.500 -> ENCRYPTION (char*)
03:48:21.535 -> Encryption took: 17672us
03:48:21.569 -> Encrypted Result: FwcQoFzbzDo51EvgqVER9zk0ELNATY5caOikrc82SY4=
03:48:21.637 ->
03:48:21.637 ->
03:48:21.637 ->
03:48:21.637 ->
03:48:31.497 -> 2 : Secured Data : 2
03:48:31.497 -> AES IV:
03:48:31.532 -> 16,48,144,16,48,19,48,48,23,48,48,48,48,48,48,48
03:48:31.565 -> ENCRYPTION (char*)
03:48:31.599 -> Encryption took: 17674us
03:48:31.632 -> Encrypted Result: jRYLsopth1T7jsmXbYN9nHw6OTvrVbxLgZox4KrkdZg=
03:48:31.666 ->
```

Fig. 3 Second data

```
COM3
Send
03:48:21.581 -> Received: FwcQoFzbzDo51EvgqVER9zk0ELNATY5caOikrc82SY4=
03:48:21.649 -> DECRYPTION (char*)
03:48:21.649 -> FwcQoFzbzDo51EvgqVER9zk0ELNATY5caOikrc82SY4=
03:48:21.722 ->
03:48:21.722 -> Decryption took: 17672us
03:48:21.755 -> AES IV:
03:48:21.755 -> 16,48,144,16,48,19,48,48,23,48,48,48,48,48,48,48
03:48:21.789 -> Decrypted Result: 1 : Secured Data : 1
03:48:31.652 -> Received: jRYLsopth1T7jsmXbYN9nHw6OTvrVbxLgZox4KrkdZg=
03:48:31.686 -> DECRYPTION (char*)
03:48:31.719 -> jRYLsopth1T7jsmXbYN9nHw6OTvrVbxLgZox4KrkdZg=
03:48:31.753 ->
03:48:31.753 -> Decryption took: 17672us
03:48:31.786 -> AES IV:
03:48:31.786 -> 16,48,144,16,48,19,48,48,23,48,48,48,48,48,48,48
03:48:31.856 -> Decrypted Result: 2 : Secured Data : 2
```

Fig. 4 Receiver decrypts the data

Fig. 5 Changing the key sender's side

Fig. 6 Changing the key receiver's side

References

1. H.S. Deshpande, K.J. Karande, A.O. Mulani,. Efficient implementation of AES algorithm on FPGA, in *2014 IEEE International Conference Communications and Signal Processing (ICCSP)* (2014)
2. F.J. D'souza, Advanced encryption standard (AES) security enhancement using hybrid approach, in *2017 International Conference on Computing Communication and Automation (ICCCA)*, pp. 5–6 (2017)
3. A. Ishfaq, A. Muhammad, Applying security patterns for authorization of users in IoT based applications, in *2018 International Conference on Engineering and Emerging Technologies (ICEET)*, pp. 22–23 (2018)
4. J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, M. Pettissalo, On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology, in *14th International Conference on ITS Telecommunications (ITST)*, pp. 55–59 (2015)
5. M.O.A. Kalaa, W. Balid, N. Bitar, H.H. Refai, Evaluating bluetooth low energy in realistic wireless environments, pp. 1–6 (2016)

Safeguarding Cloud Services Sustainability by Dynamic Virtual Machine Migration with Re-allocation Oriented Algorithmic Approach

Saumitra Vatsal and Shalini Agarwal

Abstract Data centres are networking platforms which exhibit virtual machine workload execution in a dynamic manner. As the users' requests are of enormous magnitude, it manifests as overloaded physical machines resulting in quality of service degradation and SLA violations. This challenge can be negotiated by exercising a better virtual machine allocation by dint of re-allocating a subset of active virtual machines at a suitable destined server by virtual machine migration. It is exhibited as improved resource utilization with enhanced energy efficiency along with addressing the challenge of impending server overloading resulting in downgraded services. The aforesaid twin factors of enhanced energy consumption and enhanced resource utilization can be suitably addressed by combining them together as a single objective function by utilizing cost function based best-fit decreasing heuristic. It enhances the potentials for aggressively migrating large capacity applications like image processing, speech recognition, and decision support systems. It facilitates a seamless and transparent live virtual machine migration from one physical server to another along with taking care of cloud environment resources. The identification of most appropriate migration target host is executed by applying modified version of best-fit decreasing algorithm with respect to virtual machine dynamic migration scheduling model. By executing the selection algorithm, the hotspot hosts in cloud platform are segregated. Subsequently, virtual machine-related resource loads are identified in descending order with respect to hotspots. The resource loads pertaining to non-hotspot hosts are identified in ascending order. Next, the traversing manoeuvring in non-hotspot hosts queue is exercised for identification of the most appropriate host to be reckoned as migration target host.

Keywords Data centre · Load balance · Virtual machine migration · Cloud computing

S. Vatsal (✉) · S. Agarwal
Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India
e-mail: s.vatsall@gmail.com

S. Agarwal
e-mail: shalini.cs@srmu.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_40

1 Introduction

The robustness of the data centre is indispensable as it has to address business processing peak demands at certain times which require high-resource availability. But it culminates into waste of resources in off-peak periods. This challenging problem can be alleviated by scalable and dynamic resource allocation [1]. It is achieved by dynamically expanding or contracting server hosts' related allocation of resources according to fluctuating scenario of users' demand load. It serves to conserve effectively waste of resources along with addressing server host-related problem of being infested with highly loaded or too lightly loaded situation. Resource allocation is adjudged by infrastructure-related physical resource allocation along with virtual resource allocation at task or application level [2, 3]. The resource allocation with respect to first start-up time of the system is addressed by virtual machine placement by selecting and tapping the physical resources of most appropriate physical machine. Subsequently, it is addressed by exercising re-allocation of virtual machine from one to another physical machine by using migration techniques as and when required. Live migration permits virtual machine migration within cloud with a transparency in its running state. It addresses server virtualization with an added perk of judicious load balancing along with avoidance of overloaded "hotspot" physical machines, in running state, thus avoiding system-related down-time because virtual machines are not turned off during maintenance. The virtual machines on lightly loaded hosts can be clubbed together for them to be placed on fewer physical machines, thus securing avoidance of hotspot physical machines and resource requirement considerations. As a consequence freed up physical machines can be safely turned off for securing power saving. It results into new virtual machines-related high-resource availability [1, 4]. In addition, a new strategy can be added further by clubbing migration-worthy virtual machines into groups on the basis of their physical machines and magnitude of memory shared between them, leading to live gang migration [5]. The traditional static resource scheduling algorithms like destination hashing scheduling, round-robin scheduling, source-hashing scheduling, and weighted round-robin scheduling do not consider user requests' related system load changes or dynamic scheduling of resources and hence results into unnecessary virtual machine migrations [6–9]. It can be better addressed by utilizing energy-aware virtual machine migration algorithms which are capable of virtual machine migration to suitable target hosts for de-activation of idle physical hosts resulting in minimization of energy consumption. Since virtual machine dynamic migration scheduling represents as NP-hard combinatorial optimization bin-packing problem [10, 11], it is negotiated by optimization methods which enhance the efficacy of bin-packing algorithms and associated heuristics and meta-heuristics.

The paper is organized is as follows:

- Section 2 incorporates related work.
- Sections 3 and 4 are pertinent to the problem statement and system model, respectively.

- Section 5 describes a relevant algorithmic approach for dynamic virtual machine migration.
- Section 6 puts forward the performance analysis and the results.
- Section 7 concludes the paper.

2 Related Work

Dynamic virtual machine allocation challenge basically originates from virtual machine placement strategy. It is suitably addressed by static approaches represented by round-robin scheduling and destination hashing scheduling along with dynamic approaches incorporating bin-packing approach through first-fit (FF) algorithm, best-fit (BF) algorithm, first-fit-decreasing (FFD) algorithm, and best-fit decreasing (BFD) algorithm [12]. By using live migration-related clustering technique live gang migration, the optimization of memory and network bandwidth usage can be secured along with minimization of transfer of identical shared data between co-located virtual machines [5]. By addressing a clustering co-location mechanism through the minimization of multiple co-located virtual machine migrations, the system-related load balanced state can be preserved along with continuing the migration process by virtue of utilizing system standard deviation [13]. Virtual machine re-allocation problem subsequent to migration can be reckoned as integer programming quadratic constraint for minimization of data centre-related aggregation communication costs [14]. The servers' related cooling cost within a data centre can be minimized by server placement heuristic approach [15]. The minimization of power consumption of a data centre can be secured by resource aware virtual machine placement algorithm which addresses the resource usage factor [16]. The challenge in virtual machine placement with respect to available bandwidth can be suitably addressed by the use of whale optimization algorithm [17]. To address the energy efficiency issue, the approach has to be profile-based dynamic virtual machine placement which secures minimization of total energy consumption of all physical machines during dynamic virtual machine placement [18]. In order to maximize resource utilization along with minimization of energy consumption, the multi-objective functions are coalesced as one objective function in order to secure an optimal point existing between resource utilization and energy consumption by taking Euclidean distance into consideration [19]. On the basis of multi-objective genetic algorithm and Bernoulli simulation, the minimization of resource wastage pertaining to each individual host along with number of active hosts can be obtained simultaneously [20]. For securing best virtual machine migration-related re-allocation manoeuvring, a multi-objective weighted sum function as mono-objective addresses energy cost reduction along with minimized energy cost and SLA penalty cost [21]. A cross entropy algorithm addressing virtual machine migration-related re-allocation can address the issue of data centre-related total thermal cost minimization by dint of reduced energy consumption [22]. By addressing twin factors of energy consumption and migration cost, modified particle swarm optimization can facilitate re-allocation of migrated virtual machines

on the basis of fitness function [23]. By working on relaxed convex optimization framework, a multi-level join of migration and re-allocation algorithms can facilitate optimal solution for addressing migration costs, load distribution, and cross traffic-costs by clubbing them into mono-objective weighted sum function [24].

3 Explication of Problem Statement

Cloud data centre-related prime functionality ensures creation, migration, and cancellation of host-related virtual machines. Further, it substantiates deployment of virtual machine on to a suitable host by re-allocation in order to address the processing of users application-related service requests. It involves certain scheduling strategies with respect to virtual machine dynamic migration for executing load monitoring programme by virtue of retrieving load data of CPU, bandwidth, memory along with other resources related with each virtual machine and physical host. The cloud data centre retrieves resource usage information by executing load monitoring programme. The virtual machine migration scheduling is executed in accordance with platforms load balance strategy whenever host-related load is too heavy or too light. The system processing performance requirements are safeguarded by dynamically migrating virtual machines from one host to another suitable host in order to give the quality of service fully compatible with users' demands. The process of virtual machine migration scheduling is confronted with challenging factors which include right assessment of time for determining most appropriate time instant for virtual machine migration and assessment of most suitable inherent migration scheduling strategy. It also demands a judicious judgement to ascertain instant of time when to commence the migration operation and to confirm if whether the server-related platform is competent enough to sustain load balanced state.

4 System Model

The virtual machine dynamic migration scheduling model as in Fig. 1 encompasses the ability to process users-related service requests by the scheduling main controller which evokes the execution of host-related task. Each host is provisioned with its own load monitoring programme. The associated load monitoring program addresses load state of host and virtual machines with respect to real time. Thus, it provides load data for identifying in order to select hotspot hosts which are listed into a queue. The selected virtual machines undergo migration from this hotspot hosts queue to selected target hosts in order to achieve a load balanced state in data centre platform. This whole process thus involves hotspot host selection algorithm programme along with virtual machine scheduling strategy for ensuring dynamic migration.

The core strategy of cloud computing dynamic resources scheduling inherits the concept of migrating virtual machines amongst data centres' hosts in such a way that

Fig. 1 Dynamic migration scheduling mechanism of virtual machines

each host remains enabled to run at optimal load state. The SLA violation results if host is overloaded resulting in degradation of quality of service (QoS) due to degraded task execution. On contrary, if host is loaded too lightly, it culminates into underutilized host. It exhibits as a useless energy waste which is consumed by an underutilized host. This challenge is addressed by load monitoring programme which monitors resource utilization with respect to each and every host. It identifies the hotspot host on the basis of identifying a host if it exceeds the maximum threshold or the minimum threshold and designating it as a hotspot host.

Specific criteria are expressed as follows in equation set 1 where $A_{\mathrm{CPU}}(b_j)$, $A_{\mathrm{RAM}}(b_j)$, and $A_{\mathrm{Band}}(b_j)$ represent CPU-related load state, memory, and network bandwidth in relation to physical host b_j whilst predetermined maximum load threshold with respect to CPU, memory, and network is represented as $\mathrm{Max}_{\mathrm{CPU}}$, $\mathrm{Max}_{\mathrm{RAM}}$, and $\mathrm{Max}_{\mathrm{Band}}$. Likewise predetermined load threshold related with CPU, memory, and network bandwidth is represented as $\mathrm{Min}_{\mathrm{CPU}}$, $\mathrm{Min}_{\mathrm{RAM}}$, and $\mathrm{Min}_{\mathrm{Band}}$, respectively.

$$\left.\begin{array}{l} A_{\mathrm{CPU}}(b_j) > \mathrm{Max}_{\mathrm{CPU}} \; or \; A_{\mathrm{CPU}}(b_j) < \mathrm{Min}_{\mathrm{CPU}} \\ A_{\mathrm{RAM}}(b_j) > \mathrm{Max}_{\mathrm{RAM}} \text{ or } A_{\mathrm{RAM}}(b_j) < \mathrm{Min}_{\mathrm{RAM}} \\ A_{\mathrm{Band}}(b_j) > \mathrm{Max}_{\mathrm{Band}} \text{ or } A_{\mathrm{Band}}(b_j) < \mathrm{Min}_{\mathrm{Band}} \end{array}\right\} \tag{1}$$

The fulfilment of any aforesaid condition confirms the host to be a hotspot host.

5 Proposed Algorithmic Approach

By applying hotspot host selection algorithm virtual machine re-allocation algorithm (VMRA), the identification of hotspot hosts is achieved for their arrangement in a queue. But the real challenge is to achieve load balancing amongst hosts whilst using virtual machine dynamic migration scheduling strategy. The algorithm identifies hotspot host as an item and the target host as a packing bin. Subsequently, hotspot hosts are sorted in descending order according to size of items along with due consideration of load state of CPU, network bandwidth, and memory of virtual machines in each hotspot host which forms the basis of load state assessment of all virtual machines by cumulative analysis. The sorting of hotspot hosts in descending order is exercised according to its resource load state. By traversing non-hotspot host queue, the most appropriate target for migration is identified as a packing bin for migratory virtual machines. It infers that the difference between present host-related load state and the maximum threshold value should be minimal subsequent to re-allocating virtual machine on to the target host.

A hotspot host queue is drawn in accordance with following equation subsequent to combining host load state data and hotspot host selection algorithm.

$$\left.\begin{array}{l} HS_b = \{b_1, b_2, b_3 \ldots, b_m\};\ m \in N^* \\ A_{b_j} \geq \text{Max or } A_{b_j} \leq \text{Min} \forall j \in [1, m] \end{array}\right\} \tag{2}$$

In Eq. 2, the physical host is represented as HS_b.

Subsequent to ascertaining the hotspot host queue, the sorting in descending order of virtual machines in hotspot hosts is exercised taking their resource load state in consideration. The virtual machine resource load status is calculated according to following equation.

$$C_{\text{VM}} = x_1 c_{\text{CPU}} + x_2 c_{\text{RAM}} + x_3 c_{\text{Band}} \ \& \ \sum_{d=1}^{3} x_d = 1 \tag{3}$$

In Eq. 3, the virtual machine-related CPU, memory, and network bandwidth are represented as c_{CPU}, c_{RAM}, and c_{Band}, respectively. Different weights are represented by x_1, $x_{2,}$ and x_3.

The sorting of virtual machines in hotspot hosts zone in descending order is secured with respect to resource load state by following equation.

$$\left.\begin{array}{l} Y_{VM} = \{vm_1, vm_2, \ldots vm_r\};\quad VM \in HS_b, r \in N^* \\ C_l \leq C_k \quad \forall l, k \in [1, r] \end{array}\right\} \tag{4}$$

The virtual machine queue is represented as Y_{VM}. The virtual machine resource load is represented by C_l.

The sorting of all the non-hotspot hosts in ascending order is exercised considering their resource load state according to following:

$$\left.\begin{array}{l} W_b = \{b_1, b_2, b_3, \ldots, b_t\}; \quad t \in N^* \\ A_{b_j} \geq A_{b_l} \quad \forall j, l \in [1, t] \\ \text{Min} < A_{b_j} < \text{Max} \end{array}\right\} \tag{5}$$

The target host sorted in the queue is represented by $W_{b.}$

The selection of most appropriate target whilst migrating virtual machines is procured by traversing manoeuvring of target host queue for identifying the most suitable destination. It is achieved by applying equation:

$$\left.\begin{array}{l} FW_b = b_j \\ \text{Min}\{\text{Max} - (A_{b_j} + c_l)\} \\ C_l \in Y_{VM}, b_j \in W_b \end{array}\right\} \tag{6}$$

The load balancing degree is calculated for measuring the load balance magnitude of old data centre subsequent to exercising virtual machine dynamic migration scheduling. It is expressed by following equation:

$$(\text{Balance})^2 = \frac{1}{g}\left[\left(A_1 - \overline{A}\right)^2 + \left(A_2 - \overline{A}\right)^2 + \cdots + \left(A_g - \overline{A}\right)^2\right] \tag{7}$$

The average value of load resource state is represented by $\overline{A}$.

The virtual machine dynamic migration scheduling algorithm is as under:

Algorithm: VMRA

Input: Vir_Mchn_List, Host_List

Output: Dest_Host

```
Vir_Mchn_List.sortDecreasingCPUUtilization();
   for Vir_Mchn in Vir_Mchn_List do
      maximum_memb←0;
      Dest_Host←NULL;
      for Host in Host_List do
         if Host has sufficient resource for Vir_Mchn then
            memb←estm_Power_Utilization(Host, Vir_Mchn);
            if memb≥maximum_memb then
               maximum_memb←memb;
               Dest_Host←Host;
            if Dest_Host≠NULL then
               assign Vir_Mchn to Dest_Host;
return Dest_Host;
```

The following steps are involved in virtual machine dynamic migration scheduling process:

1. Identification of hotspot hosts: It is load monitor programme which determines each physical host-related resource load state in order to ascertain the virtual machine to accept the migration on to a host.

2. Selection of target host: Sorting of non-hotspot hosts in ascending order to determine the most suitable host fulfilling all pre-conditions necessary for a migration.
3. Virtual machine dynamic migration: The network connection information, memory, and load state of CPU of the virtual machine identified for migration on to a target host are taken in to consideration.
4. The success of migration is adjudged by confirmation message from hotspot hosts to the target confirming that migration process is finished. This message infers that target is enabled with a replica image of virtual machine which was destined to undergo migration from hotspot host. This concludes a successful migration.

6 Evaluation of Performance

6.1 Simulation Setup

The evaluation of virtual machine management and resource provisioning is based on CloudSim simulator [25, 26]. This simulation is exercised considering five physical hosts with identical configuration providing 50 identical virtual machines simulating 50 tasks in cloud environment. The load status of each host is monitored by assuming that one cycle is made up of ten times of experimental execution. The host load-related maximum threshold is considered akin to 80% of system-related resource load whilst the minimum threshold is considered to be 20% of system resource load. The dynamic migration of virtual machines is executed in accordance with hosts-related current resource load state.

6.2 Results

According to Table 1, it is evident that VMRA exhibits SLA violation percentage to be lesser as compared to its competitors along with minimization of virtual

Table 1 Different approach-related energy consumption

Approaches	Energy consumption (in kWh)	SLA violation (in %)	No. of migrations
VMRA	1.39	9.97	297
CA1	1.69	11.29	329
CA2	1.60	10.73	323
NPA	4.57	–	–
DVFS	2.39	–	–

Fig. 2 Different approaches-related comparative illustration of energy consumption

machine migrations. It is evident that strategy model for VMRA incurs enhanced energy savings with respect to NPA, DVFS, CA1, and CA2 in terms of less energy consumption, thus exhibiting its superiority over its competitors.

The comparative experimental analysis as shown in Fig. 2 is conducted by incorporating dynamic voltage and frequency scaling (DVFS) [27], non-power aware (NPA) [28, 29], and simple threshold (ST)-based competitive algorithm 1 (CA1) and competitive algorithm 2 (CA2) for depiction of benefits pertaining to minimized energy consumption, reduced number of migrations, and minimized service level agreement (SLA) violations.

The comparison of load balance degree as shown in Fig. 3 with respect to virtual machine dynamic scheduling clearly indicates that load balance degree of VMRA achieves superior load balancing with due consideration of system overhead parameters which include migration number, energy consumption, and SLA violation percentage.

7 Conclusion

For the sake of conserving the data centre-related networking viability, the functioning of virtual machines in cloud environment has to remain uninterrupted by virtue of countering all challenging threat factors which may jeopardize its functioning. The dynamic virtual machine migration and re-allocation on the basis of identifying its load status of target host along with the resource reservoir are the factors of great importance. This approach revamps the concept of green cloud computing. A

Fig. 3 Comparative performance analysis of load balance degree

minimized number of virtual machine migrations are ensured by turning-off the idle hosts subsequent to a judicious dynamic virtual machine migration and re-allocation on a suitable target host, thus conserving the energy consumption. The assigning of a virtual machine to a target host is ensured in accordance with present and future resource requirements. The algorithmic approach addressed here exhibits its effectiveness encompassing minimized number of migrations, energy consumption along with less SLA violations.

Acknowledgements I am thankful to all concerned with this endeavour.

References

1. M. Mishra, A. Das, Dynamic resource management using virtual machine migrations. IEEE Commun. Mag. **50**(9), 34–40 (2012)
2. W. Hashem, H. Nashaat, R. Rizk, Honey bee based load balancing in cloud computing. KSII Trans. Internet Inf. Syst. (TIIS) **11**(12), 5694–5711 (2017)
3. M. Gamal, R. Rizk, H. Mahdi, Bio-inspired load balancing algorithm in cloud computing, in *Proceedings of International Conference on Advanced Intelligent Systems and Informatics (AISI)*, Cairo, Egypt, pp. 579–589 (2017)
4. A. Strunk, Costs of virtual machine live migration: a survey, in *Proceedings of 8th IEEE World Congress on Services (SERVICES)*, Honolulu, HI, USA, pp. 323–329 (2012)
5. U. Deshp, X. Wang, K. Gopalan, Live gang migration of virtual machines, in *Proceedings of 20th International Symposium on High Performance Distributed Computing*, San Joes, CA, USA, pp. 135–146 (2011)
6. R.V. Rasmussen, M.A. Trick, Round Robin scheduling—a survey. Eur. J. Oper. Res. **188**(3), 617–636 (2008)
7. Weighted round robin, Available online at: http://en.wikipedia.org/wiki/Weighted_round_robin

8. B.X. Chen, X.F. Fu, X.Y. Zhang, L. Su, D. Wu, Design and implementation of intranet security audit system based on load balancing, in *Proceedings of IEEE International Conference on Granular Computing*, pp. 588–591 (2007)
9. K.S.J. Hielscher, R. German, A low-cost infrastructure for high precision high volume performance measurements of web clusters, in *Computer Performance Evaluation. Modelling, Techniques and Tools*. Lecture Notes in Computer Science, vol. 2794, pp. 11–28 (2003)
10. C. Chekuri, S. Khanna, On multi-dimensional packing problems, in *Proceedings of 10th Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics*, pp. 185–194 (1999)
11. H. Youssef, S.M. Sait, Iterative computer algorithms with applications in engineering, Chapter 2
12. A. Beloglazov, R. Buyya, Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. J. Concurr. Comput. Pract. Exp. **24**(13), 1397–1420 (2012)
13. N. Ashry, H. Nashaat, R. Rizk, AMS: adaptive migration scheme in cloud computing, in *Proceedings of 3rd International Conference on Intelligent Systems and Informatics (AISI2018)*, Cairo, Egypt, vol. 845 (Springer, 2018), pp. 357–369
14. D. Zeng, S. Guo, H. Huang, S. Yu, V.C. Leung, Optimal VM placement in data centres with architectural and resource constraints. Int. J. Auton. Adapt. Commun. Syst. **8**(4), 392–406 (2015)
15. H. Sun, P. Stolf, J.M. Pierson, G. Da Costa, Energy-efficient and thermal-aware resource management for heterogeneous datacenters. Sustain. Comput. Inf. Syst. **4**(4), 292–306 (2014)
16. M.K. Gupta, T. Amgoth, Resource-aware virtual machine placement algorithm for IaaS cloud. J. Supercomput. **74**(1), 122–140 (2018)
17. M. Abdel Basset, L. Abdle Fatah, A.K. Sangaiah, An improved Lévy based whale optimization algorithm for bandwidth-efficient virtual machine placement in cloud computing environment. Cluster Comput., **22**, 1–16 (2018)
18. F. Alharbi, Y.C. Tian, M. Tang, W.Z. Zhang, C. Peng, M. Fei, An ant colony system for energy efficient dynamic virtual machine placement in data centers. Expert Syst. Appl. **120**, 228–238 (2019)
19. N. Sharma, R.M. Guddeti, Multi-objective energy efficient virtual machines allocation at the cloud data center. IEEE Trans. Serv. Comput. **12**, 158–171 (2016)
20. M. Riahi, S. Krichen, A multi-objective decision support framework for virtual machine placement in cloud data centers: a real case study. J. Supercomput. **74**(7), 2984–3015 (2018)
21. A.F. Antonescu, P. Robinson, T. Braun, Dynamic SLA management with forecasting using multi-objective optimization, in *Proceedings of IFIP/IEEE International Symposium on Integrated Network Management*, pp. 457–463 (2013)
22. X. Chen, Y. Chen, A.Y. Zomaya, R. Ranjan, S. Hu, CEVP: cross entropy based virtual machine placement for energy optimization in clouds. J. Supercomput. **72**(8), 3194–3209 (2016)
23. S.E. Dashti, A.M. Rahmani, Dynamic VMS placement for energy efficiency by PSO in cloud computing. J. Exp. Theor. Artif. Intell. **28**(1–2), 97–112 (2016)
24. T.H. Duong Ba, T. Nguyen, B. Bose, T.T. Tran, A dynamic virtual machine placement and migration scheme for data centers. IEEE Trans. Serv. Comput. (2018)
25. R.N. Calheiros, R. Ranjan, A. Beloglazov, C.A. De Rose, R. Buyya, Cloudsim: a toolkit for modelling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Pract. Exp. **41**(1), 23–50 (2011)
26. CloudSim: A framework for modeling and simulation of cloud computing infrastructures and services. Available online at: http://www.cloudbus.org/cloudsim/
27. K.H. Kim, A. Beloglazov, R. Buyya, Power-aware provisioning of cloud resources for real-time services, in *Proceedings of 7th International Workshop on Middleware for Grids, Clouds and e-Science*, pp. 1–6 (2009)
28. P. Bohrer, E.N. Elnozahy, The case for power management in web servers, in *Power Aware Computing* (Kluwer Academic Publishers, US, 2002), pp. 261–289
29. A. Wierman, L. Andrew, A. Tang, Power-aware speed scaling in processor sharing systems: optimality and robustness. Perform. Eval. **69**(12), 601–622 (2012)

Survey on Cloud Auditing by Using Integrity Checking Algorithm and Key Validation Mechanism

M. Mageshwari and R. Naresh

Abstract Cloud computing can be used to access and storing data and delivery of different services over the internet. Using cloud storage, users can remotely store their data. Cloud service provider (CSP) provides data owners to store and access their valuable data in the cloud server and offers them to make use of on-demand data access without maintaining a local copy of their data. Even though this service avoids the data owners from making use of their third-party auditor, it certainly possesses serious security threats in maintaining the data owners cloud data. In addition, integrity is also an important issue in maintaining the data owners data stored in the cloud server. This survey presents an overview of integrity check and continuous auditing. The review work based on creating secure clouds by continuous auditing and cloud certification system (CCS) is used for high level security. This survey helps to provide the security by continuous auditing and overcome the integrity issues and to avoid the attacks by integrity checking algorithm and key validation mechanism. In this survey paper, various researchers' ideas based on integrity checking and key schemes have been analyzed as literature review.

Keywords Cloud computing · Third-party auditor · Integrity checking algorithm · User ID · Key validation mechanism

1 Introduction

Cloud storage can be used to save the files remotely. Different types of services are offered by cloud such as Software as a service, Platform as a service, and Infrastructure as a service. When the organizations increased, the data can be stored by multiple ways. But when we are using cloud services which provide data security, trust is good when compared to other storage device [1]. Cloud auditing can be defined as the on-site verification activity. Cloud security audits must check the security. Auditing is taken out by a third-party auditor in cloud computing. The TPA is mainly used for

M. Mageshwari · R. Naresh (✉)
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Chennai, Tamil Nadu 603203, India
e-mail: nareshr@srmist.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_41

IT auditors. The IT auditor does not access cloud resources directly. The auditor must understand what can be audited and tested. To obtain information, all reports can be collected via TPA which is the main role of TPA. Based on that, integrity checking algorithm hash value can be created for each file from cloud servers, then key validation mechanism provides the separate key for all the data users to avoid the attacks [2]. However, cloud services are important to changing environment, and the organizations are used because of the safety, anonymity, and dependability of cloud services. Cloud service can be used to establishing the trust and increase the transparency. Certification services and companies that sell cloud-related products, such as Amazon, give cloud computing a lot of publicity and exposure. Amazon Web Services and Google Microsoft and VMware are two companies that work together [3]. Various security management methods have been proposed in business auditing in the past literatures for ensuring the reliability and privacy. Continuous auditing can maintain the original and accurate details in business sides. When using index hashing algorithm, data tag replacement algorithm, proof of retrieval algorithm computation cost is high. When using data tags, it required more space.

1.1 These Review Contributions Are Summarized as Follows

- The establishment of a private key is done to reduce the danger of data owner theft.
- An integrity checking algorithm is designed to generate the hash key value before sending to the cloud. The algorithm does not need the high storage space, therefore, reduces the cost of the computation.
- Integrity and security can be maintained by continuous auditing.
- Security and privacy of the data are analyzed and proved.
- To reduce different attacks when using multiple data users.
- When using key validation mechanism, correct user can be able to access the data.

1.2 Cloud Auditing

Cloud auditing must check the data integrity, and it also improves the security and efficiency. The user needs to generate the key. Third-party auditor gives the key to all the data users, and it verifies if the key is sent to correct user or not [4].

1.3 Privacy Protection

While users access the data in cloud easily, cloud storage providers access their privacy information. Privacy security mechanisms are used to guarantee that these

data can be secret under malicious clients. It should allow the multiple data users without worrying attackers [5].

1.4 Continuous Auditing

As a result, continuous auditing (CA) of accreditation standards is necessary to ensure that the process is simple and reliable. Continuous auditing is maintaining accurate and original details that will be stored in cloud server [6].

1.5 Integrity Checking Algorithm

For protecting client data in a cloud environment, remote data integrity must be verified. Clients used it to submit and upload critical documents to the cloud. The uploaded file is divided into small blocks using the dynamic block age algorithm. If an intruder taints data in a multi-cloud environment, the continuous auditing process forces the verifier to perform file-level check. When data is stored in a remote storage location, the cloud user reduces control of the data; at this point, consumers may be unaware of security policies, vulnerabilities, and malware information. An intruder might, for example, use a back channel attack and a Denial of Service attack to gain access to cloud data [7].

1.6 Key Generation Mechanism

Users' public keys are generated using their unique identities, but their private keys are generated by either a private key production center or a private key production center (PKG). Once hacked, PKG may easily impersonate any user in order to forge tags while remaining undetected. Cryptography, on the other hand, is a secure alternative because the user's private key contains a concealed value as well as a partial key. The partial key is generated via the key generation procedure (KGM) [8]. Because the user's public key is provided by, certificateless cryptography further avoids the problem of certificate administration. The bulk of cloud auditing solutions in certificateless environments are primarily concerned with personal data. When people wish to share their data, other obstacles arise [9]. When users want to share data with other members of their community, other obstacles arise. Routine data audits, such as address tracking, may result in major assaults for the time being. As a result, it is critical to keep the names of community members disguised during data audits. Confidentiality, on the other hand, protects group members' identities while still revealing them [10].

2 Outline of the Review Work

Integrity checking gives the assurance for safe data in the cloud. The data table describes various algorithms methodologies advantages and disadvantages (Table 1).

3 Prevention of Integrity Security Attacks and Computation Time

This table describes to reduce and prevent various attacks like forgery attack, malicious attacks, and the computation time (Table 2).

4 Various Computation Cost of Auditing

See Figs. 1 and 2.

5 Conclusion

The paper based on a various surveys of different techniques is used to provide data integrity and verify the keys for each data user. It also provides the various techniques used, such as security reliability and privacy. Main thing of this survey is to achieve data integrity and verification of computation cost. As future work, it is in the way of achieving integrity and reducing attacks and computational cost.

Table 1 Outline of review work

S. No.	Author	Methodologies	Algorithm	Advantages	Limitations
1	Lin [11]	Embedded audit module and interceptor approach	Homomorphic algorithm	Automatic data monitoring with improved transparency and efficiency	EAM does not fulfill CA's technical requirement
2	Yang [12]	High-level design of the SAaaS architecture and business flow language can be implemented	Tag generation algorithm, prove and verification algorithm	Shared resources, multi-tenancy, infinite scalability and resources	Denial of service attack, cloud-wide incident detection

(continued)

Table 1 (continued)

S. No.	Author	Methodologies	Algorithm	Advantages	Limitations
3	Zhu [13]	Dynamic audit service for outsourced storage and integrity verification	Periodic sampling algorithm, index hash table	Low cost, scalable. To improve the performance of TPAs	Internal audits damage the integrity
4	Liu et al. [14]	Support formal analysis and fine-grained update requests	Ranked Merkel hash tree, bilinear mapping algorithm	Efficient data security, flexibility	Quality of service is inefficient when using low-level metrics
5	He et al. [15]	Bilinear pairing	Ternary hash tree algorithm, ranked Merkel hash tree algorithm	Verification shows the provable security in the cloud	High computation cost
6	Shen et al. [16]	Auditing is used for group user along with bilinear map, and Hellman is used to reduce the computational cost	Bilinear map algorithm	Decryption time reduced for the user	Computational cost is high
7	Ren et al. [17]	To reduce corrupted data	Privacy preserving algorithm	To recover the data corruption	It can increase data integration efficiency
8	Stephanoward and Fallenbeck [18] [R3]	Bottom up approach and low-level metrics can be used to provide the security	Index hash table	Auditing is trust. Continuous auditing provides the security	Low-level metrics formed complex metrics
9	Shen et al. [19]	Public auditing protocol and blockless verification	MD5 algorithm	Communication cost is less	The proof and verification cost is high
10	Sridhar and Smys [20]	AES encryption and hybrid multilevel authentication algorithm	AES algorithm	Against internal and external attacks by the AES encryption method	The computation and correctness could be checked

(continued)

Table 1 (continued)

S. No.	Author	Methodologies	Algorithm	Advantages	Limitations
11	Yang and Jia [21]	Well-organized storage protocol can be used to provide security Privacy	Proof of retrieve	Privacy is better compared other techniques	Risks can be identified difficult
12	Praveena and Smys [22]	To secure the cloud data. The owner of the private key performs the decryption	Encryption and decryption algorithm	Possible for high security	To be evaluated based on different attacks
13	Shen et al. [23]	To check the data integrity by using algebraic signature	Ternary hash tree algorithm	High efficiency. Computation cost is less	More privacy
14	Yu et al. [24]	Identity-based key homomorphic cryptographic primitives to develop a system for remote integrity checking	Remote integrity, cryptography algorithm	The computation cost is low	Computation cost needs to be reduced
15	Shen et al. [19]	To prevent key exposure based on strong key exposure resilient	AES algorithm	High efficiency. More privacy	High storage overloads. More download time to be needed
16	Aujla et al. [25]	Cloud, safe storage verification, and auditing are used	Merkel hash tree algorithm	Storage and performance should be considerable	High communication and low effectiveness
17	Sebastian [26]	Create a conceptual architecture identify key components	SHA algorithm	Increasing openness and building confidence in the cloud sector to shorten the auditing process	It should not clarify manage certification violations

(continued)

Table 1 (continued)

S. No.	Author	Methodologies	Algorithm	Advantages	Limitations
18	Wu [27]	POS system and POR protocol to prove the security and authentication	Multiple precision algorithms	It provides security and privacy. Evaluation time and cost are low	To achieve security, privacy preserving does not support validation of time stamp
19	Yang [28]	Using the block chain technology to create a compelling opportunity for data visitors	Index hashing algorithm	The TPA and consumers are stateless. They do not need to keep track of data index information	Malicious attack occurred when group of user access the data 0at same time. Malicious attack affects the internal and external audit
20	Li [29]	Lattice-based privacy preserving implemented stateless auditing scheme used	Privacy preserving algorithm	To avoid quantum attack. To achieve provable security	If any hardware device failure, malicious hacks and software bugs raised

Table 2 Security attacks in integrity checking

Author	Security against attack	Computation time
Cao et al. [30]	Forgery attack	Less computation time
Apolinario et al. 31]	Spoofing attack	Minimizes storage overhead
Mahalakshmi and Suseendran [32]	Malicious attack	High storage overhead
Liu et al. [33]	Denial attack, malicious attack	Lower computational cost
Tang et al. 34]	Prevents from replay attack	Minimize computation and storage overhead
Hong et al. [35]	Prevents from forgery attack	Lower computational cost
Lu et al. [36]	Prevents from Dos attacks	Lower computation cost
Ramanan and Vivekanandan [37]	Reduced malicious attack	Lower storage overhead
Sasikala et al. [38]	Prevent quantum computer attacks	Lower computational overhead
El Ghoubach [39]	Prevents from forgery attack	Computational and communication overhead
Rady [40]	Prevents from DDOS attack	Computational overhead is more

Fig. 1 Computation cost

Fig. 2 Level of integrity checking

By reviewing previous studies using TPA, key validation mechanism and integrity checking algorithm are the best solution for cloud auditing.

References

1. A. Saranya, R. Naresh, Cloud based efficient authentication for mobile payments using key distribution method. J. Ambient Intell. Human. Comput. (2021). https://doi.org/10.1007/s12652-020-02765-7
2. R. Naresh, P. Vijayakumar, L. Jegatha Deborah, R. Sivakumar, A novel trust model for secure group communication in distributed computing. Special Issue for Security and Privacy in Cloud Computing, J. Organizational End User Comput. **32**(3), 1–14 (2020). https://doi.org/10.4018/

JOEUC.2020070101
3. A. Saranya, R. Naresh, Efficient mobile security for E-health care application in cloud for secure payment using key distribution. Neural Process Lett (2021). https://doi.org/10.1007/s11063-021-10482-1
4. K. Yang, X. Jia, An efficient and secure dynamic auditing protocol for data storage in cloud computing. IEEE Trans. Parallel Distrib. Syst. **24**(9), 1717–1726 (2013)
5. R. Naresh, M. Sayeekumar, G.M. Karthick, P. Supraja, Attribute-based hierarchical file encryption for efficient retrieval of files by DV index tree from cloud using crossover genetic algorithm. Soft Comput. **23**(8), 2561–2574 (2019). https://doi.org/10.1007/s00500-019-03790-1
6. Y. Yu, M.H. Au, G. Ateniese, X. Huang, W. Susilo, Y. Dai, G. Min, Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage. IEEE Trans. Inf. Forensic Secur. **12**, 767–778 (2017)
7. P. Vijayakumar, R. Naresh, L. Jegatha Deborah, S.K. Hafizul Islam, An efficient group key agreement protocol for secure P2P communication. Secur. Commun. Netw. **9**(17), 3952–3965 (2016). https://doi.org/10.1002/sec.1578/abstract
8. P. Vijayakumar, R. Naresh, S.K. Hafizul Islam, L. Jegatha Deborah, An effective key distribution for secure internet Pay-TV using access key hierarchies. Secur. Commun. Netw. **9**(18), 5085–5097 (2016). https://doi.org/10.1002/sec.1578/full
9. R. Naresh, A. Gupta, Sanghamitra, Malicious URL detection system using combined SVM and logistic regression model. Int. J. Adv. Res. Eng. Technol. (IJARET) **10**(4), 63–73 (2020)
10. M. Meenakshi, R. Naresh, S. Pradeep, Smart home: security and acuteness in automation of IOT Sensors. Int. J. Innov. Technol. Explor. Eng. (IJITEE) **9**(1), 3271–3274 (2019)
11. C.-C. Lin, F. Lin, D. Liang, An analysis of using state of threat technologies to implement real-time continuous assurance. in *Proc. 6th World Congress Serv.*, Miami, FL, USA, pp. 415–422 (2010)
12. N.K. Yang, X. Jia, Data storage auditing service in cloud computing: challenges, methods and opportunities. World Wide Web **15**(4), 409–428 (2012)
13. Y. Zhu, G.-J. Ahn, H. Hu, S. Yau, H. An, C.-J. Hu, Dynamic audit services for outsourced storages in clouds. IEEE Trans. Serv. Comput. **6**(2), 227–238 (2013)
14. C. Liu, J. Chen, L. Yang, X. Zhang, C. Yang, R. Ranjan, K. Ramamohanarao, Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates. IEEE Trans. Parallel Distributed. Syst. **25**(9), 2234–2244 (2014)
15. K. He, C. Huang, J. Shi, J. Wang, Public integrity auditing for dynamic regenerating code based cloud storage, in || *IEEE Symposium on Computers and Communication (ISCC),* (2014)
16. W. Shen, J. Yu, H. Xia, H. Zhang, X. Lu, R. Hao, Lightweight and privacy-preserving secure cloud auditing scheme for group users via the third party medium. J Netw. Comput. Appl. **82**, 56–64 (2017)
17. Z. Ren, L. Wang, Q. Wang, M. Xu, Dynamic proofs of retrievability for coded cloud storage systems. IEEE T Serv. Comput. **11**(4), 685–698 (2015)
18. P. Stephanow, N. Fallenbeck, Towards continuous certification of Infrastructure-as-a-service using low-level metrics, in *Proceedings of the 12th IEEE International Conference on Advanced Trusted Computing*, Beijing, China, pp. 1–8 (2015)
19. J. Shen, D. Liu, D. He, X. Huang, Y. Xiang, Algebraic signatures-based data integrity auditing for efficient data dynamics in cloud computing, in IEEE Trans. Sustain. Comput. (2017)
20. S. Sridhar, S. Smys, A hybrid multilevel authentication scheme for private cloud environment, in *Proceedings of IEEE 10th International Conference on Intelligent Systems and Control (ISCO)*, pp. 1–5 (2016)
21. K. Yang, X.H. Jia, An efficient and secure dynamic auditing protocol for data storage in cloud computing, (in English), IEEE Trans. Parallel Distrib. Syst. **24**(9), 1717–1726 (2017)
22. A. Praveena, S. Smys, Ensuring data security in cloud based social networks, in *Proceedings of IEEE International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 2, pp. 289–295 (2017)
23. W. Shen, J. Qin, J. Yu, R. Hao, J. Hu, Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage. IEEE Trans Inf Forensic Secure **14**(2), 331–346 (2016). http://dx.doi.org/10.1109/TIFS.2018.2850312

24. J. Yu, H. Wang, Strong key-exposure resilient auditing for secure cloud storage. IEEE Trans. Inf. Forensic Secure **12**(8), 1931–1940 (2017). http://dx.doi.org/10.1109/TIFS.2017.2695449
25. G.S. Aujla, R. Chaudhary, N. Kumar, A.K. Das, J.J. Rodrigues, SecSVA: secure storage, verification, and auditing of big data in the cloud environment. IEEE Commun. Mag. **56**, 78–85 (2018)
26. W. Sebastian, J. Yu, H. Xia et al., Light-weight and privacy-preserving secure cloud auditing scheme for group users via the third party medium. J. Netw. Comput. Appl. **82**, 56–64 (2018)
27. H. Yu, Y. Cai, R.O. Sinnott, Z. Yang, ID-based dynamic replicated data auditing for the cloud. Concurrency Comput. **31**(11), e5051 (2019)
28. Y. Yang, J. Yu, R. Hao, C. Wang, K. Ren, Enabling efficient user revocation in identity-based cloud storage auditing for shared big data. IEEE Trans. Dependable Secure Comput. **17**(3), 608–619 (2020)
29. J. Li, H. Yan, Y. Zhang, Efficient identity-based provable multi-copy data possession in multi-cloud storage. IEEE Trans. Cloud Comput. (2020)
30. L. Cao, W. He, Y. Liu, X. Guo, T. Feng, An integrity verification scheme of completeness and zero-knowledge for multi-cloud storage. Int. J. Commun. Syst. 1–10 (2017)
31. F. Apolinario, M.L. Pardal, M. Correia, S-audit: efficient data integrity verification for cloud storage, in *IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 465–474 (2018)
32. B. Mahalakshmi, G. Suseendran, An analysis of cloud computing issues on data integrity, privacy and its current solutions. Research Article, pp. 467–482 (2019)
33. A. Liu, H. Fu, Y. Hong, J. Liu, Y. Li, LiveForen: ensuring live forensic integrity in the cloud. IEEE Trans. Inf. Forensics Secur. **14**(10), 2749–2764 (2019)
34. X. Tang, Y. Huang, C.-C. Chang, L. Zhou, Efficient real-time integrity auditing with privacy-preserving arbitration for images in cloud storage system. IEEE J. **7**, 33009–33023 (2019)
35. H. Jiang, M. Xie, B. Kang, C. Li, L. Si, ID-based public auditing protocol for cloud storage data integrity checking with strengthened authentication and security. Wuhan Univ. J. Nat. Sci. **23**(4), 362–388 (2018)
36. X. Lu, Z. Pan, H. Xian, An integrity verification scheme of cloud storage for internet-of-things mobile terminal devices. J. Comput. Secur. **92**, 1–17 (2019)
37. M. Ramanan, P. Vivekanandan, Efficient data integrity and data replication in cloud using stochastic diffusion method. Cluster Comput. 14999–15006 (2018)
38. C. Sasikala, C. Shoba Bindu, Certificateless remote data integrity checking using lattices in cloud storage. Neural Comput. Appl. 1513–1519 (2018)
39. I.E. Ghoubach, R.B. Abbou, F. Mrabti, A secure and efficient remote data auditing scheme for cloud storage. J. King Saud Univ. Comp. Inf. Sci. 1–7 (2019)
40. M. Rady, T. Abdelkader, R. Ismail, Integrity and confidentiality in cloud outsourced data. Ain Shams Eng. J. Sci. Dir. 275–285 (2019)

Privacy-Preserving Min–Max Value Retrieval Using Data Obfuscation Algorithm in Distributed Environment

Bhople Yogesh Jagannath

Abstract A comprehensive and efficient p-based algorithm is being proposed to retrieve minimum and maximum of particular data field of database tuple distributed as a horizontal partitioning of single database table among the multiple servers without losing the privacy of individual server. Data obfuscation technique is used in the proposed algorithm to preserve the privacy of individual server. This algorithm is useful in many practical scenarios where privacy-preserving data computing is desirable.

Keywords Privacy-preservation · Distributed environment · Data obfuscation · Horizontal partitioning · Min–max value retrieval · Distributed servers

1 Introduction

Innovation has brought the drastic fall down in semiconductor prices so that system planner over the globe is more inclined toward the distributed system design connected through network link instead of single stand-alone system. Distributed system has multiple merits like scalability, fault-tolerant, reliability, load balancing, and improved response time. The necessities like multiuser configuration, resource sharing, and some form of communication between the workstations have created a new set of problems with respect to privacy, security, and protection of the system as well as the user and data [1].

B. Y. Jagannath (✉)
Department of Information Technology, Government Polytechnic Washim, Washim, India
e-mail: ybhople90@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_42

2 Proposed Data Obfuscation Algorithm

2.1 Simulation Example

Consider a class with 10 students, each student is having his/her unique rank. Teacher wants to retrieve minimum and maximum rank in his class. No student wants to share his/her rank with anyone, even not with the class teacher. Students are ready to offer any help until and unless their privacy is preserved. Using the following data obfuscation algorithm, teacher retrieves the minimum and maximum rank values with due respect to each student's privacy.

Steps of the proposed data obfuscation algorithm [3] for simulated example.

1. Generate a random number 'X' and pass it to first student. Ask him/her to add his/her rank with 'X' and pass the sum $X + R_1$ to 2nd Student.
2. Second student will add his/her own rank with sum he/she got and pass the sum $X + R_1 + R_2$ to 3rd student, 3rd will pass to 4th, 4th will perform similar action and so on till the 10th students. 10th student will pass the whole sum, i.e., $X + \sum_{i=1}^{10} R_i$), to teacher.
3. Teacher will simply subtract X from the sum and will get sum as $\sum_{i=1}^{10} R_i$.
   Calculate average of all rank as

$$\text{Sum} = \sum_{i=1}^{10} R_i$$

$$\text{Avg} = \text{Sum}/10$$

   Initialize the counter variable CountMin = 0, CountMax = 0, SumMin = Sum, SumMax = Sum, AvgMin = Avg, AvgMax = Avg and pass these value students along with following instruction.

```
(i)   if(R_i ≥ AvgMin) {
            SumMin=SumMin-R_i ;
      }
      else
      CountMin++;
(ii)  if(R_i ≤ AvgMax) {
      SumMax=SumMax-R_i ;
      }
      else
      CountMax++;
```

   For steps 1, 2 and 3, refer Fig. 1.
4. Calculate new average values as AvgMin = SumMin/CountMin
   AvgMax=SumMax/CountMax.

Fig. 1 Flowchart of minimum–maximum rank retrieval

Pass AvgMin, AvgMax. CountMin=0, CountMax=0, SumMin=Sum, SumMax=Sum to students along with above mentioned two instruction to do.

5. Repeat Step 4 along with two instructions in step 2 until we get CountMin = 1 or 0, CountMax = 1 or 0.
6. Value in SumMin indicates Minimum Rank and value in SumMax indicates Maximum Rank.

Minimum–maximum rank retrieval of steps 4, 5, and 6 are given under the synchronization bar in Fig. 1.

2.2 Illustration of Proposed Algorithm with Working Data Set Example

Consider a set of 10 students with their Ranks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, respectively. Applying the above-mentioned algorithm stepwise on this data set. Purpose is to find minimum and maximum rank of student without losing the privacy of students as their rank is secret for them, and they are ready to reveal it to anyone, but they are ready to offer any help as long as their privacy is preserved.

1–2. Pass a random no. X to 1st student and get $(X + \sum_{i=1}^{10} i)$. Subtract X

3. Get Sum $= \sum_{i=1}^{10} R_i$ Calculate Avg = Sum/10 = 5.5

   Initialize the counter variable CountMin=0, CountMax=0, SumMin=55, SumMax=55 . And pass this to students along with Instruction in Step 3 of above algorithm.

4. Get CountMin = 5, SumMin $= (55 - \sum_{i=6}^{10} i) = 15$,

   CountMax = 5, SumMax $= \left(55 - \sum_{i=1}^{5} i\right) = 40$,

   Calculate AvgMin=SumMin/CountMin=15/5=3,

   AvgMax=SumMax/CountMax=40/5=8.

   Next Iteration-

   Pass the following values along with two instructions mentioned in step 3 of Sect. 2.1.

   CountMin=0, CountMax=0,

   SumMax = sum, SumMin = sum,

   AvgMin = 3, AvgMax = 8.

   Get CountMin = 2, SumMin $= \left(55 - \sum_{i=6}^{10} i\right) = 3$, CountMax = 2,SumMax $= (55\text{-}\left(55 - \sum_{i=1}^{8} i\right)) = 19$, AvgMin = SumMin/CountMin = 3/2 = 1.5, AvgMax = SumMax/CountMax = 19/2 = 9.5

   Next Iteration-

   Pass the following values along with two instructions mentioned in step 3 of Sect. 2.1

   CountMin = 0, CountMax = 0,

   SumMin = Sum, SumMax = Sum,

   AvgMin = 1.5, AvgMax = 9.

5. Get CountMin = 1, SumMin $= \left(55 - \sum_{i=2}^{9} i\right)$, CountMax = 1, SumMax $= \left(55 - \sum_{i=1}^{9} i\right)$.

6. CountMin = 1, so value in SumMin indicates minimum Rank.

   CountMax = 1, so value in SumMax indicates maximum Rank.

EmpId	Name	Dept	Salary	Rank	DoJ	
123	John	Manu.	20000	100	2010	Partition 1
234	Alice	Prod.	30000	200	2009	
456	Bob	HR	40000	300	2001	
231	Coop	Acc	20000	301	2002	Partition 2
233	Sanjay	Invent.	30000	201	2010	
465	Yogesh	Markt.	40000	302	2009	
654	Om	Manu.	20000	400	2001	Partition 1
789	Vikas	Prod.	20000	501	2002	
987	Bob	HR	30000	700	2011	
125	Sanjay	Manu.	40000	800	2012	

Fig. 2 Horizontal partitioning of data sets

2.3 Analogy with Distributed Servers

In the above-simulated example, consider that instead of students, 10 different distributed servers are having a copy of horizontally partitioned data set, of which particular data field (e.g., rank in the above-simulated example) is private for that server and does not want to disclose with any client or server. Aim is to retrieve the minimum and maximum value of that particular data field without compromising the privacy of any individual server. The above-mentioned algorithm when executed in iterative manner produces minimum and maximum values without compromising the privacy of any server in very efficient manner as these servers publish obfuscated data, not the real data. [4]

Figure 2 explains the concept of horizontal partitioning of data sets, where first partition consists of 3 tuples, second partition consists of 3 tuples, and third partition consists of 4 tuples.

3 Complexity Analysis of Proposed Data Obfuscation Algorithm

3.1 Time Complexity Analysis

Assuming that there are n servers, and rank values are uniformly distributed between minimum and maximum, $\sqrt{n}$ iterations are carried out, and each iteration has to go

through n servers, so average-case and worst-case time complexity of the algorithm will be $O(n\sqrt{n})$. [5]

In the best case, where all servers having same value of rank, in just 2 iteration, the algorithm produces minimum and maximum rank $O(2 * n) \equiv O(n)$.

Assuming that there are 8 to 10 finite number of servers, average-/worst-case time complexity will be $O(\text{Constant} * \sqrt{n}) \equiv O(\sqrt{n})$, and best case complexity will be $O(\text{constant}) \equiv O(1)$.

3.2 Space Complexity Analysis

Each server requires constant space (space to store 4 integer variables). So, n servers will require $4 * n$ integer space. So time complexity will be $O(4 * n) \equiv O(n)$. Assuming that there are fixed no. of 8 to 10 distributed servers, space complexity will be $O(4 * \text{Constant}) \equiv O(1)$.

4 Conclusion

This p-based data obfuscation algorithm (worst/average time complexity $O(\sqrt{n})$) can have different kinds of utility in distributed environment where a particular relational schema is horizontally partitioned and distributed among n number of servers. Privacy-preserving min-max value retrieval is one of those utilities. In the future, if some distributed system design needs this utility, it can be easily implemented. It can be a most suitable algorithm for distributed key-value stores where privacy-preserving min-max value retrieval is desirable. As its time complexity is $O(\sqrt{n})$, it is computationally efficient, algorithmic latency will be less (ignoring network delays/latency).

References

1. M.R. Ogiela, L. Ogiela, U. Ogiela, Security and privacy in distributed information management, in *2014 International Conference on Intelligent Networking and Collaborative Systems*, INSPEC Accession Number: 14985068. https://doi.org/10.1109/INCoS.2014.108
2. S. Rao, Distributed system: an algorithmic approach. IEEE Distrib. Syst. **9**(11) (2008)
3. D.E. Bakken, R. Parameswaran, D.M. Blough, A.A. Franz, T.J. Palmer, Data obfuscation: anonymity and desensitization of usable data sets. IEEE Secur. Privacy Mag. **2**, 34–41 (2004)
4. J. Hamm, A.C. Champion, G. Chen, M. Belkin, D. Xuan, Crowd-ml: a privacy-preserving learning framework for a crowd of smart devices, in *IEEE International Conference on Distributed Computing Systems* (2015)
5. T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, 2nd edn. (The MIT Press, 2001)

Intelligent Virtual Research Environment for Natural Language Processing (IvrE-NLP)

Prashant Chaudhary, Pavan Kurariya, Shashi Pal Singh, Jahnavi Bodhankar, Lenali Singh, and Ajai Kumar

Abstract In the 21st century, natural language processing (NLP) has obtained much prominence for human–machine interaction (HMI). With this interest in natural language processing (NLP) has grown significantly, numerous NLP tools (e.g., morphology, the tagger, and a parser, etc.) have been developed all over the world. Despite having huge importance and requirements, we have noticed gaps for having a comprehensive single framework or platform, which encompass all NLP-related tools and technologies for promoting the research in NLP and sharing the knowledge and resources among NLP researchers required for understanding and building the solution for HMI. Our objective is to apply Software engineering in natural language processing with the concept of an object-oriented model by using a collection of reusable objects by defining the communication protocol, consisting of a set of rules that must be applied to exchange data between two NLP modules. We proposed state of art ivrE—A virtual environment for creating, modifying, executing, and analyzing various NLP solutions and technology. The proposed idea is broadly based on to define own ivrE-NLP object framework model that permits the developer to create, modify, and execute the application and analyze their outcomes by operations on visual representations of the modules. A variety of NLP-based applications (tools, modules, and plugins) already exist, they can publish into store available with environment so it can be used by research community at large. To develop complete NLP

P. Chaudhary (✉) · P. Kurariya · S. P. Singh · J. Bodhankar · L. Singh · A. Kumar
Centre for Development of Advanced Computing, Pune, India
e-mail: cprashant@cdac.in

P. Kurariya
e-mail: pavank@cdac.in

S. P. Singh
e-mail: shashis@cdac.in

J. Bodhankar
e-mail: jahnavib@cdac.in

L. Singh
e-mail: lenali@cdac.in

A. Kumar
e-mail: ajai@cdac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_43

framework or platform, we require much more than just assembling or collecting these tools or modules at one place, no matter how good any tool or module is working individually. It requires not only the standards and a set of protocols, but also requires a compliant composition than a pre-defined algorithms and their implementation. In brief, we require a comprehensive open framework to bundle, manage, and integrate set of NLP tools, modules, components, applications, algorithms, and define their associated rules, comprehensive data structures, and knowledge.

Keywords Natural language processing (NLP) · Human machine interaction (HMI) · Intelligent virtual research environment for natural language processing (ivrE-NLP) · Object-oriented model (OOM)

1 Introduction

The original motivation for Intelligent Visual Research Environment for natural language processing (ivrE-NLP) is to support reusing of the existing software and to encourage NLP research groups to explore, analyze, modify, enhance, and develop NLP tools/applications/modules, etc. by combining NLP components already developed. The principal objective of a virtual environment is to accelerate the research in NLP field for developing fast, robust, reusable, and authentic NLP tools/modules/applications which in turn help in the development of the application/system required for HMI.

The objective is to apply Software engineering in NLP with concept of object-oriented model by using collection of reusable objects. Basic idea is to provide such a compliant framework that one can design an NLP system in the best possible way for any given purpose. Furthermore, the ability to insert module into an existing application enables a system to grow in a flexible and modular fashion.

Proposed ivrE-NLP framework/platform will be helpful in developing, executing, and evaluating the outcome of complex and multi-functional NLP systems furthermore provided Visual environment to analyze the components belongs to the built system. Its efforts to meet the following objectives:

- Define information interchange protocol between components/modules of a built NLP system providing lose coupling without changing original logic or theoretical approach. It will allow modules to define their own custom information share protocol mutually accepted by components of system
- Define the integration protocol to incorporate modules in the NLP system written in any source language includes source code, binary code, or executable format
- Support comprehensible graphical interface provide facilities for managing NLP resources during building a system/application.

2 Literature Survey

Dedicated development environments for NLP have been considered or developed for various purposes for the NLP research in the past. Some development environments are conceptualized to aim at the syntactic aspect of natural language and are targeted toward language experts as audience rather than software developers and less work has been done to capture the entire paradigm of NLP. One of the similar works INTEX platform [1] is designed for the language experts to use natural language in general, while LexGram [2] intends on categorical grammars. Some of the IDEs that are most related to development are GATE [3], VisualText4, and NL-OOPS [4]. However, GATE focuses on a uniform platform for NLP-related tasks and Visual Text is specialized in the development of user-defined text analysis applications. However, lips [5] use them in the background thereby considering integrating natural language as a major language for the design flow of software and systems. NL-OOPS is the method nearer to lips, however, it does not insist on the integration with other modeling and programming languages in a common Integrated Development Environment [IDE].

The idea of intelligent information extraction for natural language specifications has been intensively presented in past research work. Saeki et al. [6, 7] studied an approach like [8, 9] that associates lexicon in English text to software concepts found in the object-oriented programming paradigm. Much similar research approaches were considered for pre-programmed methods and restricted the input language to achieve this goal. Some approaches make use of an algorithm [10] or determine recurring designs in property specifications [11].

3 Research, Development, and Integration Environment for NLP

Introducing a Virtual research environment in the field of natural language processing in which researcher allows developing their own application using various NLP components, which are either pre-existing within setup or their own plugin or in the form of services. Thus, ivrE-NLP works as a platform for combining software programming with theoretical knowledge. ivrE-NLP provide initial building blocks and comprises three principal elements:

- Configuration Manager: defines communication protocol based on the object-oriented data model approach, information about various applications is stored in configuration file in form of XML file, which contains annotations for each application, i.e., unique id, a type and dependency file (required resource to run the application).
- Developer Kit: Provide access to various applications and collection of reusable components or modules in the form of object, these objects communicate with each other based on protocol defined by the ivrE-NLP.

- Graphical User Interface: Provide an integrated environment of all available NLP-based tools/modules/applications for design development and execution. It is an easy-to-use visual interface with a 'plug-and-play' design.

Most NLP modules are developed by Iterative and incremental development software models, which require regular analysis and improvement repetitively. Every NLP module must define its dependency in the configuration file along with input and output format so another module can use that information to communicate with it without knowing the internal algorithm.

The researcher can assemble a system with the help of a visual representation of NLP Modules in form of object. Each object can be executed individually or in an integrated pipeline design by the developer. ivrE-NLP provides the facility to analyze results systematically for building, running, and investigating various NLP modules.

We designed ivrE-NLP as a virtual development and integration research environment to support the visual representation of NLP modules for developing analysis and executing entire systems such as machine translation, speech, information extraction, and retrieval.

The visual environment offers executable data objects in form of a tree data structure connected by a leaf or communication protocols, data dependency declarations. Each module contains a configuration file for linking with other modules and generates a log file for viewing and comparing modules results by relating data produced by modules, thus an integrated visual development environment leads to the rapid development of NLP-related research (Fig. 1).

ivrE-NLP allows researcher to develop new applications with the help of pre-existing modules comes with in environment by 'drag and drop' objects and connecting the dots by flowcharts diagram. ivrE-IDE offers various features such as:

- A graphical user interface (GUI) that allows users to organize the NLP tools
- A complete environment with pre-existing NLP modules that work together seamlessly
- Project explore to create/ update application with version control
- Flexible architecture for combining multiple modules
- Review or evaluation facility within development environment (Fig. 2).

ivrE-NLP IDE enables contributors other than the creators of the application to use, enhance, and analyze the NLP modules. The benefits of this facility include:

- Stimulated research in NLP field with reusable components
- Organized and managed engineered application with version control
- Evaluation and review for improvising performance of individual module
- Building profession and maintainable applications for real world
- Tight integration of modules, which improve overall productivity.

Fig. 1 ivrE-NLP IDE package

Fig. 2 ivrE-NLP IDE component integration

4 System Architecture

4.1 An Overview of Workflow Model

ivrE-NLP consists of Language resources and Engineering Solutions as an object. Both data and Grammar are types of Language Resource (LRs); all LRs have a data structure linked with them that stores information in form of attribute/value pairs.

Each Document = content + annotations + features where Annotations* where annotations have a start symbol and an end symbol, an ID, a type, and a Map (Fig. 3).

*Annotations (start Node, end Node, ID, type, Map).

Workflow model is the distinct object model which has been used for the ivrE-NLP. This model has been extensively used in natural language processing for a variety of applications. In the above model, modules are typically represented as visual diagrams and connecting lines between the diagrams. Each NLP module performs specific task ranging from preprocessing of text to parsing of sentences and generation of whole texts. The user equipped various facilities to assemble the entire system by using visual diagrams and even execute the system by simply run a command and analyzing results by examining log files generated by the NLP module, so an integrated visual environment helps users for building, running, and analyzing NLP modules.

Fig. 3 ivrE-NLP workflow

4.2 ivrE-NLP Paradigm

An ivrE-NLP is a visual environment into which various NLP modules reside. These NLP modules are loosely coupled with each other, they communicate entirely by means of the standard protocol defined by the framework. Inter-modules communication is handled by the workflow model described above along with NLP tools, language resources, and events.

The basic concept that needs to be understood by the researcher before using ivrE-NLP environment is the NLP data model which defines all the resources including input and output for all modules. It uses a standard model (Document = content + template + features) where template (start Node, end Node, ID, type, Map) into the sources document which is characterized by a type and a set of information represented in Map as attribute-value pairs. The Map stores information datasets. The data format is not dependent or restricted to any NLP formalism, to enable the use of modules based on different indigenous algorithms. This generalization helps to facilitate the representation of a wide variety of linguistic information, ranging from very simple tokenizer to very complex parser (Fig. 4).

Fig. 4 A ivrE-NLP paradigm

Fig. 5 A ivrE-NLP framework detailing

4.3 ivrE-NLP Framework: (Model, View, Controller Approach)

To provide simple and convenient Graphical Environment in ivrE-NLP framework an integrated architecture having three major sub-components has been provided:

- Language Resources (Model)—represent entities such as lexicons, corpora, and grammar
- Visual Resources (View)—represent visualization and user interfaces components that participate in GUIs.
- Programming Resources (Controller)—represent entities that are primarily algorithms/logic (Fig. 5).

The set of resources provided with ivrE-NLP is known as NLP objects which is a collection of reusable components for natural language processing-based research. All the resources are in the form executable.lib along with the configuration xml file. While using IDE to develop any application, the developer needs to use the toolbar that comes with the framework that involves Graphical User interface (GUI), business logic, and language resources altogether (Fig. 6).

4.4 Communication Protocols and Data Dependency

An ivrE-NLP model consists of NLP tools and technology that could interact with each other in two ways. First, they could express their dependencies using the required statement in their plugin.xml. If plugin A requires plugin B, plugin A can see all the

Fig. 6 A ivrE-NLP IDE framework detailing

NLP classes and resources from B, LP class visibility conventions. Each module had a namespace, and they could also specify the versions of their dependencies.

In the NLP data model, the pipeline to be executed is not specified explicitly. Instead, only the requirement of the individual module along with pre-conditions and post-conditions are specified in the configuration file. These dependencies are analyzed, and then a module execution sequence is arranged.

Module dependencies require a commutation protocol to be established for interacting between modules. The configuration file provided along with each module includes a set of annotations or attributes which required being set before execution.

There is a provision to set dependencies via parameters in configuration file, i.e., config.xml.

- Each parameter has one or more attributes, where each attribute has a set of the possible input value
- Each parameter has one or more attributes, where each attribute has a set of the possible output value
- One or more parameter included with set of attributes where each attribute may have a specified value.

An ivrE-NLP model consists of NLP tools and technology that could interact with each other in two ways. First, they could express their dependencies using the required statement in their plugin.xml. If plugin A requires plugin B, plugin A can see all the NLP classes and resources from B dependencies (Fig. 7).

Basically, declarations of pre-and post-conditions are required for each module. A concrete example of these declarations, as specified in an XML configuration files in ivrE-NLP. The pre-conditions indicate that parameter with attributes of object Tagger must be available before it will run. The post-conditions indicate that the attribute

Fig. 7 A ivrE-NLP communication protocol and dependency

that the parser has run will be set and that parameter of object Parse Derivation Tree with attributes of type Derived Tree (produced by Generator) will be produced.

4.5 Developing New NLP System in ivrE-NLP

The Researcher creates a new project by selecting various options available in ivrE-NLP IDE. If necessary, the developer needs to configure the config.xml file to the registered dependency of the system. Once the project has been constructed developer can add various.lib or jar file with the project which he can select from NLP lib section. Developers can write specific wrappers to invoke the open-source tools available with the framework. There is a configuration file associated with each module in which the module's data dependencies must be defined, as well as the specifications for obtaining its result specified. Each project has a configuration manager interface and the developer need to specify database connectivity and port before execution of the system. Once the configuration file has been completed, the developer selects "Run Command" from the NLP IDE top-level menu to compile and execute the application. Developer can "save" the existing project in workspace provided by IDE for future use.

4.6 Assemble Customized System via ivrE-NLP

Allow researchers to build customized NLP systems easily from a widely available language processing components/modules is a key feature of an ivrE-NLP platform.

In any research or development environment the set of compatible modules are obtainable and only few of them are relevant for other modules to develop system. ivrE-NLP provide an option for researchers where it views the set of available modules, their data dependency and can select the application specific module could be integrated straightforwardly. The visual approach taken to implement following functionality:

- Super-graph visualizes instantly available modules with their linked data dependencies.
- A sub-graph visualizes the sub modules details can be easily achieved by simple mouse operations on sub-graph symbol in IDE and yields a system that can be run directly.
- A new system module can be incorporated visually into the existing graph hierarchy of modules.

4.7 Executing NLP System in ivrE-NLP

This is a helpful feature for NLP component developers, where developer possible to obtain instant comment on changes made to the underlying code or resource having dependency exist with system components. It provides a facility to execute the module using IDE, obtain result, evaluate observation, rest the module and replay execution on the upgraded requirement. It encompasses system graph thought of as a detailed assembled graphical process of executable system and useful to evaluate the workflow of system. The system graph uses the color symbols to indicate the current execution state of modules such as green modules indicate ready to execute, red modules tell modules have already been executed, and cannot be executed further. To execute module in ready mode, user should select the ready option with one mouse click and perform execution on second click. In the processing, module symbol becomes red, and the dependent downstream modules automatically turn green.

4.8 Viewing and Analyzing Results in ivrE-NLP

ivrE-NLP provide framework cover one of the important aspects where the researcher can visualize the result of system modules help to analyze the integrated NLP system. NLP systems are complex includes components, interaction protocols, and their development efforts consists of iterative refinement on test data. It provides a provision to define separate results viewers that can be reusable by various NLP modules. It includes the comparation interface to evaluate the outcome of two modules based on the input data, useful to improve the modules of the system. ivrE-NLP also gives the plugin option to integrate the standard tool encourage reusability to compare results among sets of output from a module improve efficiency on outcome of system.

4.9 Statistics and Report Generation in ivrE-NLP IDE

The Report generation feature is designed to provide a user-friendly interface for managing system evaluation reports within ivrE-NLP. Moreover, the Report generation is also useful for comparing various versions of applications with the help of graphs and charts, so the researcher is able evaluate the performance of individual modules in terms of compilation time and memory. In addition, the researcher also allows assembling system with open-source available modules and investigating which module is working well in pair with their NLP system.

5 Conclusions

In this paper, we presented ivrE-NLP a virtual environment for NLP-based solutions and technologies to support software reuse and knowledge sharing within the NLP research community to facilitate NLP researchers for exploring, analyzing, and developing NLP applications. We have also described the recent studies and efforts for building a library to develop NLP application, but our motivation is to design a common unified framework by applying software engineering with NLP for accelerating research and development.

We have also introduced an integrated development environment for NLP components, algorithms, and their associated technology to build a complete NLP solution. We also defined a standard communication protocol where various modules can exchange their information with each other without knowing their internal algorithms; furthermore, developer can execute the NLP solutions and analyze the results of associated modules in isolation and publish their solutions make it available for NLP research community.

References

1. M. Silberztein, INTEX: a corpus processing system, in *Int'l Conf. on Computational Linguistics*, pp. 579–583 (1994)
2. E. König, LexGram-a practical categorial grammar formalism. CoRR, vol. cmp-lg/9504014 (1995)
3. H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, N. Aswani, I. Roberts, G. Gorrell, A. Funk, A. Roberts, D. Damljanovic, T. Heitz, M.A. Greenwood, H. Saggion, J. Petrak, Y. Li, W. Peters, Text Processing with GATE (2011)
4. L. Mich, Nl-oops: from natural language to object oriented requirements using the natural language processing system lolita. Nat. Lang. Eng. **2**, 161–187 (1996)
5. O. Keszocze, M. Soeken, E. Kuksa, R. Drechsler, Lips: an IDE for model driven engineering based on natural language processing, in *2013 1st International Workshop on Natural Language Analysis in Software Engineering (NaturaLiSE)*, San Francisco, CA, USA (2013)

6. M. Saeki, H. Horai, H. Enomoto, Software development process from natural language specification, in *Proceedings of the 11th International Conference on Software Engineering*, ser. ICSE '89 (ACM, New York, NY, USA, 1989)
7. I.S. Bajwa, B. Bordbar, M.G. Lee, OCL constraints generation from natural language specification, in *International Conference on Enterprise Distributed Object Computing*, pp. 204–213 (2010)
8. M. Soeken, R. Wille, R. Drechsler, Assisted behavior driven development using natural language processing, in *International Conference on Objects, Models, Components, Patterns*, pp. 269–287 (2012)
9. M. Soeken, R. Wille, E. Kuksa, R. Drechsler, *Supporting the Formalization of Requirements Using Techniques from Natural Language Processing*, in preparation (2013)
10. M.B. Dwyer, G.S. Avrunin, J.C. Corbett, Patterns in Property Specifications for Finite-State Verification, pp. 411–420 (1999)
11. M.E.C. Hull, K. Jackson, J. Dick, Requirements Engineering, 2nd edn. (Springer, 2005)

Analyzing the Performance of Object Detection and Tracking Techniques

Suchita A. Chavan, Nandini M. Chaudhari, and Rakesh J. Ramteke

Abstract Objects are detected in computer perspective widely in many real-world applications. In case of video processing, detection and tracking of objects should be very proper and effective. Objects are detected and tracked by traditional methods such as background subtraction, optical flow, and frame differencing method. Convolution neural network, which is a deep learning-based approach, is recently adopted by many developers to identify the object. In this paper, methods of object detection are implemented, analyzed, compared, and discussed. Out of which a robust method has been suggested which satisfies the parameters of precision, recall, and accuracy, also the visualized parameters like object localization, classification, and forms a bounding box to the object are observed and analyzed. It is observed that convolution neural networks detect all relevant objects more accurately than traditional methods. CNN locates the identified object in a video frame using a bounding box that extracts the feature and trains the image for classification. Here, CNN is considered the most promising method for object detection and tracking and can be used in further study where complex work to be handled based on object detection like video inpainting or video restoration.

Keywords Object detection · Foreground mask · Convolution neural network

1 Introduction

Object detection belongs to the techniques of computer vision. It traces the object in a digital image, stored video, or CCTV footage. Object detection plays a significant role in video surveillance, defense and border security, medical image processing, astronomy [1], traffic monitoring system, counting object, image annotation system, extraction of object, and face detection system [2]. In video processing, object

S. A. Chavan (✉) · R. J. Ramteke
KBC North Maharashtra University, Jalgaon, Maharashtra, India
e-mail: suchita19.c@gmail.com

N. M. Chaudhari
KSET, DRS Kiran and Pallavi Patel Global University, Vadodara, Gujarat, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_44

tracking is a crucial task because detected objects at the first frame should be tracked properly throughout the consecutive frames of the video stream with its motion. Objects can be a person, car, bus, or any moving thing. In video inpainting systems, object detection and tracking are the prior process to perform on a sequence of frames [3]. A video is a collection of consecutive frames. This frame is divided into two parts. One is the foreground object that changes from frame to frame and can be static or moving. Another is the background object that appears in the background of an image.

Background subtraction for object detection can be performed by median filtering, Kalman filtering, and a mixture of Gaussian distributions [4], but a mixture of Gaussian background subtraction techniques is widely used and can easily apply on dynamic background scenes [5]. Optical flow method detects the object by Lucas-Kanade method [6] as well as dense optical flow method. The Lucas-Kanade method computes only sparse feature sets, where dense optical flow computes all points of frames [7]. In this paper, conventional and deep learning-based object detection methods are implemented, analyzed, and compared.

2 Methods of Object Detection and Tracking

Object detection can be possible with methods background subtraction, frame differencing, and optical flow method [8] as well as convolution neural networks.

Background Subtraction Method: In this method, video or image is segmented into foreground object and background object. Each frame is processed to subtract background and update the background, known as background modeling. The general flow of background subtraction is shown in Fig. 1. Foreground mask is generated by subtracting the background image. Foreground mask is a binary image of moving object with learning rate 0 and 1. Background subtraction is formulated with formula.

$$F_m = \begin{cases} 1, \text{ if } I_c - B_m > T \\ 0 \quad \text{Otherwise} \end{cases} \tag{1}$$

Fig. 1 Generic flow of background subtraction method

Fig. 2 Working flow of frame differencing method

where F_m is foreground mask with respective to time t, I_c is current image/frame at time t, B_m is background model at time t, and foreground mask is computed by a Gaussian mixture-based model where every pixel is a fusion of K-Gaussians [5].

Frame Differencing: In this method, pixel-wise difference between successive frames is calculated from the sequence of video to detect the objects [9]. Each colored (red, green, and blue) frame is converted into a gray image. Gaussian filters are applied to reduce blurring of the image. The previous frame is converted into gray image. Gray image and the current frame are compared. The measured variation between both images is then checked with threshold. If the value of pixel becomes bigger, then it is allotted as white, else it is assigned as black [10, 11]. Figure 2 describes the working flow of the frame differencing method.

Optical Flow: Optical flow is the movement of object in between two successive frames caused by displacement of object or camera. If first image is $I(x, y, t)$ and movement of its pixel at t, then new image will obtain $I(x + dx, y + dy, t + dt)$. Dense optical flow estimates the two-dimensional vectors for all pixels in the frame of video based on Gunner Farneback's algorithm [7]. The magnitude and direction of optical flow from a 2D channel exhibit of flow vectors that are processed for the optical flow issue. Hue and value of HSV (hue, saturation, and value) color code are used for visualization of angle of flow, i.e., direction and distance of flow, i.e., magnitude, respectively. The strength of HSV is constantly set to a limit of 255 for ideal perceivability [12].

All above methods are a traditional approach of object detection and tracking. For efficient object detection, classification and localization of objects are very important. Deep learning-based algorithms for detecting objects in all circumstances of video capturing come under recent trends of research.

Convolution Neural Network: CNN is a deep learning algorithm able to detect the objects of image/video. Image is reduced by using ConvNet, so that it becomes easy to process an image without losing features and to achieve good prediction. In a convolution neural network, input image is processed through the convolution layer, then its output passed to the pooling layer, dimension of output feature flatten to one dimensional, and fully connected network shows output as shown in Fig. 3, then fully connected layer with softmax function identifies the classes of objects with probabilistic value [13].

Fig. 3 Architecture of convolution neural network

Convolution layer extracts the feature of input image with volume size $H * W * C$ (height, width, and channel) with filter size $F_h * F_w * C$. Feature map is passed to the pooling layer, and here, the parameters count is reduced by preserving main information about convolved feature. Max pooling is used to reduce the dimension of the image and also suppress the noise. Image is flattened into a single column vector. In fully connected networks, these feature vectors combine together on which softmax function is performed to classify the objects like person, truck, bus, etc.

Object detection using convolution neural network trains the model for the presence of a specific class object. The feature of the object gets localized and classified in output layer with the class information and bounding boxes. For object detection using CNN, basic parameters need to be considered, that is, anchor size, number of class with its name, stride, padding, and filter size.

Object localization finds the value of a bounding box having probability of object detection in terms of 0 and 1, x-coordinate, y-coordinate of the object, height and width of the bounding box.

3 Quantitative Analysis

Implementation of object detection methods explained in the previous section has been performed on Python 3.7 version. After implementing the methods of object detection, its evaluation has been done, and results are retrieved in the form of precision, recall, and accuracy. As per Table 1, confusion matrix is used with the values of true positive, false positive, false negative, and true negative.

Here, the meaning of TP, FP, FN, and TN related to object detection is described below.

True Positive: An object is available in the original frame; also, it is able to detect with the method applied on frame.

False Positive: An object exists in the original frame; however, it is not able to detect or partially detected with the method applied on frame.

Table 1 Confusion matrix for prediction of objects

True positive	False positive
False negative	True negative

Table 2 Analysis of object detection methods through precision, recall, and accuracy

	Background subtraction	Frame differencing	Optical flow	CNN
Precision	0.80	1.00	0.85	0.98
Recall	1.00	0.65	0.89	1.00
Accuracy	0.80	0.65	0.77	0.98

False Negative: An object is not available in the original frame; however, it is able to detect with the method applied on frame.

True Negative: An object not exists in the original frame, and it is not able to detect with the method applied on frame.

From the output of ground truth video of all object detection methods, all the values are calculated as per the number of objects detected correctly, partially, and wrongly and then precision, recall, and accuracy are traced out as per given table. Here, precision and recall are formulated by Eqs. (2) and (3).

$$\text{Precision} = \frac{\text{Actual Object Detetcted}}{\text{Actual Object Detetcted} + \text{Object not/Partially Detetcted}} \tag{2}$$

$$\text{Recall} = \frac{\text{Actual Object Detetcted}}{\text{Actual Object Detetcted} + \text{False Detected Object}} \tag{3}$$

Precision shows only relevant detection of objects from the image, whereas recall shows all detected objects including relevant or non-relevant objects. As per Table 2, the observation shows that the precision and recall of CNN are greater than the traditional methods, which means the object detection results of CNN are more relevant.

Another observation is that the accuracy of CNN algorithm is 0.98 which is approximately close to 1. After the CNN, background subtraction is also showing better accuracy than frame differencing and optical flow method, but its output is detected in terms of binary image with 0 and 1 values. It has only one channel. Detected objects represent white portion, and the remaining image has a black portion in the background. CNN gives the most appropriate identification of objects using three channels of RGB image. Accuracy of all methods is shown in Fig. 4.

4 Qualitative Analysis

The video used for this experiment is "vtest.avi" having 795 frames [14], which is given as input to all object detection methods. Video is about many people walking on the road. Following Figs. 5a–d show the output of conventional methods and convolution neural network object detection and tracking methods.

Fig. 4 Accuracy graph of object detection methods

Fig. 5 Methods of object detection. **a** Background subtraction method. **b** Frame differencing method. **c** Optical flow method. **d** CNN-based object detection algorithm

As per the implementation of object detection and tracking methods shown in Figs. 5a, b, objects are detected according to the shape of persons with their foreground mask. Output of background subtraction shows the binary image. Objects are detected with the shadow having a foreground mask. In this method, some objects are detected partially. Occluded objects are detected, but their detection is not very

Table 3 Comparison of object detection and tracking methods

Parameter	Object detection method			
	Background subtraction	Frame differencing	Optical flow	CNN
Object detected	✓	✓	✓	✓
Classification of objects	✗	✗	✗	✓
Selection of regions of interest with bounding box	✗	✓	✗	✓
Result in RGB image with object detection	✗	✗	✗	✓

clear. Frame difference method also shows the binary image as an output. Object has a bounding box in the original frame. Objects are detected without shadow. In this method, occluded object detection does not seem to be appropriate as well as objects are detected more instead of actual total objects present in the frame. The optical flow method has detected and visualized the object through HSV color code shown in Fig. 5c. In this method, some objects are not detected as per the presence of total objects in the input frame.

Convolution neural network object detection method detected all objects correctly. The occluded objects are also detected properly in the output video. Each object has been identified by object localization, and it also classifies the object as per its domain. All classified objects are fitted into the bounding box according to the region of interest of that object with its prediction scores shown in Fig. 5d.

Table 3 shows comparison of object detection and tracking methods with the parameter of object detection, classification of objects with specific domain, bounding box formation according to the region of interest of the object, and the ground truth video having all mentioned parameters with a colored RGB image.

5 Conclusion

In security and surveillance systems, object detection and monitoring are a very significant activity. This paper explores both traditional and deep learning approaches for object detection and tracking. After implementation, the output of those techniques has been discussed.

The quantitative analysis shows precision, recall, and accuracy of all the methods, where the convolution neural network gives more accurate detection of objects compared to background subtraction, frame difference, and optical flow. The conventional approach offers only a foreground mask to distinguish objects, which is not efficient every time for applications where detailed information of objects is required.

Object localization and classification are also important parameters for the process of object identification. This goal is achieved when convolution neural networks are

used for object identification; also, bounding boxes appear in ground truth video while performing the method on video. Quantitative and qualitative analyses prove that a convolution neural network is a more promising approach for object detection and tracking. This approach can be further used in video inpainting or video restoration to build efficient and robust mechanisms.

References

1. A. Raghunandan, P. Raghav, H.V. Ravish Aradhya, Object detection algorithms for video surveillance applications, in *2018 International Conference on Communication and Signal Processing (ICCSP)*. IEEE (2018)
2. A. Vahab, et al.,Applications of object detection system. Int. Res. J. Eng. Technol. (IRJET) **6**(4), 4186–4192 (2019)
3. S.A. Chavan, N.M. Choudhari, Various approaches for video inpainting: a survey, in *2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (IEEE, 2019)
4. D.H. Parks, S.S. Fels,Evaluation of background subtraction algorithms with post-processing, in *2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance* (IEEE, 2008)
5. A. Shahbaz, J. Hariyono, K.-H. Jo, Evaluation of background subtraction algorithms for video surveillance, in *2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)* (IEEE, 2015)
6. L.Y. Siong, et al., Motion detection using Lucas Kanade algorithm and application enhancement, in *2009 International Conference on Electrical Engineering and Informatics*, vol. 2 (IEEE, 2009)
7. G. Farnebäck, Two-frame motion estimation based on polynomial expansion, in *Scandinavian Conference on Image Analysis* (Springer, Berlin, Heidelberg, 2003)
8. P.H.L. Shilpa, M.R. Sunitha, A survey on moving object detection and tracking techniques. Int. J. Eng. Comput. Sci. **5**(4) (2017)
9. G. Thapa, K. Sharma, M.K. Ghose,Moving object detection and segmentation using frame differencing and summing technique. Int. J. Comput. Appl. **102**(7), 20–25 (2014)
10. S.H. Shaikh, K. Saeed, N. Chaki, Moving object detection approaches, challenges and object tracking, in *Moving Object Detection Using Background Subtraction* (Springer, Cham, 2014), pp. 5–14
11. N. Singla, Motion detection based on frame difference method. Int. J. Inf. Comput. Technol. **4**(15), 1559–1565 (2014)
12. J. Huang, et al.,Optical flow based real-time moving object detection in unconstrained scenes. arXiv preprint arXiv:1807.04890 (2018)
13. S. Mane, S. Mangale,Moving object detection and tracking using convolutional neural networks, in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)* (IEEE, 2018)
14. vtest.avi Video. Available at https://github.com/opencv/opencv/tree/master/samples/data. Accessed on 11 Sept 2021

A Novel Area Efficient GF(2^m) Multiplier

Asfak Ali, Ram Sarkar, and Debesh Kumar Das

Abstract Galois field multiplication has received a lot of attention from researchers due to its use in encryption, channel coding, and digital signal processing. This paper proposes an area-efficient Galois field multiplier. A sequential approach is adopted here to implement the multiplier in field programmable gate arrays (FPGA). The proposed design gets a 64.58% improvement in an area with respect to 16-bit combinational implementation and 48% with respect to 16-bit sequential implementation. The proposed method also implemented using fixed polynomial, where the design gets a 73.9% improvement in an area with respect to 16-bit combinational implementation and 55% with respect to 16-bit sequential implementation. In comparison to early methods, the time delay is also decreased in most cases.

Keywords Galois field · Multiplier · Error control coding · Finite field multiplication

1 Introduction

Efficient polynomial base GF(2^m) multipliers are needed due to the continuous improvement in the field of cryptography [1], channel coding [2], and digital signal processing [3], where m is number of bits. Since the emergence of Internet of things (IoT) in 1999, a huge amount of data is generated every second, for which security must be ensured. For this purpose, the basic block of crypto technology which is nothing but GF multiplier must be area-efficient.

A large number of architectures have been proposed for the Galois Field (GF) multiplier; some of them are combinational, and some are sequential [4–9]. Tang et al. [4] proposed two methods for computation using a pre-computed lookup table.

A. Ali (✉)
ETCE Department, Jadavpur University, Kolkatta, West Bengal, India
e-mail: asfakali.etce@gmail.com

R. Sarkar · D. K. Das
CSE Department, Jadavpur University, Kolkatta, West Bengal, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_45

Imaña et al. [5] proposed a transpositional method of GF multiplier only for trinomial polynomial. Matsumoto et al. [6] introduced a linear polynomial reduction (LPR) algorithm to multiply two m-bit field elements, where elements were passed through a multiplier array which generates $(2m-1)$ bit partial prodcuts and divided by primitive polynomial of the field to evaluate GF multiplication. Iliev et al. [7] introduced a parallel programmable Galois field multiplier in which users can use different primitive polynomial as input. Chen et al. [8] proposed a combinational-based GF(2^m) multiplier. They used AND and XOR gates to develop a special cell, and this cell is used in a systolic array to compute the GF(2^m) multiplication, which can be reconfigurable in different sizes. Sangole and Ghodeswar [9] proposed a sequential-based GF(2^m) serial-parallel multiplier. They used aggregation operation using T-FF and D-FF hooked loop structure along with XOR gates to compute the GF(2^m) multiplication.

In this paper, we propose a novel GF multiplier, we consider also its implementation considering fixed polynomial. This design is proved to be better than other methods in terms of area. Besides, in terms of delay also, the proposed design performs well in most of the cases.

2 Fundamental of Galois Field Multiplication

The GF is a set of finite elements denoted as GF(q), where q represents the number of elements in the Galois field. It is also extended using irreducible polynomials of degree strictly less than m over GF(p) as GF(p^m) with respective polynomials shown in Table 1, where p is a prime number, and m is degree of the polynomial. Irreducible polynomials also known as primitive polynomial in a field GF(p^m). For binary fields, we consider a field GF(2^m). Let us consider two elements on GF(2^m) be expressed as:

$$A(x) = a_{m-1}x^{m-1} + a_{m-2}x^{m-2} + \cdots + a_1x + a_0, \quad a_i \in \mathrm{GF}(2) \tag{1}$$

$$B(x) = b_{m-1}x^{m-1} + b_{m-2}x^{m-2} + \cdots + b_1x + b_0, \quad b_i \in \mathrm{GF}(2) \tag{2}$$

where GF(2) $\in$ (0, 1).

For given primitive polynomial or field polynomial $P(x)$ over GF(2^m) , the product of two elements is normal multiplication followed by modulo P(x) and defined as:

$$S(x) = A(x)B(x) \mod P(x) \tag{3}$$

where,

$$P(x) = x^m + p_{m-1}x^{m-1} + \cdots + p_1x + p_0, \quad p_i \in \mathrm{GF}(2) \tag{4}$$

Table 1 Field polynomial

m	Polynomials
4	$1 + x + x^4$
8	$1 + x^2 + x^3 + x^4 + x^8$
15	$1 + x + x^{15}$
16	$1 + x + x^3 + x^{12} + x^{16}$

This $S(x)$ becomes a polynomial of degree $m - 1$ as follows

$$S(x) = s_{m-1}x^{m-1} + s_{m-2}x^{m-2} + \cdots + s_1 x + s_0, \quad s_i \in \text{GF}(2) \tag{5}$$

3 Proposed Architecture

In this section, algorithm and architecture of the proposed multiplier are discussed.

3.1 Galois Field Multiplier

Algorithm : 1

Input: $A(i)$, $B(j)$, $P(j+1)$ **Output:** $S(i)(i, j = 0, ..., m-1)$

Step 1: GF Multiplication
$for\ i = m - 1\ down\ to\ 0\ \ loop$
$for\ j = m - 1\ down\ to\ 0\ \ loop$
$(a)\ M_j = B(j)\ AND\ A(i)$
$(b)\ D(j)\ = (M_j\ AND\ P(j))\ XOR\ Q(j-1)\ XOR\ Q(m-1)$
//D and Q are i/p and o/p of a Delay flip-flop respectively
$endloop\ endloop$

Algorithm The multiplier takes m number of clock cycles to compute GF(2^m) multiplication using Algorithm 1. In step 1.a, it calculates bit-wise multiplication as

$$M_j = A(x_i) \cdot B(x_j) \tag{6}$$

Polynomial reduction is done in step 1.b. Firstly, AND operation between primitive polynomial and multiplication output of every iteration is calculated, then XOR between multiplication result, previous multiplication output and MSB of previous multiplication output is calculated for polynomial reduction as,

$$D(x_i) = (M_j \text{ AND } P(j)) \oplus Q(x_{i-1}) \oplus Q(X_{m-1}) \tag{7}$$

$A(x_i)$ is taken bit by bit from MSB to LSB. Multiplication result is stored in a D flip-flop. Result shifts one bit from LSB to MSB in every iteration. If multiplication output has "1" in MSB, then only it performs polynomial reduction operation in the next iteration. It happens due to the following reason. We know every primitive polynomial has MSB as "1". Hence, when multiplication output has "1" in MSB, only then the modulus operation is performed, and the result is shifted one bit. Moreover, while doing this, we may skip XOR operation of this "1" with $P(x_m)$. After m iteration, results are loaded in the output register $S(j)$ from output of D flip-flop $Q(j)$ in step 1 (b). A top-level view of the proposed model is shown in Fig. 1. Sequential Galois field multiplier structure with variable polynomial is shown in Fig. 2.

Fig. 1 Top-level view of Galois Field multiplier structure with polynomial input

Fig. 2 Sequential Galois Field multiplier structure with variable polynomial

Fig. 3 Sequential Galois field multiplier structure for $m = 4$ with fixed polynomial $x^4 + x + 1$

3.2 *Fixed Primitive Polynomial Multiplier*

Next, we consider the primitive polynomial as fixed. The serial-bit multiplier is developed using the steps shown in Algorithm 2.

Algorithm : 2
Input: $A(i), B(j), P(j+1)[Fixed]$ **Output:** $S(i)(i, j = 0, ..., m-1)$
Step 1: GF Multiplication $for\ i = m-1\ down\ to\ 0\ \ loop$
$for\ \ j = m-1\ down\ to\ 0\ \ loop$
$(a)\ M_j = B(j)\ AND\ A(i)$
$(b.i)\ if\ \ p(j) = 1\ then$
$D(j)\ \ = M_j\ XOR\ Q(j-1)\ XOR\ Q(m-1)$
//D and Q are i/p and o/p of a Delay flip-flop respectively
$(b.ii) else$
$D(j) = M_j\ XOR\ Q(j-1)$
$endif$
$endloop$

Algorithm The multiplier takes m number of clock cycles to compute GF(2^m) multiplication using Algorithm 2. This algorithm is similar to Algorithm 1, except the step 1.b is designed in two parts. In step 1.b, it calculates bit-wise addition as well as polynomial reduction with respect to $P(x)$. If $P(j)$ is "1", then only it calculates modulus as well as addition; else, it only calculates addition operation is i.e when $P(X_i) = 1$ (step 1.b.i)

$$D(x_i) = M_i \oplus Q(x_{i-1}) \oplus Q(X_{m-1}) \tag{8}$$

Fig. 4 Area comparison plot of Galois field multiplier

And, when $P(X_i) = 0$ (step 1.b.ii)

$$D(x_i) = M_i \oplus Q(x_{i-1}) \tag{9}$$

After m iteration, results are loaded in the output register $S(j)$ from output of D flip-flop $Q(j)$. For $m = 4$ considering fixed polynomial $x^4 + x + 1$, the multiplier is shown in Fig. 3.

4 Experimental Results

We compare our work with both combinational [4–8] and sequential [9] GF multipliers. For delay evaluation, the authors of [5–7] and [8] used Xilinx-virtex-V100FG256. In [4], authors used Altera-Stratix-II-EP2S130, in [8] in addition to Altera-Stratix-II-EP2S130, another device Xilinx-virtex4-VXC4VSX55 was used. However, the work [10] showed that Xilinx-virtex4-VXC4 VSX55 has close performance to Altera-Stratix-II-EP2S130. To compare our results, we use the Xilinx devices used in many previous papers. Time delay of the combinational architectures [4–8] and the proposed design is recorded in Table 2, where the first column represents the different designs, second represents the device name. The third, fourth, and fifth columns, respectively, represent the time delay of the GF(2^4), GF(2^8) and GF(2^{16}) multipliers for different designs. From this table, we can clearly see the maximum path delay in the proposed design gets improved than those of past methods.

For area calculation, the authors of [5–8] used Xilinx-virtex-V100F G256 which are indicated in Table 3, where the first column shows GF multiplier with the different bits; second, third, fourth, fifth, sixth, and seventh columns represent the area for the designs mentioned in [5–8] proposed work without and with fixed polynomial, respectively.

Table 2 Comparison of time delay (ns) with combinational architecture

Work reference	Device	GF (2^4)	GF (2^8)	GF (2^{16})
Matsumoto et al. [6]	Xilinx-Virtex V100FG256	12.314	17.516	–
Iliev et al. [7]	Xilinx-Virtex V100FG256	12.794	24.311	–
Tang et al. [4] GLUT	Altera-Stratix II EP2S130	11.106	–	–
Tang et al. [4] GPB	Altera-Stratix II EP2S130	19.133	–	–
Chen et al. [8]	Xilinx-Virtex V100FG256	12.215	18.146	–
	Xilinx-vertix4 VXC4VSX55	6.300	8.168	–
Proposed design with variable polynomial	Xilinx-Virtex V100FG256	7.418	7.760	8.462
	Xilinx-vertix4 VXC4VSX55	3.810	3.874	3.955
Proposed design with fixed polynomial	Xilinx-vertix4 V100FG256	7.184	7.418	7.418
	Xilinx-vertix4 VXC4VSX55	3.806	3.810	3.810

Table 3 Comparison of area (slice) with combinational architecture in Xilinx-virtex-V100FG256

Method	Matsumoto et al. [6]	Iliev et al. [7]	Imaña et al. [5]	Chen et al. [8]	Proposed design with variable polynomial	Proposed design With fixed polynomial
GF(2^4)	6	9	–	6	8	6
GF(2^5)	10	15	10	10	10	8
GF(2^6)	15	22	–	14	12	9
GF(2^7)	22	33	19	20	14	11
GF(2^8)	30	43	–	30	17	12
GF(2^9)	37	52	32	34	19	14
GF(2^{15})	105	153	83	90	32	24
GF(2^{16})	–	–	–	96	34	25

Table 4 Comparison of sequential architecture in Spartan3 xc3c50-5pq208

Family: Spartan3

Target device: xc3c50-5pq208

S. No.	Parameters	Conventional design	Sangole et al. [9]	Proposed design with variable polynomial	Proposed design with fixed polynomial
1	No. of bits	8	16	16	16
2	No. of clocks	No. of bits	No. of bits	No. of bits	No. of bits
3	Delay (ns)	7.31	6.280	6.878	6.318
4	No. of slice	26 out of 786	20 out of 786	18 out of 786	9 out of 786
5	No. of slice flip-flop	34 out of 1536	31 out of 1536	16 out of 1536	16 out of 1536
6	No. of 4 input LUTs	59 out of 1536	53 out of 1536	31 out of 1536	16 out of 1536
7	No. of bonded IOBs	36 out of 124	36 out of 124	50 out of 124	34 out of 124

It can be seen that in the proposed design, the area decreases for number of bits >5 in comparison to other works. For GF(2^{16}), the number of slices of the proposed design is decreased by 64.5% with respect to [8] which increases to 73.9% for fixed primitive polynomial. An illustration of area comparison of the different GF multiplier architectures is shown in Fig. 4, where the x-axis represents the number of bits, and the y-axis represents the area in the number of slices. From this, it can be clearly seen that the result of the proposed design outperforms the past results.

Comparison with sequential architecture of [9] is shown in Table 4 using Spartan3-xc3c50-5pq208. From this table, we observe that our design is always better than conventional combinational design. In comparison to [9], although time delay increases by 10% number of slices is decreased by 10%, number of flip-flop slices by 48%, number of 4 input LUTs by 42%. For fixed polynomial, with the increase of delay by only 0.6%, number of slices, number of flip-flop slices, number of 4 input LUTs are decreased by 55%, 48%, 70%, respectively. Number of bounded IOBs is also decreased.

5 Conclusion

In this paper, we have presented a new multiplier hardware architecture over the GF(2^m) field based on their primitive polynomials. Given a primitive polynomials, our design become fixed depends on that polynomial. We have used D flip-flop, XOR, and AND gates to implement GF(2^m) multiplier. We showed that area and path delays are reduced in the proposed design. As the slice and number of LUTs are reduced, the power consumption would also reduce.

Acknowledgements A part of this work is funded by DST(ICPS) CPS-Individual/2018/403(G).

References

1. N. Koblitz, Hyperelliptic cryptosystems. J. Cryptol. **1**(3), 129, 150 (1989)
2. R.E. Blahut, *Theory and Practice of Error Control Codes* (Addison Wesley, 1983)
3. R. Lidl, H. Niederreiter, *Introduction to Finite Fields and Their Applications* (Cambridge University Press, Cambridge, 1994)
4. H. Yi, S. Tang, L. Xu, *A Versatile Multi-input Multiplier over Finite Fields* (International Association for Cryptologic Research (IACR), Cryptology ePrint Archive, 2012), p. 545
5. J.L. Imaña, Reconfigurable Implementation of Bit-Parallel Multipliers over $GF(2^m)$ for Two Classes of finite Fields, in *International Conference on Field-Programmable Technology (ICFPT)*, pp. 287– 290 (2004)
6. M. Matsumoto, K. Murase, Multiplier in a galois field. U.S. Patent, No. US4918638 A (1990)
7. N. Iliev, J.E. Stine, N. Jachimiec, Parallel programmable finite field $GF(2^m)$ multipliers, in *Proceedings of IEEE Computer Society Annual Symposium on VLSI Emerging Trends (ISVLSI 04)*, pp. 299–302 (2004)
8. R.J. Chen, J.W. Fan, C.H. Liao, Reconfigurable Galois field multiplier, in *2014 International Symposium on Biometrics and Security Technologies (ISBAST)*, pp. 112–115 (2014)
9. A.M. Sangole, U. Ghodeswar, Design of bit serial parallel multiplier using finite field over $GF(2^p)$, in *2015 International Conference on Advances in Computer Engineering and Applications (ICACEA)*, IMS Engineering College, Ghaziabad, India, pp. 961–965 (2015)
10. Altera, Stratix II vs. Virtex-4 Performance Comparison. White Paper, ver. 2.0 , Altera, Inc (2006)

Virtual Testbed of Vehicular Network for Collision Simulation and Detection on SUMO and OMNeT++ Simulators

Tatini Shanmuki Krishna Ratnam and C. Lakshmikanthan

Abstract Vehicular safety technologies play a vital role in preventing or minimizing the impact of vehicle collisions to reduce life-threatening injuries and keep down vehicle collision-related casualties. One such application is connected vehicles, powered by vehicle-to-infrastructure (V2I) technology to enhance safety on road. It enables all the vehicles on a road within its range to communicate their speed, position, and heading direction to roadside unit (RSU) through cooperative awareness messages (CAM). This process needs three major operations. The first one is receiving the data from the vehicles, the second one detects the collision, and the third one communicates it with the vehicle in case of an impending collision. In this study, we developed a sophisticated algorithm to detect collisions. On detection of the impending collision, RSU sends a warning message to the concerned vehicle. This alerts the driver to take control measures like brake and speed limiting. Here, we implemented the intersection and rear-end collision scenarios using simulation of urban mobility (SUMO) traffic simulator and developed vehicular network (VANET) on network simulator OMNET++ . Veins framework combines both traffic and network simulator. Now using this computerized testbed, we can simulate the collision scenarios on the connected network and evaluate the timeline and data delivery rate with which the latter received the signal in order to take control actions like brake or halt the vehicle.

Keywords Simulation · Testbed · Collision · Detection · SUMO · OMNeT++ · VEINS · Vehicular network · V2I communication · Roadside unit

1 Introduction

According to the World Health Organization (WHO), nearly, 1.35 million people are dying every year due to road traffic crashes. These crashes cost 3% of the gross domestic product for most countries. WHO's global status reports on road safety

T. S. K. Ratnam (✉) · C. Lakshmikanthan
Department of Electronics and Communication Engineering and Department of Mechanical Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2ael19023@cb.students.amrita.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_46

emphasize the need for strong policies, smart road design, enforcement, and technological developments as preventive measures to help in reducing road incidents [12]. Collision detection is one of the most significant as well as the challenging safety application. The common causes are over speeding and exceeding safety distance. In the present time, VANET applications enable the rapid exchange of awareness and safety messages among the connected vehicles, help to reduce collisions [8].

VANET can connect vehicle-to-infrastructure (V2I), vehicle-to-vehicle (V2V), and vehicle-to-everything (V2X) [2]. So to address its complex nature, a simulation testbed is needed to provide precision and repeatability of the test cases mimicking the real-time events and study their capacity and capabilities. Simulation testbed not only reduces development and testing periods but also very economical and boosts rapid development cycle. These computerized simulations for communication networks are offered by softwares such as network simulator-1, ns-2, ns3, and OMNeT++; driving testbeds are CORSIM, PARAMICS, VISSIM, and SUMO. Here, collision scenarios are simulated on the SUMO software and collision detection algorithm designed under the supervision of vehicular network on OMNeT++. We choose SUMO and OMNeT++ simulators as they are open-source applications with robust performance [9].

2 Literature Survey

Vehicular network research started around 2000 in the scope of increasing traffic safety. Connected vehicular networks are an extensive part of building an intelligent transportation system. Author Validi's paper focused on study of VANETS in combination with traditional vision-based sensors for quick and reliable safety features [7]. In urban areas, with increased traffic flow, there is a need for smart methods for handling traffic to avoid accidents. So, author 'Senthil Kumar Mathi' proposed a model to monitor and regulate the traffic flow at intersections through RSU to alert the driver with upcoming dangers like collisions [1].

Author 'Rajeswar Reddy' developed traffic simulation in SUMO and communication network in OMNeT++, so vehicles can communicate with a RSU in a V2I fashion. He concluded that, with increase in density of traffic the communication network will be busy, which may lead to loss of data and delay in communication [3]. Author 'Sivraj P' studied Wi-Fi, DSRC, and modified DSRC communication protocols for VANETs on SUMO and NETSIM. He analyzed that among the other variants modified DSRC with IEEE 802.11 g is having good performance [2, 10].

3 Methodology

The papers mentioned in the above literature survey section focused mainly on the study of comparison on various technological solutions for vehicular networks for

enhancement of safety and comfort of the vehicle and its occupants. The work done to understand the specific road events such as different collision scenarios and the system behavior for these collisions at various speeds, traffic conditions, and road conditions are minimal. In this work, we used V2I communication through DSRC protocol to establish a vehicular network. RSU connected automatically to all the vehicles entering the vicinity of this communication range. This RSU is part of infrastructure that receives the CAM from vehicles showing their speed, position, and heading direction. We implemented a sophisticated collision detection algorithm in RSU to detect the collision chances using CAMs. And recorded the time taken for detection of collision to generate an alert or control action halt. We identified the collision scenario where the collision detection is fastest. This whole scenario is to be simulated using the below software tools [1, 6, 9].

1. Python 3.8,
2. Simulation of urban mobility simulator (SUMO),
3. Veins (integrated solution),
4. OMNeT++ (network simulator).

SUMO for Traffic Simulation

In SUMO, we imported real-world road maps and simulated traffic conditions to replicate the roads and road infrastructure like traffic lights, one-way, and two-way lanes that can accommodate denser and less dense traffic conditions, also integrated with other network simulator OMNeT++ and an integration simulator veins to provide a vehicular network for simulating connected vehicles on realistic traffic demands and incidents. To simulate traffic mobility in SUMO, we needed two inputs, they are street network, i.e., map and traffic demand, i.e., vehicle routing as shown in Fig. 1.

In SUMO, there are three ways to generate these road maps, they are possible by

(i) Using 'netgenerate' command, we compiled the subscripts that contain information of lanes, junctions, etc., to generate road network file.
(ii) 'NETEDIT' is a visual network editor; offers powerful select, edit, and highlight capabilities for user defined road and traffic designing.
(iii) Using 'netconvert' command, we converted real road maps imported from external source like OpenStreetMap database into network files (Fig. 2).

We generated the road network and added traffic flow to create an active map with moving vehicles by compiling the network file(net.xml), demand file(.route.xml or trips.xml), and poly files(.poly.xml) to generate a SUMO executable configuration file with extension.sumo.cfg. We can visualize it by running it on the SUMO-GUI interface. Figure 3 depicts the real-world roads with traffic lights [11].

Collision Simulation

Through the SUMO application, we designed road maps and generated the traffic entities. Now, we need to simulate the collision events to test the communication protocol on network simulator OMNeT++ software. SUMO software is a controlled

Fig. 1 Flowchart for simulation of road traffic

Fig. 2 Creation of network file—defines road map

Fig. 3 Real-world traffic network

environment that automatically detects any malicious events and controls the traffic to make it free of collisions. Therefore, the following operations we achieve vehicular collisions:

(i) Halt vehicle on edge (lane) and route another vehicle on the same route, cause a rear-end collision
(ii) Add variable speed signs to the edges (lanes), causing ambiguity in vehicle routing and leads to accident scenarios like intersection collisions, etc.
(iii) Add variable speeds to vehicle definition and make overtaking scenarios that lead to a collision.
(iv) Set the minimum gap parameter of SUMO, so whenever the gap between the vehicles reduced below the minimum gap, collision is detected.

After identifying the collision, SUMO provided a scope to take actions like warn, stop, and remove a vehicle to identify the collision. In Fig. 4, vehicle collision is indicated by a warning message in red color. This warning contains information about vehicles involved in a collision, minimum gap between them, lane identity, and type of collisions. We are able to succeed in simulating intersection and rear-end type collisions. In the warning message, rear-end collision type is specified as a move for stage parameter and intersection collision is identified with the same name [8].

To implement a collision-free algorithm in network simulator, we need vehicles information. As SUMO can also connect to the other software through traffic control interface (TraCI), it helped in controlling the simulation of road traffic and getting

Fig. 4 Collision events in SUMO-GUI interface

Fig. 5 SUMO and OMNeT++ interface with veins

FCD data during runtime to share it with the network simulator through a middleware framework called veins [6].

OMNeT++ for Network Simulation

OMNeT++ is an event-based network simulator, and it can prototype vehicular networks to communicate with road traffic simulated on SUMO using the veins a middleware framework. Veins will enable SUMO's traffic network models to work under the influence of OMNeT++ communication network as shown in Fig. 5. OMNeT++ offers a platform for implementing and re-using existing communication protocols, and it also supports automotive communication standards. So we can build a compatible and reliable vehicular network. [8] We developed a wireless network with dedicated short-range communication (DSRC) in V2I fashion by specifying medium access control (MAC) layer with IEEE 1609 standard and physical layer with IEEE 802.11p standard [3].

Messages are exchanged between RSU and vehicles within the maximum interference distance range of 2600 m. RSU will get the speed and position information as cooperative awareness message (CAM) from vehicles. The algorithm shown in Fig. 6 checks the chances for collision through current vehicle's position and speed as inputs, represented by vectors × 0 and v1, and takes previous inputs in B. Initialize the set C for holding the vehicles that are to be collided with the current vehicles and calculated the current vehicle's position for its changeover time. Every vehicle generated CAM that is to be received by the detector. The algorithm assigned this CAM to CAM b ∈ B (mapping messages with current vehicles). The vehicle's position over time and calculated the difference between this vehicle and the current vehicle $d(t)$. We computed $d(t)$ as scalar by taking it as squared value of distance $d(t)$, i.e., $D(t) = |d(t)|$^2. We must look for $D(T)$ minimum value over time t; now take time t where $D(t)$ is at its minimum value. If $t < 0$, vehicles are moving away from each other, it means there is no chance of collision. Else the vehicles are moving toward each other. Then, take distance d between vehicles. Check if the d value is lower than the $d$min threshold. If so, then add b to C, i.e., mapping alert to the collision vehicles set, b identifies the vehicles to send alert message, that will be involved in the collision and sends an alert before the collision happens. In our scenario, we are sending a message to halt vehicles for 50 s [5, 6, 9].

Fig. 6 Flowchart of collision detection

4 Results

Veins support parallel execution of traffic mobility on SUMO and vehicular network on OMNeT++. Red signals on vehicles indicate the threat of collision as depicted in Fig. 7 and 8.

On detection of a collision, RSU sends a message to vehicles and halts for a time duration of 50 s, at their current position to avoid accidents. Figure 9 shows the position of vehicles for time graph. Constant straight lines depict that the vehicles are in a halt state.

For any communication network, its reliability is highly dependent on data (packets) delivery time and data transfer, without any losses and delays. Here, we measured network reliability through packet delivery ratio (PDR), which tells how much percent of data packets received out of all the data packets transferred [4].

$$\text{PDR} = \frac{\text{data packets received} * 100}{\text{data packets transmitted}}$$

Fig. 7 Vehicular network on OMNeT++

Fig. 8 Mobility model on SUMO-GUI at run time

We got a maximum of 86% PDR. We calculated it by accessing the network performance data generated after complete execution. In Fig. 10, bar chart of the total received packets by individual nodes (vehicles) is plotted. In Fig. 11, bar chart of total lost packets by individual nodes (vehicles) is plotted. We took average values of all those parameters to compute PDR.

The average time delay for our vehicular network is calculated by taking the variation in total busy time and channel busy time between RSU and individual nodes (vehicles). The average result is 200 ms. Figure 12 illustrates the bar chart of a busy time for the vehicles.

Fig. 9 Vehicles position versus time plot

Fig. 10 Bar chart of total received packets

Fig. 11 Bar chart of total packets lost by vehicles

Fig. 12 Bar chart of busy time for vehicles

5 Conclusion

This paper aims to prevent or minimize the impact of vehicle collisions using vehicle-to-infrastructure fashioned VANET technology. We simulated traffic collision scenarios on traffic mobility application SUMO, communication network with collision detection algorithm on OMNeT++ network simulator, and created the VANET functionality by integrating the OMNeT++ communication network to simulated traffic environment on SUMO via veins. We achieved supervisory control over the vehicles within the network range through an exchange of messages between the roadside unit and vehicles. We used standard wireless DSRC protocol to communicate the data. We computed the reliability of the network through PDR and average time delay for an individual vehicle to the roadside unit channel. We attained 86% PDR and 200 ms for average individual time delay for vehicles. We improved PDR for this proposed VANET more than the existing methods [4]. This scenario was effective for short to medium range applications, as vehicles exceeding the network range, communication is lost. To extend the range, we can strategically place multiple RSUs on lane starting and ending based on high-probability accident zones.

References

1. Dr. M. Senthil Kumar, E. Joseph, S, D., V, M. Karthik, S, H., "Design and Implementation of message communication to control traffic flow in vehicular networks", Int. J. Eng. Adv. Technol. (IJEAT), **9** (2019)
2. M. S., S. P., "Performance comparison of communication technologies for V2X applications," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 356–362 (2020). https://doi.org/10.1109/ICCES48766.2020.9137879
3. G. Rajeswar, R. Ramanathan, An empirical study on MAC layer in IEEE 802.11p/WAVE based vehicular Ad hoc networks. Procedia Comput. Sci. **143**, 720–727 (2018)
4. A. Mukhopadhyay, V. A. Bharadwaj, "V2X based road safety improvement in blind Intersections,"*2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, (2020), pp. 964–968. https://doi.org/10.1109/ICIRCA48905.2020.9183253

5. M. Schratter, M. Hartmann, D. Watzenig, Pedestrian collision avoidance system for autonomous vehicles. SAE Int. J. of CAV **2**(4), 279–293 (2019). https://doi.org/10.4271/12-02-04-002
6. G. Avino et al., "Poster: a simulation-based testbed for vehicular collision detection," 2017 IEEE Vehicular Networking Conference (VNC), (Torino, 2017), pp. 39–40. https://doi.org/10.1109/VNC.2017.8275655
7. A. Validi, T. Ludwig, C. Olaverri-Monreal, "Analyzing the effects of V2V and ADAS-ACC penetration rates on the level of road safety in intersections: evaluating simulation platforms SUMO and scene suite," *2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, (Vienna, 2017), pp. 38–43. https://doi.org/10.1109/ICVES.2017.7991898
8. F. Lyu et al., "Towards rear-end collision avoidance: adaptive beaconing for connected vehicles," in IEEE Conference, Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2020.2966586
9. M. Malinverno, J. Mangues-Bafalluy, C.E. Casetti, C.F. Chiasserini, M. Requena-Esteso, J. Baranda, An edge-based framework for enhanced road safety of connected cars. IEEE Access **8**, 58018–58031 (2020). https://doi.org/10.1109/ACCESS.2020.2980902
10. "Documentation-Veins." [Online]. Available: https://veins.car2x.org/documentation/. Accessed 25 May 2020
11. "SUMO User Documentation -Sumo." [Online]. Available: https://sumo.dlr.de/wiki/SUMO_User_Documentation#Introduction. Accessed 06 Jun 2020
12. "WHO Road Safety Status Report" https://www.who.int/violence_injury_prevention/road_safety_status/report/en/. Accessed 2 Aug 2020

Artificial Intelligence-Based IoT Clustered Smart Vending Machine

Ravivanshikumar Sangpal, Suhas Khot, Pratiksha Vallapure, Rajkumar Mali, Rasika Kumbhar, and Aparna Hambarde

Abstract Vending machines are a vital part of normal people's everyday life in countries like Japan or the USA. India is lacking in the use of effective methodologies in vending machine space. The smart vending machine is designed to work ambiently with IoT hardware and the proprietary design architecture of the physical machine. This vending machine is designed to reduce the mechanical complications present in the current machinery and use technically advanced systems based on IoT; which will help in remotely interfacing with the machine as well as intelligently compiling the data which our artificial intelligence system interprets, giving us predictions for customers as well as the vendor. The machine will be connected to the cloud, responsible for comprehensive data collection and processing. Our system will be having a cluster of vending machines that will be interconnected to each other. The singular vendor can view the data generated by all these machines on a single console and will be able to control and monitor the aspects such as product management, machine state check, and many more from the console itself. The machine will not be having any physical interface for users, and we have designed an utterly end-to-end architecture for this system which will revolutionize the vending machine space in the industry.

Keywords Vending machine · IoT · Artificial intelligence · Cloud computing · IoT cluster · Embedded controller systems

1 Introduction

The vending machines we are considering in this monograph are the product vending machines, i.e., snacks or beverage vending machines which are used in public exposure. The spectrum of the market for these product vending machines is quite wide, and we will be focusing on the corporate usage for explaining in perspective. These vending machines are mostly imported from other countries, we consider recent

R. Sangpal (✉) · S. Khot · P. Vallapure · R. Mali · R. Kumbhar · A. Hambarde
Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, Maharashtra 411048, India
e-mail: programmer0608@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_47

market drift, and we can observe some manufacturers which are locally based in India. The machines manufactured by these manufacturers are quite effective in terms of general use cases, but they have not addressed all the issues in the existing systems. To understand our solution, let us just understand the present condition of snacks/beverage vending machines.

1.1 Introduction to Standard Machinery

Vending machines in general are quite bulky. The purpose of those machines was to reduce the need for a physical vendor at the place. If you look at some of these machines, they are very identical to the beverage fridge, and we see at local stores. This machine has three major parts or sections namely:

Controller Board: The controller board is the brain of the system which integrates every mechanical component and system to function properly. This board is having controller logic coded in and is mainly used by the number-pad present at the front of the vending machine.

Product Storage: This phase of the machine is where every product is stored for users to buy. A very generic mechanism known as spring mechanism is used as the main dispatching mechanism which pushes the product indicated by the controller. This spring mechanism has its own set of disadvantages.

Dispatchment Block: This part of the machine is where the products get relocated so that users can pick them up. Usually, the products which are pushed by the spring mechanism fall in this area. This block is open access, and users can easily pick their items as they fall.

1.2 Problems Derived from Existing Systems

The product vending machines available in the current market are designed and built considering the old and traditional design followed globally. These machines now need to be cashless considering the Digital India transition and also the COVID-19 restrictions [1]. Cashless vending machines are available in the market, but they are having some issues which need to be re-addressed: (a) Those machines do not provide adequate feasibility for the user to use them fluently. (b) It needs physical intervention for the transaction to be completed and the purchase to be made by the user. (c) The complete process is not automated, which makes the user think that directly providing cash is better than these cashless vending machines. These were only some of the issues faced by the pre-existing machines. This pre-existing system creates a lot of issues for the people using them on a regular basis.

Usually, these machines are to be seen in a commercial environment surrounded by people who work there; the small issue like you need to physically touch the machine to complete the purchase of the product is now going to have a lot of trouble considering this COVID-19 situation [2, 3]. Also, these machines work on a singular unit basis, i.e., it is a single machine having the program installed on it and provides all the data through that machine itself. This generates issues in a larger corporate environment where a lot of machines are needed in that campus itself. These machines need to be based on an interconnected server-client system which will ensure centralized data is used throughout the range of machines used in that campus.

This will ensure that the company does not need to take note of every machine placed, separately. Instead, they can just update the central database and the changes will be automatically reflected to all the machines under that network. Also getting to the point that the user needs to physically intervene in the system for the process to complete is going to be changed. Users need to touch the small display of the machine to select what product they have to buy and also what quantity is needed by them to be bought from the vending machine. This will no longer be true because our system makes this process completely automated and you can stay away from the machine by not going near or touching it to get your purchase done. We will create a secondary interface available to you so that interact with that 2nd interface to get your things done rather than touching the machine. This interfacing methodology is elaborated in subsection 2.1.

1.3 Approach to Smart Vending Machine

We have completely redesigned the overall infrastructure of these vending machines in perspective of how they are used by the user, vendor, and the client as well. We have considered all the aspects of the points stated above and designed this system which will revolutionize the vending machine market.

The smart vending machine is built with the integration of IoT. The vending machine will have controller boards that will be connected to the cloud and completely synchronized with the server. The IoT implementation is not what highlights this project but the use of IoT clustering for managing multiple machines under the same network is the breakthrough in the design. The use of clustering makes sure that all the machines in the cluster or network stay connected with each other and pass the data to the cloud. This will not affect drastically for the end-user but the vendor will be having a tremendous advantage in managing all those machines from the cloud console. The smart vending machine (SVM) also has mechanical design changes which will eliminate the bulky part of the current vending machines and make our resulting product to be modern and compact. The central architecture is having three components as depicted in Fig. 1.

The first module is the physical machine itself which is having a controller board connected to IoT sensors. The machine will have all the data generated and pass it

Fig. 1 Central architecture of smart vending machine

to the database on the cloud. The cloud acts as a mucilage of the system. The final component of the system is the mobile application. This application will be used to interact with the vending machine without touching it. The secondary interface is the mobile application that will fetch data from the cloud, and the user will update the requirement on which the machine will trigger a specific item to be dispatched. The dispatched item will then be collected by the user to complete the purchase.

SVM will use online transactions for all the purchases made by the users. The user will select the desired item on the mobile application, and the transaction needs to be completed from the mobile using any convenient method such as UPI, net banking, debit, or credit cards. The completion of the transaction will trigger the mechanical machine, i.e., once the transaction is successful, the authorization token is collected by the server, and the token authenticity is maintained in the database. Once this token is received, the trigger will be launched which will result in triggering the product dispatching mechanism.

2 The Architecture of Smart Vending Machine

As explained in the above subsection, the central architecture of the vending machine is a combination of IoT-based hardware, interactive mobile application, and cloud systems for data collection and interpretation. Let us understand the flow of the system:

The user approaches the vending machine to purchase a specific item from the machine. The machine does not have any kind of manual input, and the user does not need to touch anything. The SVM will have a QR code on the front glass of the machine. The QR code is responsible for machine identification and product information fetching. Once the QR code gets scanned, the user is redirected to our application in which he/she can view all the products present in the machine.

All the data will be visible as if you are on an e-commerce Website. You can easily select the products you intend to purchase and proceed with payment. Once the payment is successful and the token is received on our cloud, the product dispatchment triggers, and your product will be accessible to you in the bottom section of the machine. This complete process runs for all the machines present in the cluster, and each bit of data is stored on the cloud for further computation.

2.1 Software Architecture and Integration

The software end of the system is where the magic happens. The computing of all the data is done in a cloud server. The base of this system is the cloud which will be provided by services such as AWS, Azure, or Google Cloud. The base of all the software is the database which contains the data of the complete system. The schematics of the database are very crucial for other components of this system.

The second layer on top of the barebones database schema is the software stack known as the MERN stack [4]. The application will be designed using the frameworks present in the MERN stack, i.e., MongoDB for database, express, and Nodejs for logical programming in the system, and React.js + REACT Native for Web application and native mobile application designed for vendor and end-user, respectively.

Database Schematics: The database is the most important component of this software system as all the data will be originating from here and the machine-generated sensor data will be stored here. The base schema of the database includes primary data such as (a) number of machines present in the cluster, (b) products listed with details provided by the vendor, (c) the number of products present in any specific machine (inventory data), and (d) the shell structure of data going to be generated in future such assensor data, transactional data, user logs, and product update logs.

User Application: This application will be designed to display the data fetched from the server. The application will have a QR code reader to identify the target machine in front of the user. After that, the interface will be quite simple and intuitive so that users will not face any issues while using the application [5]. We have tried to eliminate the learning curve for this application as it is based on UI trends followed by top applications such as Amazon and Flipkart.

This UI will further initiate the process of the transaction. The transactions will be handled by an API in JavaScript, and the data will securely pass through the cloud server. Once the Auth token is received in the database, the trigger will be launched

which will communicate with the controller board on the machine and pass the signal to dispatch the specific product asked by the user.

Vendor Console: This is the most interesting end of the system. This console will have the same UI as the mobile application and will contain administrative privileges so that intricate data can be accessed on this console. The vendor can view all the transactions done on any specific machine and the inventory condition of the products in all present machines in the cluster. This console will view statistical data reports of any period selected depicting the growth or downfall in sales on any single product or a complete range of products.

The statistical data will also display the health of the machines which will be calculated based on the sensor data provided by the machine itself. The data will also depict the need for a specific vending machine in a certain location.

2.2 Distinctive Approach

The software architecture of the system is very peculiarly designed with consideration of nearly a year of granular research on multiple parameters of the vending machine ecosystem. The MERN stack approach referred to above is selected based on this contemporary research. The other options or approaches of the software stacking process are MEAN, LAMP, Meteor stack, and many more. The reason for choosing the MERN over the others is that the MERN stack is the most up-to-date and agile framework structure which enables users to experience a lag-free and intuitive application process. Also, the chosen software framework will ensure the feasibility, scalability, atomicity for transactions, and highly encrypted data flow structure so that the system stays robust.

Looking forward to the hardware perspective of the SVM, this design and methodology are pioneered by us. The existing systems do integrate technologies such as IoT and smart connectivity but the balance which we have created in the feature set for customers, and vendor is a phenomenal achievement. We have focused on the elimination of bulky, traditional designs and have adopted modular and modern material choices. This system will surely change the thought of vending machines in the eyes of the world.

2.3 Hardware Architecture and Integrations

The majority of the data is gained from this hardware system. Let us dive deeper into the intricacies of it. The hardware of the system is the actual vending machine. The vending machine will be fully functional but as stated before, most of the complicated mechanisms are eliminated from the traditional design because the use of those mechanisms is no longer needed and it would increase the bulk of our system

unnecessarily. The SVM will have its brain as controller boards such as Arduino, ESP32, Node MCU, etc. [6], for purposes such as main computation and control, connectivity, and wireless communication, respectively.

The other components of the system are the motors which control the tray system holding the products. The tray system is also decreased in size and materials so that the weight of the overall chassis is drastically reduced. Moving on to the circuitry of the main system, the below figure depicts the architecture of the hardware integrated into the smart vending machine (Fig. 2).

The components in this system are interconnected to each other for the completion of the designed modules. Each module is interconnected to each other in such a way that data that is generated in a single module can be passed to other components with minimum latency so that machine is fast and fluent when a transaction is triggered by the user.

Every product dispatched and transaction made will generate the data which will, in turn, help the vendor to manage the elements of this system in order of his profit. The machine's overall construction will not be done using the general aluminum chassis but with a new approach.

The complete chassis and external paneling of the vending machine will be constructed using composite materials made out of recycled plastic. This has cut down our cost by miles, and the structural integrity of the system is also not hampered

Fig. 2 Hardware integration architecture

as the internal bulk is eliminated by us previously. This complete approach leads us to a cheap to manufacture and easy-to-control and access system of smart vending machine.

3 Safety and Security

As these machines are intended to be placed in public environments, the safety and security of the products in the vending machine, as well as the vending machine itself, are very important.

3.1 Hardware Security

All the vending machines will be equipped with sensors that transmit the data to the cloud. This enables a new gateway to SVM security. We have efficiently placed several sensors in the vending machine which will ensure the safety of the vending machine and the products in them. If any sensor detects an intrusion or abnormal activity in the general workflow of the machine, an immediate alert will be sent to the vendor [7]. This will ensure that everything stays under the control of the vendor. The vendor can also set up periodic checks of the systems which ensures everything stays safe and maintains a sense of security. The below figure depicts the structure of the array of sensors in the SVM.

3.2 Software Security and Safety

The components of the software are the most secured part of this system. The application is built on REACT Native and runs on Android which ensures the safety of the application. The application before launching on the Play Store will undergo a brief security check by the Google Play Store team. The passing of our application from the Play Policy ensures that our application does not intercept any kinds of malware or viruses.

The cloud part of the data is also secured by the cloud provider which will be AWS or GCP or Azure but if we consider the flow of data through these layers, then it will be encrypted in a 256-bit encryption algorithm for which the key will be local for the individual user and the server configuration. We assure the safety and security of transactional data as there are multiple layers of security through which the data will be passed into the system.

Talking about the physical safety of end-users, we have taken into consideration the norms which are implicated by the Government of India following COVID-19

and sanitization regulations. The system will be having a UV lamp-based sanitizer in-built so that every product will stay sanitized throughout the process.

4 Artificial Intelligence Integrated with IoT

On top of all these integrations among IoT hardware and software stack, our system adds a layer of artificial intelligence which proves to be the cherry-on-top for all vendors and users.

The AI system will be running in the cloud and will predict certain parameters which are useful for the vendor [8]. As all the data collected by the sensors is stored, we have our private data set on which our AI model will predict parameters such as (a) the predictive growth and fall in terms of sales of products, (b) the list of products which are top-selling and should be stocked more, (c) the list of products which are not sold often and hence need, less amount of stock, (d) the prediction which states that which machine should be turned down for small regular intervals based on the timestamps of the transactions so that components remain healthy, and (e) list of products which are stocked in the machine for some amount of days but are about to expire which includes products based on dairy and baked goods. This data will help the vendor to efficiently manage the overall machines and products and gain profit.

Also, an AI model is designed for users as well which predicts the products you intend to buy based on the past transactions you have done. This is similar to suggested products on Amazon. This will help users to reduce the steps while buying anything as they see their regular products right on top of the others.

5 Fundamental and Supplemental Enhancements

Summarizing the overall improvements which are made in smart vending as compared to other proprietary or generic product vending machines are:

5.1 Customer-Focused Enhancement

In older vending machines, users needed to carry money in form of cash or coins, or tokenized pieces. Whereas in SVM, they do not need to carry any kind of money with them. In older machines, the users would use manual input for ordering the product and in SVM, and they do not need to touch the machine to get the product. This vending machine can be considered as a step forward in the direction of the digital era with ubiquitous computing.

5.2 Vendor-Focused Enhancement

Vendors are having the most precious advantage in this system of a console control. The one-stop portal where a vendor can access any information in terms of statistical data or conditional report of any specific machine or inventory status of all products, etc. Also, vendors can access this data remotely, if the vendor decides to leave for a weekend to a holiday, he can keep an eye on the console which will provide all the needed information. Also, the insights provided by the AI model stand helpful for getting the most out of the vending machine infrastructure in terms of profits, savings, inventory, etc. The vendors are in fine fettle if they use our smart vending machine system.

6 Circumscription of Smart Vending Machine

As the complete architecture of this machine is based on the Internet, the absence of the Internet will leave the process unrendered. The location of the establishment of SVM is suggested to include an in-house WI-FI or local server connection for annexing the machines available in the network. The workaround for this issue is that the vending machines will need a similar software with local system coded so that the vending machines will access the data when necessary. But the problem of transactional data parsing will still exist in such cases. This leads to attenuation of the SVM.

Even if the structural integrity of this work is acclaimed to be better than the general metal chassis designs, deeper investigations regarding long exposure to external environments and result analysis are needed to be done so that the structural integrity is maintained. Also, if a person is not carrying a smartphone or an Internet-connected device or the device they have does not have adequate camera resources; the user will be unable to use the machine at all. The vending machine is also having multiple parameters which are required to be tested for long periods like the collection of data affecting the server scaling, cloud stability, and sustainability in the longer run, coherence among the IoT hardware in the system is also needed to be tested for a longer duration for assuring the quality and feasibility of this work.

7 Conclusion and Future Work

We can conclude on an amicable note that the smart vending machines can be used in a wide variety of commercial and non-commercial infrastructures benefitting the larger group of people including both sides of the coin, users, and vendors. We believe that this architecture will be a revolutionary start to a new normal in our Indian markets. We also have some design improvements to be considered before

wrapping up this monograph, but we need to invest a tremendous amount of testing and quality analysis before deploying this vending machine in the open market.

This research and development are performed under Prof. Aparna Hambarde. We would like to thank her for assistance with documentation, intellectual aspects, and for comments that greatly improved the manuscript. This research is sponsored by our college, i.e., K J College of Engineering and Management Research, Pune; for which we would like to thank our principal sir Dr. Suhas S. Khot for helping us in multiple aspects of this research. We would also like to show our gratitude to them for sharing their pearls of wisdom with us during the course of this research, and we thank our reviewers for their so-called insights. We are also immensely grateful to our parents for their comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

References

1. D.T. Wiyanti, M.N. Alim, Automated vending machine with IoT infrastructure for smart factory application. J. Phys.: Conf. Series. https://doi.org/10.1088/1742-6596/1567/3/032038
2. N.Y.N. Izuddin, H.A. Majid, N.M.A. Al-fadhali, IoT monitoring system for vending machine. Progress Eng. Appl. Technol. **2**(1), 514–521 (2021). https://doi.org/10.30880/peat.2021.02.01.051
3. T. Dahanayaka, D. Chinthaka, A. R. Lokuge, I. Atthanayake, Cash sanitization machine. Euro. J. Adv. Eng. Technol., **7**(11), 24–25 (2020), ISSN: 2394–658X
4. M. Mohammadi, *How to Integrate React Native with NoSQL Database (Hosted on Azure)* (LAHTI UNIVERSITY OF APPLIED SCIENCES, Napapiiri Jukola, 2020)
5. K. Roy, A. Choudhary, J. Jayapradha: Product recommendations using data mining and machine learning algorithms. ARPN J. Eng. Appl. Sci., **12**(19) (2017), ISSN 1819–6608
6. O. H. Yahya, H. S. ALRikabi, I. A. Aljazaery, Reducing the data rate in internet of things applications by using wireless sensor network. iJOE **16**(3) (2020)
7. A. Hendra1, E. Palantei, Syafaruddin, M. S. Hadis, N. Zulkarnaim, M.F. Mansyur: wireless sensor network implementation for IoT-based environmental security monitoring. The 3rd EPI International Conference on Science and Engineering 2019 (EICSE2019). https://doi.org/10.1088/1757-899X/875/1/012093
8. P. Behera, S. Sanjay, B. Prathima, P. Singh, D. M. Sunil, Machine learning based drought prediction system using cloud and IOT. Int. J. Modern Agricult., ISSN: 2305–7246 **10**(2) (2021)

Artificial Intelligence-based Vehicle In-Cabin Occupant Detection and Classification System

M. Tamizharasan and C. Lakshmikanthan

Abstract Vehicle safety is the primary and necessary aspect of the automobile industry. An airbag is one of the passive safety systems available in an automobile. However, the airbag deployment needs to be controlled to avoid accidents due to it. In low-velocity crashes, the injury caused by airbag deployment is higher than the impact inside the vehicle. Children are more vulnerable to airbags when they sit near airbag housing without proper seatbelts or child seat arrangements. Deployment of the airbag when no occupant is sitting on a seat is unnecessary. So, it is important to detect the occupants' presence in a seat and their classes such as a child or an adult. The primary aim of the paper is to detect the occupants' presence and classify them into different classes. The occupant classes used are child and adult. In this project, we developed a technique to identify the occupants' presence and verify the data from one sensor using another. We collected the image data of the occupants using a camera in a sedan and hatchback vehicles. We analyzed the images using a deep learning algorithm. The output classified the occupants as child and adult. A load cell sensor mounted on the seat was used to measure the weight of the passenger. This data was used to confirm the occupant classification. We evaluated the model detection and classification performances with the parameters such as precision 0.95, recall 0.97, and F1-score 0.96 for image dataset, and we got 0.73 as a classification accuracy for load cell dataset. Finally, we compared both the model performances.

Keywords Detection · Classification · Deep learning · Machine learning · Bounding boxes

M. Tamizharasan (✉)
Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: tamilarasanukl@gmail.com

C. Lakshmikanthan
Department of Mechanical Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: c_lakshmik@cb.amrita.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_48

1 Introduction

Improvement in the safety of vehicle occupants is an important domain in the automotive industry. There are many safety systems available in a vehicle such as automated emergency braking (AEB), traction control system (TCS), electronic stability control (ESC), roll stability control, tire pressure monitoring system (TPMS), and adaptive cruise control (ACC). Each system has its function and purpose concerning safety. Sensors are used to measure or detect the presence of traffic lights or speed bumps, vehicles, pedestrians, and even road edge detection. Image processing algorithms are used for these applications. For example, image processing algorithm for unmarked road edge detection [1] was performed for vehicle safety. For vehicle comfort, automatic seat adjustment using face recognition [2] was performed using image processing technique. For both safety and comfort, real-time detection of speed bump and distance estimation [3] was performed using image processing technique. Typically, we classify the safety systems into two types, such as active safety system and passive safety system. An active safety system prevents the vehicle from a crash. Passive safety systems protect the occupants of a vehicle if a crash occurs. Among all, airbags are primary and passive safety systems. They typically make airbags of nylon fabric, and it is a deflated balloon that is folded inside the steering wheel in such a way that whenever there is a collision, it inflates extremely fast in about 200 miles per hour [4]. We installed a deceleration sensor in different areas, such as the front guard or motor firewall. Setting off the sensor enacts a pyrotechnic device that contains the chemical called sodium azide that touches off and delivers an enormous amount of nitrogen gas after that, it blows-up the airbag. Airbags can reduce the number of deaths or injuries that happened because of crashes. However, it can produce a range of wounds. The usually noticed wounds are slight, scraped spots especially above the chest such as neck, face, and upper arms. A portion of the primary reports was eye wounds. The distribution of airbags might separate. These wounds might be more regrettable to the patients who are wearing eyeglasses [5]. Airbags can reduce serious injuries and death rates for adult occupants who are sitting in the front seat passenger side. Occupants who are sitting too close to the airbag are more vulnerable. It recommended children who are under 13 years old to not sit in the co-driver seat [6]. Airbag deployment will strike the child's head. Besides that, it can crush the head of an infant who is sitting in the passenger seat. Suddenly stopping the vehicle causes the child to slide forward, and becoming closer to the airbag housing because of that, the deployment of an airbag affects the child's face and causes critical spine damages [7].

The primary aim of this project is to detect each occupant sitting in a vehicle and classify them into two different categories, such as adult and child. We have used the image dataset and load cell dataset. We have used YOLO version 3 object detection algorithm for detection and classification for image dataset, whereas we used support vector machine (SVM) algorithm for load cell dataset values. Finally, we studied and analyzed both model's performances.

2 Literature Survey

Author K Jost has implemented an occupant detection for airbags and the passive occupant detection system [8]. This depends on a pressure sensor fixed to the seat, bladder and connected to electronic control unit (ECU). Limitation: This system recognizes heavy objects placed on a seat and detects them as a child and deploys an airbag during a crash.

In-vehicle occupancy detection with CNN on thermal images, authors Nowruzi and Ghods have introduced a new thermal image dataset to detect and count the occupants inside the vehicle using convolutional neural networks [9]. Limitation: This method can only count the number of occupants sitting in the vehicle and not able to classify whether the occupant is an adult or a child. Since this method used thermal sensors and a thermal dataset, the cost of the project was expensive when compared to the pressure sensor.

Low-cost power in-vehicle occupant detection with mm-wave FMCW radar, authors are Alizadesh and Abedi [10]. Limitations: There was no classification part done to identify whether the occupant is a child or an adult. The cost of the radar was also expensive.

In-vehicle occupancy detection and classification using machine learning, authors Malneedi Vamsi and K. P. Soman have introduced a method to detect and classify the occupant into an adult or a child using a machine learning algorithm [11].

Limitations: This is limited only to hatchback cars and not for sedan cars, and this is not suitable when occupants are exposed to bright light conditions. They used single sensor, size of the dataset was small, and also false positive rate is also high for their machine learning algorithm.

In this project, we have used two different vehicles (sedan, hatchback) to make sure that it is not limited only to a specific type of car. We also have performed a machine learning algorithm for load cell dataset and calculated accuracy, precision, recall, and F1-score. Finally, we compared the performance of both the models.

3 Methodology

The major part of the project used an artificial intelligence technique to perform the classification of an occupant in a vehicle. Artificial intelligence is the overall technique of anything that requires some kind of decision making to take place. A part of a subset is called machine learning, and inside the machine learning, there is a subset called deep learning.

Figure 1 shows the overall idea of the project. It comprises two parts. They are based on camera and weight sensor. We have used two different artificial intelligent

Fig. 1 High-level block diagram of proposed method

models to perform classification tasks. Finally, we compared the performance of both the models.

Occupant Detection and Classification Using Image Dataset

Typically, we can find deep learning applications in image processing, videos, texts, and audio. We use the deep learning technology in this project since the project dealt with an image dataset as an input. This process involved dataset collection, image pre-processing, feeding the dataset to deep learning neural network algorithm, and finally, detection and classification task.

Occupant detection and classification system using images are shown in Fig. 1. It started with input block followed by pre-processing, detection and classification blocks, and finally, output blocks. We have captured and fed the occupant images as an input to the model. We have labeled each occupant's face in an image for class names (adult, child) before feeding into a deep learning algorithm. We have trained the images using the YOLO version 3 algorithm. Finally, we got the output bounding boxes with respective class names and a confident score on top of each image.

Image dataset collection

We have used hatchback and SUV vehicles to collect the dataset. We collected images with many combinations, such as with and without occupants. There were 16 scenarios in which occupants can sit and travel on a 5-seat vehicle. The occupant's seating arrangements are driver (D), co-passenger (CoP), passenger behind the driver (P1), passenger sitting in the middle of the rear seat (P2), and passenger sitting behind the co-passenger (P3).

We captured the images as shown in Fig. 2 from the hatchback and sedan vehicles. We have captured each image and saved it with different names at different times. We have collected datasets in different lighting conditions with variations in

Fig. 2 Sample images from the dataset

gender, variations in seat positions, and different age groups of people. The number of windows varies, which causes different lighting conditions. Further, the camera position and orientation vary, which results in a different perspective.

Data pre-processing

Once we are done with the dataset collection, the next step is to put bounding boxes on top of these images. Bounding boxes are rectangular boxes drawn on every face image. On top of each bounding box, there will be a class name associated with it. There are multiple tools available to perform bounding boxes and label the images. We have used labeling tool to do bounding boxes and labeling for each class. We have collected 1280 images and annotated using labeling tool as shown in Fig. 3.

The Algorithm Used for Image Dataset

The primary aim is to detect the location of an occupant and predict its respective class with bounding boxes for each class. To perform detection and classification, we have used the algorithm called YOLO version 3 [12] which is a part of the deep learning. YOLO applies single neural networks to an entire image, and it divides an image into grid cells. Each grid cell gives the probability values. Finally, it predicts the bounding boxes for each region. YOLO uses convolutional neural networks (CNNs).

The structure of the YOLO algorithm contains several convolutional layers and pooling layers and two fully connected layers at the end. YOLO can distinguish unique items within the picture with the corresponding area. YOLO is faster than other deep learning algorithms such as R-CNN and faster and fast R-CNN, and it applies neural networks to the entire images. It divides the pictures into boxes or

Fig. 3 Sample of an annotated image using labeling tool

regions and gives probability values for each area and picks the best one as per the probabilities. The model Darknet53.conv.74 [13] is used as a transfer learning model to train on top of our classes.

Transfer learning model: Transfer learning is a method where we reuse the learned knowledge from one task to another related task. Transfer learning is an approach to use an available model which is pre-trained for some sort of classes. Darknet 53 acts as a feature extractor in the YOLO version 3 algorithm. Darknet 53 is a convolutional neural network that has got 53 layers stacked one after the other. YOLO version 3 neural network has three types of YOLO layers at the last. We configured the last three YOLO layers for own dataset. Finally, we trained the model with the custom dataset.

Occupant Classification Performed Using Load Cell Dataset

The load cell is a cantilever beam-type load cell that is used to measure the force, tension, or pressure applied to it. Load cell contains strain gauge in a Wheatstone bridge configuration. One side of the beam is fixed, and the pressure is applied on the other side of the beam. Whenever there is a pressure applied strain gauge, it deforms and produces an electrical signal. The electrical signal produced by the load cell is directly proportional to the amount of pressure applied to it.

We have used a 200 kilogram load cell sensor for this project. The load cell setup block diagram is shown in Fig. 4.

The load cell is connected to the HX711 load cell amplifier module. HX711 is an analog to digital converter which converts the load cell value in digital, and the amplified output value is given to Arduino. The Arduino board calculates and converts the weight values in terms of grams and displays the output on the computer.

Fig. 4 Load cell setup block diagram

We positioned the load cell in between two wooden plates in such a way that any person can sit on top of it.

Collection of dataset

The load cell is connected to HX711, HX711 is connected to Arduino, and the final output is displayed on a computer. The load cell output weight values are logged in a CSV file with a timestamp. We have concentrated and placed the load cell sensor only on the front passenger seat and collected the data as shown in Fig. 5.

According to national transportation safety board, children who are below 29 kms are not supposed to sit in the front passenger seat so when the occupant weights below 29 kms were considered as a child class and weights which are above 29 kms were considered as an adult class. Overall, we collected 500 load cell values using sedan and hatchback vehicles.

Fig. 5 Load cell complete setup

The Algorithm Used for Load Cell Data

Support vector machine (SVM) [14] is a supervised learning and classification algorithm for machine learning. Supervised learning has a labeled input data for specific output classes. SVM is the model that best splits the two classes. SVM has a hyperplane, which is the widest margin that separates the two classes. So, the distance between the two classes and the lines is as far as possible. In this dataset, it splits the two classes (child and adult) in the best possible way. SVM varies from other algorithms in machine learning in a manner that it selects the decision boundary that maximizes the distance from the nearest data points of all classes. We have split the dataset into two parts. One is testing, and the other one is training. We used 30% of the dataset for testing and the remaining 70% of the dataset for training. Finally, we evaluated the performance of the model and calculated confusion matrix, precision, recall, and F1-score values.

4 Results

Image dataset results

We trained and tested the images in the dataset using the YOLO version 3 algorithm and received detections with a confident score per bounding box as an output, as shown in Fig. 6.

Fig. 6 Sample detection and classification output images

Bounding boxes tell where the occupant or a person is present in an entire image. The confident score shows how precious the model believes that the predicted bounding box contains the actual image.

Confident score can be calculated using formula:

$$C = \Pr(\text{object}) * \text{IoU},$$

IoU is a shortened form of Intersection over Union. IoU is a ratio between the area of overlap and the area of union. Whenever there are no objects in a cell, the value of a confident score will be zero. We can evaluate the performance of the model using many parameters which are IoU, precision, F1-score, and recall. The average IoU value is 82.67%. mAP is a shortened form of mean average precision. mAP value we obtained is 99.70%.

Precision is the total ratio between current predictions and total predictions, and it tells how good the model at whatever is it predicted. We obtained 0.95 as a precision value. The recall is the ratio between correct prediction and the total number of correct items in the seat. We obtained 0.97 as a recall value and 0.96 as a F1-score.

The graph shown in Fig. 7 represents the losses for each iteration. The blue curve shows the loss values, and it started from a random point and continued to move in

Fig. 7 Loss value and mAP value graph

```
 calculation mAP (mean average precision)...
 Detection layer: 16 - type = 28
 Detection layer: 23 - type = 28
164
 detections_count = 689, unique_truth_count = 561
class_id = 0, name = Adult, ap = 99.95%          (TP = 191, FP = 26)
class_id = 1, name = Child, ap = 99.45%          (TP = 353, FP = 0)

 for conf_thresh = 0.25, precision = 0.95, recall = 0.97, F1-score = 0.96
 for conf_thresh = 0.25, TP = 544, FP = 26, FN = 17, average IoU = 82.67 %

 IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
 mean average precision (mAP@0.50) = 0.997014, or 99.70 %
Total Detection Time: 6 Seconds
```

Fig. 8 Performance of image dataset model

the direction where the losses are minimal. Finally, at the end of 4000 iterations, the average loss value we obtained was 0.2876.

Class ID=0 and Class ID=1 represent adult and child class, respectively, as shown in Fig. 8.

True positive (TP) is predicted positive, and it is true. For example, the model predicted the class as an adult, and its ground truth is an adult or the model predicted the output class as a child, and its ground truth is a child. False positive (FP) predicted positive, but it is false. For example, the model predicted the class as an adult, but its ground truth is a child, or the model predicted the output class as a child, and its ground truth is an adult. False negative (FN) is predicted output class as not an adult, but it is an adult, or it predicted output class as not a child, but its ground truth is a child.

Load cell dataset results

We have shown the performance of the model in Fig. 9. We have calculated precision, recall, and F1-score for both the classes. There were 150 test sample data, which was 30% in the overall sample.

From Table 1, we can observe that in overall 150 test samples, the model predicted 78 samples and 32 samples correctly for child and adult classes, respectively. The

Fig. 9 Performance of load cell dataset model

```
[[78  0]
 [40 32]]
              precision    recall  f1-score   support

           0       0.66      1.00      0.80        78
           1       1.00      0.44      0.62        72

    accuracy                           0.73       150
   macro avg       0.83      0.72      0.71       150
weighted avg       0.82      0.73      0.71       150
```

Table 1 Confusion matrix of the model

	Predicted child (0)	Predicted adult (1)
Actual class child (0)	78	0
Actual class adult (1)	40	32

Table 2 Performance comparison

S.No	Predicted output with probability values		Actual output
	Image dataset model	Load cell dataset model	
1	Adult (1.0)	Adult 0.98)	Adult
2	Adult (1.0)	Adult (0.98)	Adult
3	Adult (1.0)	Adult (0.98)	Adult
4	Child (0.95)	Child (0.50)	Child
5	Adult (1.0)	Child (0.50)	Adult
6	Child(1.0)	Child (0.78)	Child
7	Adult (0.99)	Child (0.50)	Adult
8	Adult (1.0)	Adult (0.96)	Adult
9	Adult (1.0)	Adult (0.64)	Adult
10	Child(0.84)	Child (0.58)	Child

remaining 40 samples belong to the adult class, but the model misclassified and predicted as child classes. The overall accuracy value obtained is 0.73.

Occupant classification performance comparison

Table 2 shows the difference in the performance of both models. We can verify the output detections using image dataset of front passenger seat occupants with load cell dataset outputs. The image dataset model trained with the YOLO version 3 algorithm performs better than the model trained with the load cell dataset. In ten samples, the model trained with load cell dataset misclassified two samples. Based on the analysis, we can conclude that the image dataset is more reliable and gives better performance.

5 Conclusion

We achieved occupant detection and classification task using an artificial intelligence algorithms. We captured in-vehicle occupant images using a camera connected with the Raspberry Pi module. We performed bounding boxes for each image and classification for each bounding box using labeling application. We used sedan and hatchback vehicles to capture the occupant images and classified them into two classes (child, adult) using you only look once (YOLO) deep learning algorithm.

We displayed detection results as an output with bounding boxes for each class. Finally, we evaluated the model performance using IoU, mAP, precision, recall, and F1-score values. We performed a machine learning algorithm for load cell dataset and calculated accuracy, precision, recall, and F1-score. Finally, we compared the performance of both the models.

6 Future Scope

This project can be extended by adding more classes to the system. The classes can be an empty seat, infant, infant in infant seat, child, child in child seat, and everyday object available in the vehicle. The dataset can be trained with other algorithms, and performance can be compared. The future score of the project is to perform a sensor data fusion technique. This can be achieved by fusing both image dataset and load cell dataset values.

References

1. J. Annamalai, C. Lakshmikanthan, *An Optimized Computer Vision and Image Processing Algorithm for Unmarked Road Edge Detection: Methods and Protocols* (2019). https://doi.org/10.1007/978-981-13-3600-3_40
2. M. Vamsi, K.P. Soman, K. Guruvayurappan, Automatic seat adjustment using face recognition. Int. Conf. Inventive Comput. Technol. (ICICT) **2020**, 449–453 (2020). https://doi.org/10.1109/ICICT48043.2020.9112538
3. V.S.K.P. Varma, A. Sasidharan, K. Ramachandran, B. Nair, Real time detection of speed hump/bump and distance estimation with deep learning using GPU and ZED stereo camera. Procedia Comput. Sci. **143**, 988–997 (2018). https://doi.org/10.1016/j.procs.2018.10.335
4. C. Woodford, Airbags (2008/2020). Retrieved from https://www.explainthatstuff.com/airbags.html
5. K. Cunningham, T.D. Brown, E. Gradwell et al., Airbag associated fatal head injury: case report and review of the literature on airbag injuries. Emergency Med. J. **17**, 139–142 (2000)
6. J.A. Saveika1, C. Thorogood, Airbag-mediated pediatric atlanto-occipital dislocation. Am. J. Phys. Med. Rehabil. **85**(12), 1007–10 (2006). PMID: 17117005. [PubMed] [Read by QxMD]
7. A1. Quiñones-Hinojosa, P. Jun, G.T. Manley, M.M. Knudson, N. Gupta, Airbag deployment and improperly restrained children: a lethal combination. J. Trauma. **59**(3), 729–33 (2005). PMID: 16361920. [PubMed] [Read by QxMD]
8. K. Jost, "Delphi occupant detection for advanced airbags. The passive occupant detection system (PODS) from delphi automotive systems." Autom. Eng. Int. **108**(2000)
9. F. Nowruzi, W. El Ahmar, R. Laganiere, A. Ghods, "In-vehicle occupancy detection with convolutional networks on thermal images," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Long Beach, CA, USA, 2019), pp. 941–948
10. M. Alizadeh, H. Abedi, G. Shaker, Low-cost low- power in-vehicle occupant detection with mm-wave FMCW radar (2019)
11. M. Vamsi, K. P. Soman, "In-vehicle occupancy detection and classification using machine learning," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (2020), pp. 1–6. https://doi.org/10.1109/ICCCNT49239.2020.9225661

12. J. Redmon, A. Farhadi, An incremental improvement. arXiv (2018)
13. J. Redmon, Darknet: open-source neural networks in c. http://pjreddie.com/darknet/ (2013–2016)
14. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)

A Hybrid Approach Towards Machine Translation System for English–Hindi and Vice Versa

Shraddha Amit Kalele, Shashi Pal Singh, Prashant Chaudhary, Lenali Singh, Ajai Kumar, and Pulkit Joshi

Abstract With the rapid progress in the technology and data in the public domain, the machine translation and data science have made remarkable progress. In this paper, we discuss our specific use case of developing machine translation system for English to Hindi and Hindi to English language translation. For this system, we have used the daily proceedings of the Lok Sabha as data and developed NMT-based machine translation system on the top of already available rule-based machine translation system. Developed system has been evaluated using bilingual evaluation understudy (BLEU) as well as the human evaluation metrics using comprehensibility and fluency. In machine translation (MT), there is the trend of measuring post-editing time, and thus, we have also evaluated our system by measuring post-editing time using open-source tool.

Keywords Machine translation · Neural network · Hybrid · Parliamentary domain · Evaluation · BLEU scores · Human evaluation

S. A. Kalele (✉) · S. P. Singh · P. Chaudhary · L. Singh · A. Kumar · P. Joshi
AAI, Centre for Development of Advanced Computing, Pune, India
e-mail: shraddhak@cdac.in

S. P. Singh
e-mail: shashis@cdac.in

P. Chaudhary
e-mail: cprashant@cdac.in

L. Singh
e-mail: lenali@cdac.in

A. Kumar
e-mail: ajai@cdac.in

P. Joshi
e-mail: pulkitj@cdac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_49

1 Introduction and Background

With the advancement in machine learning technologies, the language technology has reached at an advanced level of maturity, wherein it is impacting everyone's life in the world irrespective of the language. Language technology has made the work of translation easy and less time consuming. Many machine translation tools and systems are available for use, but to best our knowledge, none of them assists for translation from English to Hindi and vice versa for parliamentary domain. In this paper, we share our experience in developing hybrid machine translation system for translating from English–Hindi and vice versa for daily proceedings in parliamentary domain.

Daily processing of the Lok Sabha produces many documents in various domains, namely papers to be laid on the table (PLOT), list of business (LOB), Bulletin Part I and Part II, synopsis, etc. Papers to be laid on table includes reports and notifications from the institutes, companies, etc. which comes under government or financed by government. Parliamentary Bulletin Part I contains a brief record of the items taken up daily like question, papers to be laid, bills introduced, etc. Parliamentary Bulletin Part II includes items of general interest to members like changes in responsibility of members, progress of Bills and matter related to security, etc. Synopsis contains a brief record of speeches of members made during the session, etc. Bulletin Part I and Synopsis are published daily, while Parliamentary Bulletin Part II is published during the session and also in intersession as well. All English data produced during the working has to be translated to Hindi, the same day for further processing. To assists the translation of daily proceedings, a hybrid system has been developed. The system uses both rule-based and NMT approach for translation.

1.1 Background

Until 2013, rule-based, statistical and hybrid approaches were used to develop machine translation systems from English to Indian languages. The system that works on pseudo-target approach is Anglabharti [1] (IIT, Kanpur, India); other was example-based approach AnglaHindi (IIT, Kanpur, India) [2] that translate frequently occurring noun and verb phrases. Anusaaraka [5, 6], a Paninian grammar-based system, was developed by Indian Institute of Technology, Kanpur. Another system MANTRA-Rajya Sabha [3] developed by CDAC Pune, India, is a machine-aided translation tool primarily designed to translate Rajya Sabha proceedings from English to Hindi language. Dave et al. [4] proposed a machine translation approach based on universal networking language (UNL). Apart from this, many SMT system has been developed for Indian language translation.

Neural machine translation is an emerging approach in machine translation, which is recently proposed [9, 10] and [11]. Unlike the traditional phrase-based translation system [12] that is made of many small sub-parts that are tuned separately, neural

machine translation attempts to build and train a single, large neural network that reads a sentence and outputs a correct translation [13].

Recently, the application of deep neural networks is making tremendous improvement in machine translation. Google came up with a very good translation system called Google neural machine translation (GNMT) system [7, 8], which used millions of samples for translation. It consists of LSTM networks. They have also implemented attention mechanism for dealing with contextual dependencies of the text and also handled the occurrence of unknown words by dividing the words into finite set of common words.

The primary problem for the machine translation research in Indian languages is unavailability of good-quality parallel sentences in abundance. For this reason, our first objective was to create a parallel corpus, which covers sentences from an untouched domain and that too of good quality and in large number. The motivation behind this was to support machine translation research by creating aligned parallel corpus. Another problem with available open-source corpus was impurities such as spelling mistakes, presence of unknown characters, data in different fonts and varied formats that makes it unusable for training or other research work. Thus, to make use of such data, it has to be pre-processed properly. The pre-processing includes various stages along with the cleaning, vetting, aligning, etc. The process of data creation is discussed in next section.

2 Data Creation

A corpus is a collection of authentic text organized into datasets. 'Authentic' in this case means text written is created by a native of the language. These datasets are important for development of any NLP application. Generally, for neural machine translation, a larger size of corpus is better. But, one cannot prioritize quantity over quality. A high-quality training data is required for good training results. This can be achieved by ensuring accuracy, completeness and timeliness. The process of data creation comprises of following steps.

2.1 Identification of Data Source

A good-quality parallel and aligned data plays a crucial role in the development of any machine learning application. For development of neural machine translation system, a huge and accurately aligned parallel corpus in Unicode is required. Many open-source datasets are available on web, but these are often problematic in terms of quality. Thus, we decided to create our own dataset for MT application. So, the obvious challenge was not only to identify the website for data creation but also handling the inconvenience and messiness of the scraping and assembling process.

For development of this system, we have chosen the Lok Sabha website (Parliament of India, Lok Sabha) for data collection. This website contains data in the form for bilingual (i.e. English and Hindi) PDF files. These files are the records of the daily proceedings of Lok Sabha, discussed in introduction and background.

2.2 Preparation of the Parallel Corpus

A parallel corpus is a corpus that contains a collection of original texts in one language and their translations into another languages. The corpus can be mono- or bi- or multidirectional in nature. English and translated Hindi files at the chosen website are kept year-wise in each subdomain separately. The data files are not aligned or parallel. The data from identified domains was downloaded.

Data transformation

After the mapping was established, the PDF data was converted to doc or txt format using optical character recognition (OCR). It was then found that the majority of the Hindi data was not in Unicode, and after doing OCR, many characters were lost or converted into a junk. In many cases, where the data was in tabular format, formatting got lost and it resulted in corrupted data. Thus, converted data required manual and semi-automatic validation after which this data was then further aligned for generation of parallel corpus.

Alignment of data

The alignment of corpus refers to storing the text of the source language and the corresponding translation and aligning the two texts at paragraph, sentence or phrase or word level. Here, we have aligned the data at sentence level. Each parliamentary subdomain has its own style of writing. In PLOT and LOB, single paragraph represents a single sentence (length up to 30 words). This may also contain embedded tables with numeric and text data as well. In synopsis domain, a single paragraph contains multiple sentences with generally no tables or other numeric data. The text up to length four was removed and was later processed and used for lexicon building. The table data was removed manually by analysing the source PDF files. The paragraphs in synopsis domain were broken into valid sentence and later were aligned with Hindi corresponding data manually. This resulted in approx. 5 lakhs of aligned parallel sentences and 20 thousand lexicon and multiword domain terms.

Data compilation

The compilation of the converted and corrected data was done with following points in the mind.

– Language Authenticity: During the process, only valid language data was taken, and other data was deleted during the validation and cleaning process.

Table 1 Showing the examples of corrupted data

Actual Data	OCR Data
सी.आर. चौधरी	सी-आर चधरी, सी-ृआर चिधरी, च धरी चरण सिंह
सं. ए-12011/1/2017 -स्था. (खंड-1), एफ सं. 10-8/2016-बीबिएंडपीए	संट ए-1201101020 17-स्था" (खंड-1), एफ सं 10-82016-बीबिएंडपीए
राष्ट्रीय कृषि विपणन संस्थान	राप्रिय कषि विपणन संस्थान
का.आ. 242 (अ); का.आ. 490(अ)	का-ृआ2 242(अ); काःआँ 490(अ)

– Data sampling: The validated, aligned and parallel data was bucketed domain-wise; the frequency of each sentence was counted, and the underlying structure types of sentences were identified using the in-house developed tool for sentence structure identification.

Challenges in Data preparation

The whole process of parallel corpus creation posed many challenges. Few of which are discussed below.

– *Similarity in the names:* The names of downloaded files were similar, and there is no identifier to establish a mapping between the English and its corresponding Hindi data. This mapping has to be established manually to create a link, by analysing content of each data file.
– *Missing files*: In certain cases, only the monolingual (English or Hindi) files are available. These single files need to be identified and marked separately. This was done by analysing the content of each file and mapping of downloaded files.
– *Corrupted data*: During the process of conversion, the Hindi characters were lost at many places or wrongly converted to some other character. The examples of which are given in Table 1.

Sentence breaks at wrong place: The English sentences were broken at wrong places leading to the mismatch of English and its corresponding Hindi data.
Stylistic variation in translation: In synopsis domain, the translators take liberty in following cases: (i) translate two English sentences together in to a single Hindi sentence; (ii) one English sentence into two corresponding Hindi sentence; (iii) translates only meaning of the English sentences without capturing the sentence structure.

All this makes the alignment of sentences much difficult and time consuming. The examples of above discussed stylistic variations are given in Table 2.

– *Repetition:* In some domains like list of business, etc., there were lot of repetition in sentences, and the sentence length and complexity of sentence were more than in other subdomain. This led to more efforts and less data in terms of unique corpus.

Table 2 Examples of stylistic variations

English	Hindi
Industrial development will push the State on the road to development automatically.	राज्य में उद्योग लगाइए विकास स्वतः होगा।
There are 35-36 sugar mills in Bihar which are lying closed today.	बिहार में 35-36 चीनी मिलें बंद हैं।
If a juvenile reports any crime to the police, it should be taken seriously and action should be taken if a police man ignores his complaint.	यदि किशोरावस्था का बालक किसी भी पुलिस स्टेशन में जाकर अपराध् की जानकारी देता है तो उसे तुरन्त ही गंभीरता से लिए जाने की जरूरत है। यदि कोई पुलिसकर्मी उनकी शिकायतों की अनदेखी करता है तो उसके खिलाफ कार्रवाई किए जाने की जरूरत है।

Addressing the Challenges

All the challenges mentioned above were handled efficiently and effectively to produce a good-quality, unique and aligned parallel data.

- File mapping issues were resolved by manually checking the content of each file and bucketing it in correct domain.
- The conversion errors were correct by semi-automatic method, wherein first few files were corrected manually which led to with the identification of the pattern or specific issues, and later, the data was corrected with semi-automatic methods and through programme.
- The breaking of English sentences during the alignment process was done both by manual efforts and sometimes with the help of available text editors.
- The process of validation and final alignment was done manually.

The data preparation took more time than expected due to the challenges mentioned above. All the efforts yield 5 lakhs of cleaned and aligned parallel sentences which was used to develop the machine translation system and can be used in many other NLP application further.

3 Development of the System

We have developed the hybrid model for machine translation from English to Hindi and vice versa. The hybrid includes the rule-based, i.e. tree adjoining grammar (TAG) and neural-based machine translation system. The TAG consists of multiple and many elementary trees, and the substitution and adjunction operations are performed on the tree to parse the source sentences to target. TAG largely handles the domain of the sentences to give better accuracy as its formalisms and dependencies defined by the elements of grammatical rules. For example, for parsing the English sentences, 'the opinion poll does not have a single question on whether the respondent received any food support from the Government or other sources (Fig. 1)'.

Neural machine translation (NMT) is method of translation source to target language based upon the artificial neural network to envisage the likelihood or the

Fig. 1 Parsed tree Source: http://nlp.stanford.edu:8080/parser/index.jsp

```
Parse [14]
(ROOT
  (S
    (NP (DT The) (NN opinion) (NN poll))
    (VP (VBZ does) (RB not)
      (VP (VB have)
        (NP
          (NP (DT a) (JJ single) (NN question))
          (PP (IN on)
            (SBAR (IN whether)
              (S
                (NP (DT the) (NN respondent))
                (VP (VBD received)
                  (NP (DT any) (NN food) (NN support))
                  (PP (IN from)
                    (NP
                      (NP (DT the) (NN Government))
                      (CC or)
                      (NP (JJ other) (NNS sources))))))))))
    (. .)))
```

probability of a given sequence of words/token, classically modelling whole input sentences in a sole integrated model. It is based upon many factors such as.

3.1 Model Building

Language model is built using the transformer deep learning model. Transformer uses the attention mechanism and weighs the input data differently according to its significance. Attention mechanism helps top detect the relationship between the words of sentence and weigh their significance separately while building model rather than taking phrase-based approach unlike its predecessor statistical-based machine translation (SMT).

3.2 Pre-processing Data for Model Building

To prepare the data for model building byte pair encoding were applied and then vocab files were generated.

3.3 Encoder–Decoder Model and Beam Search for Inference

Transformer-based encoder and decoder [16] are used to build model. Alternatively, RNN and LSTM also can be used, but transformer provides the characteristics like positional embedding, non-sequential processing and most importantly self-attention which makes the model more dynamic as compared to RNN and LSTM. Beam search [17] takes each position separately and looks for optimum result at that position and then moves to next position repeating the same for each position until it reaches the end of the sentence. At the end of the beam search completion, it returns the output with the best result possible based on the trained model.

Adam optimizer is used, and various combinations of parameters were manipulated to get the best possible result.

4 Evaluation and Results

The developed machine translation system was evaluated using both automatic and subjective evaluation methods. For evaluation, the test data a sample data of varying length covering various sentence structures from different parliamentary domains was extracted. The BLEU score for English–Hindi translation system is 21.18 and for Hindi–English translation is 41.47. Blind and open methods were used for subjective evaluation. In case of blind evaluation, the evaluators were only provided with the machine translated output, while in the open-source sentences, it was also provided along with MT output. In both open and blind evaluations, each sentence was given score for the sentence comprehensibility and fluency on minus 1 to 4 scale. Later, the overall comprehensibility and fluency were calculated using formula. Blind and open evaluation scores are given for the systems below (Table 3).

Apart from this, the post-editing time was also calculated using the open-source tool [15]. The post-editing time required for sample data was 3.34 hours for Hindi output, but for English output, it was 4.68 hours.

Table 3 Scores of subjective evaluation

Subjective evaluation				
System	Comprehensibility		Fluency	
	Blind	Open	Blind	Open
English–Hindi	60.13	52.75	46.08	42.27
Hindi–English	96.99	94.13	92.25	79.39

5 Conclusion and Future Work

In this paper, we have discussed our approach for development of machine translation system for English to Hindi and vice versa for specific domain (parliamentary). The results that we got were encouraging. Though advanced AI models are available and proven effective for other languages (English), data for Indian languages still remains a challenge. Much of Indian language data is available over the web, but none of it is readily usable due to various issues such as data formats, data encoding and unavailability of parallel corpora. Thus, a lot of manual effort is required for creation of the aligned, parallel corpus. The focus of future work is on developing tools for the pre-processing and preparation of data.

Acknowledgements We are thankful to Ministry of Electronics and Information Technology, Government of India, for supporting our work through the pilot Bahu-Bhashak project under Natural Language Translation Mission. We are also thankful to GIST Group of CDAC Pune, to carry out evaluation under CoE Bahu-Bhashak project under Natural Language Translation Mission.

References

1. R. M. K. Sinha, K. Sivaraman, A. Agrawal, R. Jain, R. Srivastava, A. Jain, ANGLABHARTI: a multilingual machine aided translation project on translation from English to Indian languages, in: Systems, Man and Cybernetics. Intelligent Systems for the 21st Century, IEEE International Conference on, 2 (IEEE, Vancouver, Canada, 1995), pp. 1609–1614
2. R. M. K. Sinha, A. Jain, AnglaHindi: An English to Hindi Machine-aided Translation System (MT Summit IX, New Orleans, USA, 2003), pp. 494–497
3. CDAC-MANTRA, Available from: https://www.cdacindia.com/html/aai/mantra.asp. Last updated 07 Aug 2017
4. S. Dave, J. Parikh, P. Bhattacharyya, Interlingua-based English–Hindi machine translation and language divergence, Mach. Transl. **16**, 251–304 (2001). https://doi.org/10.1023/A:1021902704523
5. A. Bharati, V. Chaitanya, A. P. Kulkarni, R. Sangal, Anusaaraka: machine translation in Stages, arXiv preprint cs/0306130 (2003)
6. S. Chaudhury, A. Rao, D. M Sharma: Anusaaraka: an expert system-based machine translation system, in: Natural Language Processing and Knowledge Engineering (NLP-KE), 2010 International Conference on, (IEEE, Beijing, 2010), pp. 1–6
7. M. Johnson, M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Viégas, M. Wattenberg, G. Corrado, M. Hughes, J. Dean, Google's multilingual neural machine translation system: enabling zero-shot translation, arXiv preprint arXiv:1611.04558 (2016)
8. Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. S. Corrado, M. Hughes, J. Dean, Google's neural machine translation system: bridging the gap between human and machine translation, arXiv preprint arXiv:1609.08144 (2016)
9. N. Kalchbrenner, P. Blunsom**:** Recurrent continuous translation models. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Seattle (2013), pp. 1700–1709

10. I. Sutskever, O. Vinyals, Q. V. Le, Sequence to sequence learning with neural networks. CoRR abs/1409.3215. http://arxiv.org/abs/1409.3215
11. K. Cho, B. Van Merrienboer, D. Bah-danau, Y. Bengio, On the properties of neural machine translation: encoder-decoder approaches (2014b). arXiv preprint arXiv:1409.1259
12. P. Koehn, F. J. Och, D. Marcu, Statistical phrase-based translation. In**:** Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol 1 (Association for Computational Linguistics, Stroudsburg, PA, USA, NAACL '03, 2003), pp. 48–54. https://doi.org/10.3115/1073445.1073462
13. B. Dzmitry, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate (2014a). CoRR abs/1409.0473. http://arxiv.org/abs/1409.0473
14. Stanford POS tagger**,** http://nlp.stanford.edu:8080/parser/index.jsp
15. http://www.clg.wlv.ac.uk/projects/PET/#:~:text=PET%20is%20a%20stand%2Dalone,time%20amongst%20other%20effort%20indicators
16. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is All You Need. https://arxiv.org/abs/1706.03762. Last modified 06 Dec 2017
17. Beam Search, https://en.wikipedia.org/wiki/Beam_search. Last edited 28 Oct 2021

Gender Issues and New Governance: Managing Menstrual Health and Hygiene with Emerging Multimedia Options for a Progressive Economy

Suparna Dutta and Ankita Das

Abstract A nation-building process needs to be comprehensive and progressive with strong backward and forward linkages. Hence, insight into the evolution and foresight driven endeavors to change, grow and develop further become the cornerstone for sustainable development. Women-centric issues, especially those related to health, hygiene and sanitation will play a cataclysmic role in this broad context. Here the design and implementation of an effective outreach communication for social mobilization based on substantial 'reversal of learning' is expected to lead toward more responsive, potent and sustainable governance. We are witnessing a progressive global interest in the domain and much of so in India. What is still missing is an overarching, orchestrated mechanism, framework and governance which will curate novel initiatives and ideas, filter redundancy and repetition and harness emerging multimedia options by tapping the state-of-the- art cognitive and immersive technologies to loop the population into a dialogic communication for inclusive growth. This innovative, customized, mediate communication paradigm will disseminate knowledge to spread awareness for better menstrual management practices across the nation to promote a healthy and efficient workforce to support a healthy economy.

Keywords Menstruation · Health · Hygiene · Usage and disposal of FHPs · Insinuators · Awareness · Attitudes · Taboo · Practices · Prototypes and stereotypes · Knowledge · FHPs · Outreach · Communication design · Reversal of learning · Multimedia · Cognitive and Immersive tools

1 Introduction

Menstruation is a characteristic unavoidable truth and a month-to-month event for the 1.8 billion young girls, women, transgender men and *non-binary* people of conceptive age. However, a huge number of the menstruating individuals across the world are

S. Dutta (✉) · A. Das
Birla Institute of Technology, Mesra, Ranchi (Off campus- A-7, Sector 1, Noida, India
e-mail: s.dutta@bitmesra.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_50

denied the privilege to deal with their menstrual cycle in a simple, healthy and a dignified manner. Sexual orientation imbalance, prejudicial social standards, social taboos, destitution and absence of fundamental intervention from the governing authorities frequently neglect the needs felt by the bleeding individuals. Girls from any low and middle-income countries (LMIC) including India enter puberty with huge knowledge gaps, prejudices and misconceptions about menstruation and hence remain unprepared to not only cope with it but also remain unsure of when and where to seek help from [1]. Kaur also sites gender inequality as the main propellent for this social malaise. Unequal rights usurped traditionally by men has resulted in downing the female cry for help and protest despite being the home maker, the silent workforce, and the most alert and conscious beneficiary of development initiatives. Nidhi Mallik (2013), in an exploratory study, identified that 'the awareness and knowledge about menstruation, most of the girls were dependent upon their mothers and other family members. The fact that even educated girls were not discussing these problems with their parents for reproductive health problems reflects the poor communication between them.'

'It is now common knowledge that men can easily (and have already begun to) support and influence women and girls in managing menstruation at home, schools, work, and at the community level by assuming many socio-economic roles as husbands, fathers, brothers, students, teachers, colleagues, leaders and policymakers.', [2]. Whereas, the stark ground reality shows (Namrita Rai 2019) that 'Around 63.7% women still experienced embarrassment due to the lack of facilities to dispose sanitary pads, while 72.6% found it problematic to find bins for disposal, and 80.1% claimed that their bins carrying waste were not emptied on a regular basis. And a significant majority of 84.6% women struggled to conceal the sanitary napkin while throwing it in a bin so that the male members do not get to see what they were throwing.' Whereas, Syed Uzair Mahmood (2019) shared through a survey that 'These sanitary napkins are not disposed in dustbins but are littered on the roads, free plots or burnt openly creating risk of contamination and disease transmission. In rural-urban divide, 46.89% of the urban and in 60.96% of the rural girls, burnt their soiled FHPs but 45.23% of the urban girls and 12.33% of the rural girls disposed soiled FHPs by throwing it with the routine waste. However, it was rather interesting to note that semi- skilled and semi-literate men were more comfortable in discussion issues related to contraception and pregnancy to a great degree, and slightly less on childbirth but certainly not shut toward menstruation (Singh and Singh 2014). Gragi Abhay Mudey [3] also stated in her research that 'It was found that 37.33% girls disclose only to their mothers regarding menstruation, whereas 28.67% were comfortable about disclosing to all. The various reasons given for not wanting to disclose were; irritation caused due to unnecessary remarks especially for not attending school and functions (46.26%), disgusting feeling (15.89%) and 9.35% to avoid irritation when others talk about it.'

In the Menstrual Health Management (MHM), disposal of the soiled FHPs play as critical as critical role as it is to manage the bleeding. In the absence of a state managed supportive system and ambience, women have had little choice but invent their very own ingenious methods and ways of creating crude and accessing proper

FHPs, using, changing and disposing them during their active menstruation in accordance to their individual awareness, education, economic status and social norms and ambiance. Transgender men and non-double people who discharge too regularly face more damaging segregation due to their sexual orientation and menstrual discharge in them is looked up as demonic and unnatural. This keeps them away, like menstruating females from getting access to the materials, offices, schools and services that they need urgently. Often, such exclusion and disadvantage cause serious and extensive adverse psychosomatic effects on the lives of the individuals (Aleena Baig 2019). Similar observations have come from similar researchers and scholars like [4], Naveeta Kesharwani (2010), Sushmita Malaviya (2019) and (Sumpter and Torondel 2013). Surveys and studies also reveal that most frequently the plastic used in sanitary napkins are not bio-degradable and therefore the improper, hapless littering of the soiled FHPs lead to health and ecological hazards. Stray animals scavenging this filth carry and spread the germs far and wide which few notice and often it is the unsuspecting rag pickers and street urchins and who become the first line of human contact with this virulent vector force to spread contamination and dreadful diseases. Fortunately, this dreadful, unmentored, untrained and disorderly method of metropolitan waste administration in the cities, mushrooming residential areas, markets, towns and even the villages now are progressively getting noticed by our civil society agents and getting articulated at various platforms [5].

1.1 Mediated Communication

Immersive Technology has for some time been the hottest cafe for socialization across the world. The pandemic and the consequent lockdown have only emphasized this global phenomenon provoking online chattering to the next level. The language was the god gone astray in flesh (Frantz Fanon) in the age of paper and ink and it may now be said that social media, the "hashtags" and the ever-emerging multimedia communication options have become the near equivalent of the same in the age of these virtual narratives of our times. These immersive vicarious experiences which are driven by our needs, taste and options to see and share content is propelled greatly by the evolving Cognitive and Immersive technologies. The addictive potency of such technology is arming this fast-spreading medium with the enormous power of advocacy and influence. Consequently, such mediated communication is creating the opportunity to create sharper subgroups based on interest and deliberate content. Hence, much of the critical ideological and socio-emotional matter like gendered content is shared more freely among the members making it the indispensable point of conjunction, a platform and the coming together for consumers who are keen to chat, share, mingle and do business together and believe spontaneously and consciously. This is where, the medium is gaining the power to influence and has become a tool for both advantage and ace in contentious and sensitive social communication drives and initiatives. In fact, smart governance should ideally take advantage of these online community clubs to disseminate critical social messages for greater attention

and retention by the audience. But for sharper impact, content creation for such outreach communication should ideally be based on participatory communication. Dialog and discussion are the two holy wheels of such targeted participation and such participation is primarily possible with help of innovative mediated communication.

1.2 Objective

This study is based on a field survey conducted in the year 2019–2020. The scope of the survey was limited to the metropolitan city of Bengaluru and its suburbs in the south of India. The participatory action research used qualitative research methodology and tools to explore and comprehend the worries and propensities of menstruating individuals and people connected with them in this, demystify myths on menstruation, augment self-adequacy and esteem in women, assist vulnerable menstruating individuals with capacity building aptitude and skills to overcome obstacles to their wellness, freedom, privacy, sexual orientation-based violence during this vulnerable period, menstruation induced discomfort, promote proper disposal of soiled menstrual waste and design a communication framework that seamlessly integrate into the e-governance ecosystem of urban, semi-urban and rural areas in India keeping the plural hyper-local context of our nation.

1.3 Background

Menstruation from menarche to menopause associates with new changes and vulnerabilities that needs to be collectively dealt with at all levels. Menstruating individuals require support and some more so through their menstruating life till menopause irrespective of class and education. But especially vulnerable in this are the base-of the pyramid population, the low middle class and the destitute. It is well understood today that menstrual well-being and hygiene interventions can augment gender transformative skills, performance abilities and professional competence. It also well understood that there is a need to design a dynamic ecosystem of trusteeship for this where all the stakeholders would align to create an overarching inclusive framework based on a participatory model that initiates sustainable dialog to not only ensure 'reversal of learning' but also the integration of this dynamic, dialog-based ecosystem of progressively informed trustees into the emerging e-governance ambit in India [6]. Though issues related to menstrual health and hygiene has to become a prominent and integral part of all governance to ensure the development to be glitch free and the inclusion has to be universal [7] yet, unfortunately, even the survey conducted reiterates that menstrual health, hygiene and awareness still remains a causality in our country especially in the marginalized, deprived and the lower middle class (LMC) segments of the population. The situation is not very encouraging and healthy in the middle class population as well specially in the tier 2 and tier 3 cities. However, the

upper middle class and affluent families of the metropolitan cities are much more fortunate in the sense that they have overcome much of the taboos and ritualistic Indian traditional approach toward menstruation. So prompting and sensitizing a responsive digital governance for a sound and comprehensive inclusive development becomes the responsibility of all.

1.4 Current Scenario

It is the Community-based menstrual health tips which help to battle and give succor to the menstruating individuals. So, a network of competent stakeholders from school teachers, village midwives, primary healthcare centers both in urban and rural areas, government dispensaries, dedicated district (preferably block level) administrative machinery and so on need to come out as the talkative interlocuters. But here, it is the communication paradigm used that is going to make all the difference [8]. The communication principle of 'seeing is better than hearing, but it is best to experience…' is what needs to be pegged on in this context of targeted outreach. The governance and its partners need to move out of printed flyers, and garrulity and resort to the magic of cost-effective multimedia communication. Borne by ICT, the emerging immersive and cognitive technologies like 3-D Animation, Augmented, Virtual and Mixed Reality and the virtual games are the new magic wands of effective communication. They customize and they personalize the content and they also can cater to the individual and the hyper-local through the social media. And the potency of the immersive technology can attract and impinge on the minds of the end user through the incessant use of content packed in the infotainment mode. The new governance across the country must harness this and developing content will not be difficult and there is a steady progress of academic institutions and Institutions of Higher Educations (IHE) who are showing interest in offering degree diploma program in Animation and Multimedia Studies. Working in tandem with such IHEs, the entertainment or the Animation Industry, the government can easily design such modules for their outreach communication.

1.5 Reversal of Learning

Governance and policy making share a hand and glove relationship. Unfortunately, policy decisions are made from the federal offices which are frequently cut-off from the ground reality. Of course, this does not indicate lack of competence and skill in the official functionaries who are mostly selected through the rigors of a highly competitive and complex selection procedure. But what mostly go lacking is the familiarity with the hyper-local context, traditions and nuances. Fortunately, the ongoing pandemic has brought this hyper-local to the forefront and hence addressing it will no longer remain a difficult domain for anyone. This familiarity is expected to

bring the policy makers will come in close contact with the local ethnicity, culture and typical traditions, knowledge, worldview and customs developed thereof and push them toward a deeper comprehension and develop a more holistic perception of the factors that promote most of such social malaise. It is indeed disappointing that even after 7 decades of independence and self-governance, we in India still need to frame popular gender and family-planning slogans where as in contrast, the governance has been robust and progressive enough to make India become a signatory nation to the United Nations Sustainable Development Goals that idealize a gender bias free and neutral global society at all levels and locations. So, doubting the intentions of governments is not always fair, nor is it always correct to distrust the wisdom of the government. What normally remains missing in such socio-economic evaluation context is the awareness and acceptance of the hyper-local reality and its dynamics by the educated policymakers. It is a kind of 'traditional disconnect' and a 'dissociation' that stems from ignorance often stemming from the conceit born from the modern university driven education of the policymakers. Once this 'connect' between the hyper-local context and the governance happens, the latter is expected understand, learn and accept both the subtle local social matrix that change every 100 km with the dialect in India, the local wisdom as well as the local stereotypes with all its idiosyncrasies. This learning and acceptance is of critical value for the lawmakers because it is here where 'local resistance' and thereafter the popular disconnect with laws, policies and the government itself begins. The sloth and power of local tradition in which most of such eccentricities are meshed in and borne by should not be ignored while framing our rules, policies and political agendas. Behavior change communication relies on such 'typical knowledge' of tradition, culture and worldview. So, when policies are framed, it is done with reference to this knowledge of the local order. But most importantly, it is when communication modules are designed that this knowledge of the local came handy in customization of the content which is critical for behavior change communication and thereby inclusive growth.

2 Data Analysis Derived from Survey Par Conducted in Bangaluru in 2019–20

2.1 Menstrual Health

There were significant number of respondents who knew of and suffered some of the vaginal and health diseases related and caused by improper menstrual hygiene, but they still did not have specific knowledge of those dreadful diseases like the Urinary Tract Infections (UTI) and Reproductive Tract Infection (RPI). However, almost none connected such recurring infections with FHPs and their hygienic use and disposal. The survey also found individuals who themselves or their friends who suffered from 'yeast infections' and all they understood of it was that an odor filled discharge was to be expected. It was less than 10% respondents who realized the

connection between appropriate menstrual practices sound vaginal and menstrual health, with recurring problems like irritation, fever or some other serious health issues that even lead to chronic conditions like cancer. The survey did not come across individuals who mentioned cervical cancer, hepatitis B and the like on their own while talking about menstruation and MHM.

2.2 Knowledge and Awareness

An overwhelming majority of the female respondents shared that they first got to know of menstruation from their mothers and from information and experiences shared by their friends and other classmates. Less than 20% of the female respondents admitted to learnings from menstrual health camps in school or biology subject they studied as part of their curriculum. These were also young girls who spoke about their proactive schools and teachers who talked to them and shared relevant information. Some also talked of having met some educated, emancipated relatives who introduced them to novel FHPs and talked to them of hygiene and proper disposal of FHPs. However, none of the respondents knew about insinuators and were quite startled and impressed by the smokeless mechanism that was capable of destroying regular domestic waste and the FHPs easing their challenge of waste disposal. Though less than 15% of the female respondents correlated a clean, green pollution environment with MHM, an overwhelming 63% of them agreed that schools, residential areas, transport stations and hubs, markets, etc. need insinuators.

2.3 Social Taboo

The survey revealed a varied and probably a hint of a progressive change. Mostly, menstruation was an avoidable topic for discussion as it was looked upon as natural and regularly occurring and hence to be taken in stride. Heavy bleeding and related discomfort are still a matter of individual concern. Less than 20% of the sisters and mothers were overly concerned about their young daughters bleeding heavily. Majority did not have or looked for any specific information from anyone or source on heavy bleeding and its cure nor felt the need to do so. Over 76% of the female respondents did not feel local administration had anything to do with heavy bleeding and FHPs, but they were all surprised with the workings of an insinuators. Regarding talking about menstruation. 40 percent of the female respondents spoke to each other, their mothers and neighbors and purchased and shared FHPs. More than 20 % of the male respondents admitted to purchasing the FHPs for their wives and even their sisters and other female relatives at home. These were also male respondents who were aware of the pain, fatigue and discomfort caused by menstruation and appreciated insinuators. 13% of the male respondents were open to health camps and support from local administrations. But more than 87% of all the women still

would avoid places of worship and treat menstruation as ritually polluting. There were many local customs of prohibition of both consuming and serving of food during this period and men felt no compunction to cooking to take such taboos forward. Not attending school was another overwhelming trend among the lower income families. But females, generally went back to work the moment the heavy bleeding stopped irrespective of exhaustion, weakness or irritation which never was to be considered.

2.4 Disposal

The survey revealed a universal exposure and knowledge of improper and ugly disposal of soiled FHPs. Littering was again a universal annoyance which was prosaically accepted by all. The resulting stink and animal menace though a source of embarrassment was still not overtly connected with health hazards. It is still very common to through and litter unwrapped soiled FHPs from widows, moving trains, and in public places with little guilt and responsibility. The demand for administrative support for such waste management is still less than 18% and people are still unable to differentiate between dry and wet waste. Significantly, there is still no popular protest and opinion against callous disposal of city garbage where used FHPs are crudely gathered, visible transported with other waste material. Such menstrual waste unsurprisingly is commonly found in public unisex toilets, college washrooms and dumped trough ducts into activity and public space in schools, colleges and work areas. Yet, no warning or public instructions were seen for the proper disposal of FHPs from any authority.

3 Conclusion

India as an economic super power suffers a few paradoxes which needs immediate attention. What is of advantage here is the progressive unison in accepting gender challenges at all levels and the realization that despite having reached the orbit of Mars in a maiden venture, we have still left our womenfolk in the more distant Venus. This is especially true of our Menstrual Health Management (MHM) practices. Beyond debate, it is individuals from the middle and lower social rung that suffer most and are not only affected with low self-esteem and suffer socially debilitating consequences for this required normal biological condition but are thence affecting the nation's progress to level next by spreading diseases and remaining unproductive at work and school. A simple solution to this is the governance taking recourse to multimedia communication where an overarching communication framework is created that would induce spontaneous participation of the people into open and free conversation with experts from the field. Doctors, counselors, administrators should create their specific knowledge driven content in infotainment packages and make them available to the people to answer their questions and alley their fears

and doubts in effective dialogic communication. But what needs to be seen is that both privacy, security and authenticity should be kept at the premium. Keeping in mind the open architecture of the immersive multimedia technology, security of the content uploading and the end users should be constantly kept in mind. At the same time, the potency of the emerging animation and multimedia communication should be harnessed for greater outreach, sharper reception, deeper learning and content creation and content delivery. The appeal, the playful, recreative ambience and infotainment potential that such cognitive and immersive technologies are capable of delivering should be used to create participation of the end users. What is to kept in mind is the dialogic format of communication where questions, doubts and fears could be expressed and allayed and addressed by the experts as an integral part of effective governance that is itself aware of not only the problems but the cause and context of the problem and its solution as well without bias and arrogance and insularity of education and power.

(The Authors did not include the Menstruating Third Gender within the Scope of this Study).

References

1 R. Kaur, K. Kaur, R. Kaur, Menstrual Hygiene, management, and waste disposal: practices and challenges faced by girls/women of developing countries. J. Environ. Public Health **2018**, 1–9 (2018). https://doi.org/10.1155/2018/1730964

2 A. Dasgupta, M. Sarkar, Menstrual hygiene: how hygienic is the adolescent girl? Indian J. Commu. Med.: Official Publ. Indian Assoc. Preventive Soc. Med. **33**(2), 77–80 (2008). https://doi.org/10.4103/0970-0218.40872

3 M.A. Bhausaheb, N. Kesharwani, G.A. Mudey, R.C. Goyal, A Cross- sectional Study on Awareness Regarding Safe and Hygienic Practices amongst School Going Adolescent Girls in Rural Area of Wardha District, India (2010)

4 S. Puri, S. Kapoor, Taboos and Myths associated with women health among rural and urban adolescent girls in Punjab. Indian J. Community Med., **31**, 168–70 (2006). [Google Scholar]

5 S. Ahn, K. Cho, Personal hygiene practices related to genito-urinary tract and menstrual hygiene management in female adolescents. Korean J. Women Health Nurs. **20**(3), 215 (2014). https://doi.org/10.4069/kjwhn.2014.20.3.215

6 N. Mehta, S. Dutta, Multimedia for effective communication. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (2016). https://doi.org/10.1109/scopes.2016.7955510

7 A. Kumar, K. Srivastava, Cultural and social practices regarding menstruation among adolescent girls. Soc Work Public Health., **26**, 594–604 (2011). [PubMed] [Google Scholar]

8 S. Dutta, N. Mehta, R. Pratik, Humane Digital Route to Customize Communication for Sustainable Development. INTED2013 Proceedings, (2013), pp. 3090–3096

An IoT-Based Smart Parking System for Smart Cities

Venkatesh Mane, Ashwin R. Kubasadgoudar, Raghavendra Shet, and Nalini C. Iyer

Abstract There has been an exponential growth in the vehicle population in recent times, adding to the metropolitan traffic. This increase in the number of vehicles has led to the problem of inadequate parking spaces resulting in traffic congestion. To address this challenge, a smart parking system is proposed in this paper, which makes use of TIME RESOURCE SHARING to effectively utilize the parking spaces based on peak demand time and enables prior identification and reservation of parking space with the help of unique identification. The system periodically updates the parking status. The interaction of the driver and the owner of the parking space takes place through the application connected through the cloud. The related work in this area has also been referred. The proposed system has the potential to transform the current parking method and alleviate the traffic congestion caused by insufficient parking space.

Keywords Time-shared parking · Unique code · Traffic congestion · Parking space · Arduino

1 Introduction

Smart city development is one of the major areas that need to be addressed. Smart city development refers to the efficient utilization of the available resources with the active interaction with the community leading to the improvement in the quality of life and in the overall development of the nation. In present times, with the ever

V. Mane (✉) · A. R. Kubasadgoudar · R. Shet · N. C. Iyer
KLE Technological University Hubballi, Hubballi, India
e-mail: mane@kletech.ac.in

A. R. Kubasadgoudar
e-mail: ashwinrk@kletech.ac.in

R. Shet
e-mail: raghu@kletech.ac.in

N. C. Iyer
e-mail: nalini.c@kletech.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_51

increase in vehicle population for the ease of transportation, there arises a serious problem of inadequate parking spaces. The available parking spaces prove to be insufficient to meet the growing demands. Searching for a vacant parking space becomes a tedious, time-consuming task for the drivers leading to traffic congestion and causing hindrance to the flow of traffic [1]. It also contributes to increasing the extent of pollution and increasing the risk of accidents. Hence, this issue of inadequate parking space should be addressed to achieve smart city development.

'To address the issue, many new parking lots are being built. But this means new constructions and utilization of extra space, which is again a big hassle [2]. To reduce the traffic congestion caused by inadequate parking spaces, many smart parking systems have been proposed. The conventional method adopted is that the driver simply searches for a parking space. They continue to do so until he finds a vacant space [3]. Some systems check for the availability of parking space and broadcast it to the drivers [4]. In this, there is no reservation facility, and sometimes instead of reducing traffic congestion, they worsen it. Therefore there exists a greater demand to address this issue with an effective strategy.

Our contribution to meet the demands of parking space efficiently is as follows. A smart parking system with time-shared facility is proposed, which adds a new dimension to the existing parking space by introducing the concept of time sharing, which makes use of city resources efficiently, which excludes new constructions and major investments for parking spaces. Shared parking refers to the parking facility that is used jointly by several sectors, maybe offices or residential flats, and designed to take advantage of differences in peak parking demand times, which can vary by day, week, and season [5] For instance, apartment parking space remains vacant during the weekdays when the owner goes for work and the parking space allocated to offices remain vacant during weekends [8, 9].

The proposed system has a unique device at the parking space for the vehicle identification and monitoring of parking status. The system also caters to the collection of parking fees based on the time duration and request from the customer [6, 7]. The interaction between the customer and the owner of the parking space takes place with the help of an application wherein all the information is stored in the cloud.

The organization of the rest of the paper is as follows. Section 2 describes the details of the proposed methodology. Section 3 describes the scope of implementation and the conclusion of the article in Sect. 5.

2 Proposed Design

The proposed system aims at providing a TIME-SHARED RESOURCE facility that makes use of existing parking spaces without having to build new constructions or new parking facilities to address the problem of inadequate parking spaces. Timed Shared Resource Parking basically aims at providing parking spaces during peak hours of the day. Apartment parking space remains vacant during the weekdays when the owner goes for work, and the parking space allocated to offices remains

Fig. 1 Overview of the system

vacant during weekends. Thus there exists a need to identify and utilize the spaces available in an efficient manner. Therefore an integrated system providing a solution on a time-shared basis is illustrated in Fig.1, along with the operational details.

The system mainly focuses on the effective utilization of parking space through the interaction between the user and the owner, connected through an application via a unique device.

Every parking space is connected to the Internet through the proposed application. This enables the communication between the user and the owner of the parking space. This information of vacancy or occupancy is sent to the user through the application. The user then reserves a parking space from the available spaces. The owner determines the availability and the time duration for which the parking space is available for the user to be utilized. Then the same data/information is updated in the database.

Each parking lot is deployed with a unique ID for the validation/confirmation of the user's identity and booking. The system acts as a centralized body that dynamically allocates parking spaces on a time-shared basis. When the user reaches the parking spot, the user is identified by the unique ID given at the time of signing up. Since the information of each parking lot is stored in the database of the system, the system is able to synchronize all requests from the user. Therefore, whenever a user books a parking space, that space is reserved and is unavailable for other users.

The proposed system is a combination of hardware and software modules. Details of each are detailed in the further sections.

3 System Architecture-Hardware

The functional details of the system are shown in Fig. 2.

Fig. 2 Functional block

The system basically comprises a unique identification device and an application. The interaction between these unique devices takes place through the application.

Each parking space is deployed with a hardware module, and every vehicle will have a unique identity card, which is basically an RFID. Each such hardware module comprises electronics and embedded circuits that provide unique identification determination based on the parking slot booking. The module consists of a controller which does all the computations and processing of the data. The controller is able to access the Internet through the help of a Wi-Fi module. This module enables the hosting of the application. It also consists of a real-time clock (RTC) that helps to keep track of time duration.

4 System Architecture-Software

The flow chart for the application running between the two dedicated devices is shown in Fig.3.

The software module consists of an application designed on an android platform. The user has first to create an account to be able to use the services. Once the account is created, the user can log in with his username and password. Users can then select the required destination and check the availability of parking spaces in the vicinity. If free spaces are available, then the user can reserve the parking space. One user is allowed to reserve only one space. For booking, the user has to enter his details, including his unique identification number. Once the Parking space is reserved, the database is automatically updated from vacant to occupied hence no other user is

Fig. 3 Software architecture

allowed to park in that space. Prior notifications are sent to the user's phone to indicate that reservation time is about to expire.

The owner has to sign-up for the first time and then provide the required information. The owner is also given a unique identification number. When the owner logs in to his account, he has an option to rent his parking space for a particular time interval. This information is used by the user to enable time-shared parking.

5 Implementation

The proposed system mainly focuses on assigning parking spaces on a time-shared basis. Availability of parking space before reaching the destination is made available to the user with a booking facility. As a result, this reduces the "multiple-car-chase-single-slot" phenomenon; hence the system reduces the traffic congestion, thereby leading to the decrease in the extent of air pollution.

Other salient features of the module include a number plate tracking system, Bluetooth connectivity, tracking facility, unique identification, which helps in achieving the aim proposed. The software module consists of the various locations available for parking in the vicinity, which can be monitored using GPS modules. These free spaces can be identified, booked, and utilized in a feasible way. The integration of both of these modules paves the way for the achievement of a smart parking system. Even vehicle identification, in case of theft, is also addressed through a unique ID given to each vehicle. In case of a change in a number plate, the unique ID remains the same and hence addresses the issue of identification of the stolen vehicle.

As the system involves user and owner communication through the application, we may embed the Google navigation system onto it to better direct the users to the parking space. The unique identification number assigned to each user can be made to contain all the details of the user, such as his license, registration number, and address, so that the user can be identified through a unique ID. Hence, paving the way to paperless identification in a smarter way.

Details of the application developed are described in further sections.

6 Application Details

The user and the owner are connected through the application. This application helps to know the status of parking spaces.

Figure 4 shows the home screen, which has a sign-up and login option. The sign-up page asks for the details.

After signing up, a unique ID is assigned. After logging in, an options page is displayed, as shown in Fig. 5.

The owner has an option to rent the space, and the customer has an opportunity to book a parking space, as shown in Fig. 6.

Fig. 4 Application home and sign-up page

Fig. 5 Unique ID and user options

Fig. 6 Booking and renting page

7 Discussion on Results

The proposed method offers the advantages of optimized parking space usage, improves in the efficiency of parking operations, and helps traffic in the city to flow more freely with the existing resources.

Our system works with a wide range of operators ranging from private retail operators to the government or private sector-owned large-scale parking systems.

The owner of the parking space will be benefitted by getting maximum capacity from the available space.

Space management was easier with an instant real-time reporting system. The drivers were guided quickly and easily to the nearest available parking space.

8 Conclusion

To address the need of inadequate parking space, the proposed system makes use of TIME SHARING parking facility, which enables effective utilization of city resources along with active interaction of the community. This reduces the time spent searching for parking spaces, thereby reducing the traffic congestion and paving the way to the development of the smart city.

References

1. R. Shet, N.C. Iyer, Y. Jeppu, Fault tolerant control system for autonomous vehicle: a survey. J. Adv. Res. Dyn. Control Syst. **12**(8), 813–830 (2020)
2. P.S. Pillai, R. Shet, N.C. Iyer, S. Punagin, Modeling and simulation of an automotive RADAR. In *Information and Communication Technology for Competitive Strategies (ICTCS 2020)* ed. by. A. Joshi, M. Mahmud, R.G. Ragel, N.V. Thakur. Lecture Notes in Networks and Systems, vol. 191. (Springer, Singapore, 2022)
3. N.C. Iyer et al., Virtual Simulation and Testing Platform for Self-Driving Cars. In: *ICT Analysis and Applications. Lecture Notes in Networks and Systems*, ed. by S. Fong, N. Dey, A. Joshi, vol 154 (Springer, Singapore, 2021)
4. R.M. Shet, N.C. Iyer, P.C. Nissimagoudar, A. Kulkarni, J. Abhiram, S.K. Amarnath, Motion Control and Sensor Fault Diagnostic Systems for Autonomous Electric Vehicle. In: *ICT Analysis and Applications. Lecture Notes in Networks and Systems*, ed by S. Fong, N. Dey, A.Joshi, vol. 154 (Springer, Singapore, 2021)
5. P. Shastri, A. Chauhan http://timesofindia.indiatimes.com/city/ahmedabad/Ahmedabad-lets-get-smart-link-fee-to-parking-demand/articleshow/50885898.cms
6. A. Roy. http://timesofindia.indiatimes.com/city/nagpur/New-bldgs-will-require-double-the-parking-space/articleshow/51001996.cms
7. P. White, *No Vacancy: Park Slopes Parking Problem and How to Fix It*, http://www.transalt.org/newsroom/releases/126
8. http://fresno.ts.odu.edu/newitsd/ITS_Serv_Tech/park_sys_tech/parking_systems_tech_report_print.htm
9. Parking Management for Smart Growth by Richard W. Willson

Agriculture Stakeholder's Information Need—The Survey in Gujarat, India

Axita Shah and Jyoti Pareek

Abstract Agriculture plays an important role in Indian economy. In recent era applications and technology usage is drastically increased. To fulfill the demand in Agriculture, it is important to know the information need of the farmer. Farmer's Information Need can be satisfied by the varied human stakeholders in Agriculture or by the technology use. We have prepared Agriculture Questionnaire to know the farmer's information need. Our Agriculture Survey is filed as copyright, detailed in [1], We have conducted survey to the varied demography of the farmer on different time and at different location. To improve the accuracy of the survey, we have interviewed the stakeholders on one-to-one basis directly. Our survey to the agriculture stakeholders of Gujarat also enables and facilitate us examining the channels of information communication, gathering information need and knowing varied need in natural Guajarati language. Facts and figures of the survey leads to the necessity of intelligent retrieval system.

Keywords Agriculture survey · Farmer information need in agriculture · Indian agriculture questionnaire

1 Introduction

According to World Bank Data,[1] 70% of the world's poor who live in rural areas and are poorly educated, Agriculture is the main source of income and employment. 64.93% of land area is used for Agriculture where Crop Production Index is also increasing exponentially, i.e., 143.6 in 2020 (Base 2016–2017) which proves that Agriculture plays an important role in Indian Economy. According to Agricultural

[1] http://data.worldbank.org/—The World Bank Group is one of the world's largest sources of funding and knowledge for developing countries.

A. Shah (✉) · J. Pareek
Department of Computer Science, Gujarat University, Ahmedabad, India
e-mail: axitashah@gmail.com

J. Pareek
e-mail: jspareek@gujaratuniversity.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_52

Statistics at a Glance[2] 2020 [2]. In 2011 68.9% of India's population lives in rural India for their livelihood. 263.1 million Population from total 1210.9 million population works in Agriculture from that 45.1% are cultivators and other 54.9% are Agricultural laborers. There are 8 union territories and 28 states in India. Average 57.8% of Agricultural households to rural households where in Gujarat, it is 66.9%.[3]

Farmer's Information Need can be satisfied by the varied human stakeholders in Agriculture or by the technology use. Agriculture expert stakeholder who has knowledge of farming help farmers to resolve their problem. But Several Farmers do not have an access of other human expert. Hence, Technology usage is blessings for them. At one side poor people are poorly educated while the other side Literacy rate in Adult and youth is also increasing, i.e., 74%. Literacy Rate of Gujarat is also 79.31% that shows increase in the usage of existing online and offline knowledge. In Gujarat farmers are getting the benefit of Directorate of Agriculture to increase and protect the crop production.

There are problems in getting the data/information like Static Information, More than one information source, unavailability of an expert, etc.… which cannot be resolved by using Traditional App or Expert's direct view [3]. Hence, this information leads to the necessity of an agriculture expert system which retrieves from diverse sources. We have explored existing agriculture data sources which is useful to resolve farmer's query. Sources of Information from trusted web sites and publication by agriculture expert include but are not limited to Agriculture crop cultivation textual knowledge, statistical data etc., by Ministry of Agriculture, Department of Agriculture, Government of India, Indian Agriculture Universities, Agriculture Companies and so forth publishes numerous online and offline informative documents for the welfare of the farming individuals. This is the place where data of agriculture exists on Internet in a large volume. The client needs different types of information for crop cultivation. We came to know about it from our agricultural survey [1]. And it is very important to have data to provide the information that the stake holders need. We have researched, reviewed and stated the sources of such data and the information in the *Agriculture Content-Set* section available from the government and other institutes that is required for direct access or as part of technical retrieval system research. This data can be useful to guide the farmer about crop cultivation system. These data are available on Internet for the usage of the farmers that is definite knowledge seeker.

Researchers in this field have developed technical application to fulfill the demand of information and knowledge seeker. Information and Communication Technology systems and applied artificial intelligence techniques are available at farmer's doorstep to sort out farmer's information need. We have reviewed and summarized literatures of applied AI in Agriculture and stated in *Research status of*

[2] Remark: Agricultural Statistics at a Glance is a flagship publication of the Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India containing data on a wide range of economic indicator, particularly in agriculture sector. https://eands.dacnet.nic.in/PDF/Agricultural%20Statistics%20at%20a%20Glance%20-%202020%20(English%20version).pdf.

[3] https://en.wikipedia.org/wiki/2011_Census_of_India.

applied AI in Agriculture section. It will be better if we provide this knowledge to its users on one platform.

Information need of Agriculture stakeholders mostly achieve via friend advise, expert adviser, Mobile SMS, Helpline no., Cable tv, Newspaper, Expert Adviser, Training and Demonstration, Chat, WhatsApp, video or Audio conference. To understand farmer's information need, we have surveyed few of the agriculture stakeholders of Gujarat and collected information related to their crop cultivation. Our Survey by Questionnaire is explained, summarized and analyzed in *Analysis and Interpretation of Agriculture Crop Cultivation System Survey* section. From the other established sources and from the survey, we have collected 2848 user's query in natural language. We have also prospected stakeholders' information need in a form of natural language query as a part of survey.

2 Agriculture Content-Set

Informative web access provided by various autonomous entities in Agriculture includes but are not limited to, Department of Agriculture, Cooperation and Farmers Welfare,[4] Ministry of Rural Development,[5] Area and Production Statistic by Ministry of Agriculture and farmers welfare,[6] Government Agriculture Data,[7] The Indian Council of Agricultural Research (ICAR)[8]—autonomous organization under the Department of Agricultural Research and Education (DARE), Ministry of Agriculture, Government of India, Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India At a Glance Report,[9] Indian Agricultural Statistics Research Institute (IASRI),[10] Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India,[11] National Institute of Agricultural Extension Management,[12] Food and Agriculture Organization of the United Nations,[13] The International Crops Research Institute for the Semi-Arid-Tropics,[14] Indian Agriculture College/Universities News,[15]

[4] http://agricoop.nic.in/.
[5] https://rural.nic.in/.
[6] https://aps.dac.gov.in/Home.aspx?ReturnUrl=%2f.
[7] https://data.gov.in/search/site?query=agriculture.
[8] http://www.icar.org.in/.
[9] http://eands.dacnet.nic.in/Default.htm.
[10] https://iasri.icar.gov.in/.
[11] http://agricoop.nic.in/.
[12] https://www.manage.gov.in/default.asp.
[13] http://www.fao.org/index_en.htm.
[14] http://eprints.icrisat.ac.in/information.html.
[15] http://agricollegenews.com/.

Agriculture Technical News,[16] Indian Agriculture Retailer News,[17] Agricultural and Processed Food Export Development Authority,[18] Agriculture Insurance,[19] Farmer's Portal,[20] AGMARKET,[21] mKisan—SMS Portal for Farmers by Department of Agriculture and Cooperation,[22] Indian Agriculture Jobs,[23] Consortium for e-Resources in Agriculture,[24] Indian Agricultural Research Institute (IARI)—Journal,[25] e-learning portal on Agricultural Education[26], Information in Agriculture Science,[27] Bhartiya Kisan Sangh gujarat,[28] Data bank on Agriculture and allied sector,[29] four Agriculture Universities of Gujarat,[30,31,32,33] Online Agriculture Community[34], Agriculture Research Article Search Websites[35], Agriculture Knowledge Discovery Websites[36,37], Agricultural Census[38], Agriculture Time series of India's economic indicator by EPW Research Foundation[39] and other Indian Agricultural Web Sites indiaagristat[40], khetiwadi[41], kisan[42], ikhedut.aau[43], nabard[44], agriwatch[45], etc.…

[16] http://agtechnews.com/.
[17] http://agriretailers.com/index.htm.
[18] http://apeda.in/apedawebsite/.
[19] http://agriinsurance.com/.
[20] http://farmer.gov.in/.
[21] http://agmarknet.nic.in/agmarknetweb/default.aspx/.
[22] https://mkisan.gov.in/.
[23] http://www.indiaagronet.com/Agriculture-Jobs/.
[24] http://www.cera.iari.res.in/index.php/en/.
[25] http://epubs.icar.org.in/ejournal/.
[26] https://ecourses.icar.gov.in/.
[27] https://www.agriscienceindia.in/.
[28] http://bharatiyakisansangh.org/.
[29] https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS.
[30] https://nau.in/index.
[31] http://jau.in/.
[32] http://www.aau.in/.
[33] http://www.sdau.edu.in/.
[34] http://www.agricultureinformation.com/.
[35] http://oajse.com/subjects/agriculture.html.
[36] http://www.agnic.org/.
[37] http://www.sciseek.com/Agriculture/.
[38] http://agcensus.nic.in/.
[39] http://www.epwrfits.in/Agriculture_All_India_State.aspx.
[40] http://www.indiaagristat.com/.
[41] www.khetiwadi.com.
[42] www.kisan.net.
[43] http://faq.ikhedut.aau.in/.
[44] www.nabard.org.
[45] www.agriwatch.com.

Agriculture and Horticulture Data Sources are also available for Retrieval as per the user's demand. There are around 1200 freely available textual and multimedia resources on government and private institutes websites categorized and listed as dag,[46] atma gujarat,[47] sdau,[48] aau, [49] icar,[50] iari,[51] nau,[52] jau,[53] gov farmer,[54] ikhedutaau,[55] ikhedut gujarat,[56] agricoop,[57] soilhealth.dac.gov,[58] dacnet,[59] prsindia,[60] doh.gujarat,[61] gujecostat,[62] gsfcagrotech,[63] nhb,[64] agriculturetoday,[65] youtubechannels on Agriculture[66] and topical Search engine as google.[67]

Agriculture information is published weekly or monthly in the newspapers/magazines such as National Agriculture Today, Kisan samaj in Hindi from Indor, krushik jagat[68] in Hindi from Bhopal, Krushi Jagran,[69] krushi prabhat,[70] Agro Sandesh,[71] Krushi Vignan,[72] Real Target, Krushi Govidhya by Aanand University, Krushi Goldline in Hindi from Jaypur, Kisan Sandesh, Fasal Kranti,[73] Digital agrimedia film, agro house publication.[74]

---

[46] https://dag.gujarat.gov.in/images.
[47] https://atma.gujarat.gov.in/.
[48] http://www.sdau.edu.in/.
[49] http://www.aau.in/sites/default/files/.
[50] https://icar.org.in/sites/default/files.
[51] https://www.iari.res.in/files.
[52] http://old.nau.in/files.
[53] http://www.jau.in/attachments.
[54] https://farmer.gov.in/.
[55] http://agri.ikhedut.aau.in.
[56] https://ikhedut.gujarat.gov.in/.
[57] http://agricoop.gov.in/.
[58] https://soilhealth.dac.gov.in/Content/FAQ.
[59] https://eands.dacnet.nic.in/.
[60] http://prsindia.org/sites/default/files/.
[61] https://doh.gujarat.gov.in/.
[62] https://gujecostat.gujarat.gov.in.
[63] http://www.gsfcagrotech.com/.
[64] http://nhb.gov.in/.
[65] http://www.agriculturetoday.in/.
[66] https://www.youtube.com.
[67] https://www.google.com/.
[68] https://www.krishakjagat.org/Web/single_epaper/486.
[69] https://subscription.krishijagran.com/.
[70] http://www.krushiprabhat.com/epapers-listing.
[71] http://sandeshepaper.in/edition/57311/agro-sandesh.
[72] https://krushivigyan.blogspot.com/2018/12/scintilla.html.
[73] http://fasalkranti.com/news.
[74] http://agrohouse.in/agrohouse/shop_book.php#.

Agriculture industry publishes the facts and figures with statistics by magazines, articles, periodicals, textbooks as in textual or multimedia format as data and information. Researchers from universities, industries and government institutes working on Agriculture generate or collect knowledge in database as an example, Soil Health Card,[75] faolex[76] and data warehouse. Healthy e-mail/message communication of community expert is also available as a community data. Generated predicted knowledge is also helpful to the Farmer. Indian farmer can use this available form of information via technical application to increase productivity of farming and generate revenue.

3 Research Status of Artificial Intelligence in Agriculture

Vast research in Academia has been carried out for the applications of artificial intelligence in Agriculture. Applied Agriculture scopes a variety of topics specified in [4–9] includes but are not limited to, Intelligent retrieval, Information system, expert system, IOT-based system, machine learning, geographic context-based expert system, Question-answering system, Agri Robot, Sense-based expert system, Rule-based expert system, Fuzzy logic-based expert system, etc. Private Industry and the government also inflate its agriculture framework by providing information via web portal and mobile applications. Existing providers' study has been researched, reviewed and submitted it as an article in our submitted paper [10] that talks about research status of applied artificial intelligence approaches in Agriculture and the research gap.

Various researchers have worked in the retrieval process from Agriculture content. Techniques such as Information Retrieval (IR) [11], Question-answering (QA) system [12–14], Search engine [15–17], and Natural language and keyword-based interfaces are developed to get solution of user's information request. They have worked in the optimization of crop cultivation system, from the seed plantation till the harvesting process. These techniques are implemented using data, rules, thesaurus and ontology-based knowledge. Retrieval is improved by understanding user's request and retrieval sources such as documents in IR, and QA system and web pages for search engine.

Based on our study, we hypothesized the requirement of an intelligent system that retrieves from diverse government data sources and that requires to fulfill the information need of the farmer. To know the information need, we have prepared the survey, detailed in [1], conducted the survey and its interpreted analysis and results are as follows.

[75] https://www.soilhealth.dac.gov.in/.

[76] faolex.fao.org/faolex/index.htm.

4 Analysis and Interpretation of Agriculture Crop Cultivation System Survey

To fulfill the information need of the Agriculture Stakeholder, specifically Farmers in Agriculture Crop Production System, we had conducted a Survey via well-structured Agriculture Questionnaire as copyright Filed for "Survey-Questionnaire of Indian Agriculture Stake Holder's Information Need for Agriculture Crop Cultivation Information System of India". Detail Survey Questionnaire is given in the filed copyright—[1]. Objective of the survey is to know the varied user's natural language request as a part of Man–Machine Interaction, to gather the information needs of farmer community in Gujarat, India and to examine the channels of information communication and sources of information dependence used by the farmers.

It is an agriculture survey for Agriculture Crop Production Information System of Gujarat, India in Questionnaire form. We have prepared Questionnaire in both English and Gujarati languages. This Survey is distributed among 5 categories, i.e., personal information of stakeholder, Questions regarding Business/Job in Agriculture Crop Production System, Queries for Agriculture Crop Production and Role of IT in Agriculture. Agriculture questionnaires collect information on a core set of indicators such as Basic crop production of main crops, with emphasis on improved measures of Quantification of production and Plot size, Land holdings, Farming practices, i.e., Mechanization, Soil and environmental management, Water management, Input use and technology adoption, Use of technology and farming implements, Seed varieties, Fertilizer, pesticides/herbicides applications, etc.…We have referred several Indian government agriculture websites and Indian council of agriculture research (ICAR) benchmark book published by agro house [18] and Agriculture Categories by EPW Research Foundation[77] to prepare this Questionnaire.

Kindly refer Survey-[1] to get in detail considerate category wise questions. We have conducted this survey by visiting AgriAsia-2016, 2017 and 2019, (Companies Visited in AgriAsia: Agrostar, reliance foundation, vimax rop science ltd, gnfc, Agriculture Farmers welfare and cooperation department—Government of Gujarat), Gujarat Vidhyapith KVK Seminar and Department of Agriculture, Horticulture and Statistics, Government of Gujarat. There are around 308 individuals' involvement with 272 survey responses from AgriAsia, and 36 opinions by conducting an interview of Experts, Advisors, Government Domain Employees for knowing their viewpoint in Agriculture research.

[77] http://www.epwrfits.in/Agriculture_All_India_State.aspx.

5 Results and Discussion: Analysis and Interpretation of Agriculture Crop Cultivation System Survey

Our survey analysis, after data collection and preprocessing is as follows: Diverse Age group of participants in survey below 20 Yrs is 4.5%, 20–40 is 42.5%, 40–60 is 48.5% and 4.5% are above 60. From them 88.8% are Male and 11.2% are female. Around 94% from them own the farming unit.

Here, in Fig. 1 depicts personal facts and figures of survey participant. A of Fig. 1 expresses the speech ability of local, national and international languages. More participants know Gujarati compared to Hindi, English and other languages. More than 60% participants are having basic knowledge of technology usage such as use of computer, mobile phone and application, summarized in B of Fig. 1. Varied Education Qualification of respondents in C and respondent's agriculture role in D are specified in Fig. 1

A to E of Fig. 2 depicts figurative summary of respondent's agriculture crop cultivation system practice. Most of the respondent's crop cultivation frequency in a year is twice shown in A of Fig. 2. Total counts of types of soil at farm in B of Fig. 2, total categories of crops which you are producing in farm in C of Fig. 2, total count of preferred medium of information sources in C of Fig. 2, Sources of farmer's help need summary count in D of Fig. 2 and Requirement of information need's category summary in E of Fig. 2 are demonstrated in relevant chart. These

Fig. 1 Personal informative analysis of survey participants

Fig. 2 **a–f** summary of agriculture crop production system usage by survey participants

analyzed, interpreted and demonstrated summary of Figs. 1 and 2 indicates diversified background of survey respondents.

Figure 3 identifies varied application usage by the respondents in the Bar chart. Out of these, Agrostar, Agrimedia, Mandi and mkrishi are maximum used among the community. Application usage may get changes depending on the current availability of application, latest functional and error-prone applications in the market. Around

Fig. 3 Bar chart of usage of technical application by participants in agriculture

70% respondents agree with 50–100% varied span of technology support is important in farming. 83% agree that technology supports have definite role in Agriculture where 10% disagree and other 7% do not know about it. And around 57% farmers who are using Kisan Call Center Service are satisfied with the expert's advice.

Query Set in Natural Gujarati Language: We have collected around 1339 agriculture's stakeholder's request query. Out of 1339, 566 is from AgriAsia 2016, 506 is from AgriAsia 2019, 199 is from scientists at Kisan Vigyan Kendra, and other 68 from companies like GNFC, Kisan Call Center, Reliance Foundation, Agrostar, Agriculture farmers welfare and cooperation department-government of gujarat at AgriAsia via interview. We have gathered around 2848 agriculture's stakeholder's request query analyzed based on the frequency of keywords and represented by Keyword Cloud in the Fig. 4. Out of 2848, 1339 as a part of questionnaire, 299 from Agriculture FAQ Government Book, 625 is from Agrimedia mobile app and 585 is from IKhedut FAQ. Our analysis identifies mostly asked types of the agriculture queries listed as in Table 1.

Our result of conducted survey toward Agriculture Stake holders cognize farmer's context and situation in Gujarat and their need as a part of natural language queries. It also identifies the necessities of tools and technologies as well as agriculture scientific tips.

Fig. 4 Keyword cloud of agriculture stake holder's requested query

Table 1 Analyzed categories of agriculture queries

Mostly Asked Agriculture Queries
ફળો અને શાકભાજી ઉગાડવાની રીતો
ખાતર ને લગતા પ્રશ્નો, ખેતીમાં રોગને લગતી માહિતી
પાક માટે દવા અને પાણી નો છંટકાવને લગતા પ્રશ્નો
જીવાત અને ઢોરથી પાકોનુ રક્ષણ કરવાના ઉપાયો
પાકની લણણી અને કાપણી કરવાની વિવિધ રીતો
સરકારની વિવિધ યોજનાઓ જેમકે વીમા યોજનાઓ
જંતુનાશક અને ફુગનાશકના છંટકાવ ની માહિતી
સુધારેલી પાકની જાતના ફાયદા
પાકની નીંદણ માટેનો વૈજ્ઞાનિક અભિગમ
વિવિધ પ્રકારની જમીનો ની માહિતી
ઔષધીય ખાતર તૈયાર કરવાની પધ્ધતિ
ઋતુ ને લાયક વરસાદ અને તાપમાન ને અનુરૂપ પાક ની માહિતી
વિવિધ પાકના સ્તરે ટેક્નોલોજી નો ઉપયોગ
સિંચાઈ ની સગવડ ને લગતા પ્રશ્નો

6 Conclusion

Analysis of undertaken survey shows immense potential to standardize, regulate and automate the agricultural processes and solve the problems in Indian Context in a form of search and retrieval-based artificial intelligent system. Our current analysis escort toward viable, efficient and cost-effective solution of obtaining wisdom

through the working with independent individual. Our Survey to the Agriculture Stake holders cognize us about the information need of varied farming practices to diverse demographics. Even our future research will analyze farmer's Information need that is in Natural Language Query. We will extensively explore, classify, process and analyze natural language query.

Diverse available data and information resources in natural language should be utilized on one platform via an agriculture expert information system. Even Agriculture Stakeholder is not required to know the technicality context of the resource. In future, we will propose Retrieval system that fulfill the above specified mandatary request. This Retrieval system will provide easy to access, user friendly interface, one-point access mechanism for diverse information needs, even accessible by non-technical persons, and provide relevant solution to the user's request. The complexity of this implementation would be diverse storage and access mechanism of diverse computer resources, natural language ambiguity, difficulty in defining knowledge and inference for the same.

We hope that this work will help researchers in the field to acquire the necessary information and technologies to further conduct more advanced research.

References

1. A. Shah, J. Pareek, Survey-Questionnaire of Indian Agriculture Stake Holder's Information Need for Agriculture Crop Cultivation Information System of India. Literary/Dramatic works Copyright Certified—L-113297/2022. (2021)
2. G. India, *Of: Agricultural Statistics at a Glance—2018* (Controller of Publication, New Delhi, Government of India, 2018)
3. A. Rafea, *Expert System Applications: Agriculture. Central Laboratory for Agricultural Expert Systems*, PO Box. 100, (1998)
4. B.T. Bestelmeyer, G. Marcillo, S.E. McCord, S. Mirsky, G. Moglen, L.G. Neven, D. Peters, C. Sohoulande, T. Wakie, Scaling up agricultural research with artificial intelligence. IT Professional. **22**, 33–38 (2020)
5. Y. Awasthi, Press "A" for artificial intelligence in agriculture: a review. JOIV : Int. J. Inf. Visual. **4** (2020). https://doi.org/10.30630/joiv.4.3.387
6. A. Gambhire, B.N. Shaikh Mohammad, Use of artificial intelligence in agriculture. SSRN Electron. J. (2020). https://doi.org/10.2139/ssrn.3571733
7. S. Chaterji, N. DeLay, J. Evans, N. Mosier, B. Engel, D. Buckmaster, R. Chandra, Artificial intelligence for digital agriculture at scale: techniques, policies, and challenges. arXiv preprint arXiv:2001.09786 (2020)
8. M. Pathan, N. Patel, H. Yagnik, M. Shah, Artificial cognition for applications in smart agriculture: a comprehensive review. Artific. Intell. Agricult. **4**, 81–95 (2020)
9. K. Jha, A. Doshi, P. Patel, M. Shah, A comprehensive review on automation in agriculture using artificial intelligence. Artific. Intell. Agricult. **2**, 1–12 (2019). https://doi.org/10.1016/j.aiia.2019.05.004
10. A. Shah, J. Pareek, Contemporary research status and recommendation of applied artificial intelligence in agriculture. Towards Excellence. 14, (2022)
11. A.S.Sharma, A., A framework for semantics and agent based personalized information retrieval in agriculture—IEEE Conference Publication. In: 2nd International Conference on Computing for Sustainable Global Development (INDIACom) (2015)

12. T. Kawamura, A. Ohsuga, Question-answering for agricultural open data. In: *Transactions on Large-Scale Data-and Knowledge-Centered Systems XVI*. pp. 15–28. (Springer, 2014). https://doi.org/10.1007/978-3-662-45947-8_2
13. S. Gaikwad, R. Asodekar, S.Gadia, V.Z. Attar, AGRI-QAS question-answering system for agriculture domain. In: *2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015*. pp. 1474–1478. Institute of Electrical and Electronics Engineers Inc. (2015). https://doi.org/10.1109/ICACCI.2015.7275820
14. M. Devi, M. Dua, ADANS: an agriculture domain question answering system using ontologies. In: *Proceeding—IEEE International Conference on Computing, Communication and Automation, ICCCA 2017*. pp. 122–127. Institute of Electrical and Electronics Engineers Inc. (2017). https://doi.org/10.1109/CCAA.2017.8229784
15. N. Xie, W. Wang, *Research on Agriculture Domain Meta-search Engine System*. Presented at the October 18 (2009). https://doi.org/10.1007/978-1-4419-0211-5_68
16. W. Li, Y. Zhao, B. Liu, Q. Li, *The Research of Vertical Search Engine for Agriculture*. Presented at the October 18 (2009). https://doi.org/10.1007/978-1-4419-0211-5_2
17. C.Y. Sufen, T.P., L.S., S., Study of agricultural search engine based on FAO agrovoc ontology and google API. In: *World Automation Congress*. IEEE (2010)
18. Indian Council of Agricultural Research: Handbook of Agriculture. Directorate of Knowledge Management in Agriculture, New Delhi (2012)

System Design and Implementation of Assistive Device for Hearing Impaired People

P. Nikita, B. H. Shraddha, Venkatesh Mane, N. Shashidhar, P. Vishal, and Nalini C. Iyer

Abstract This paper discusses the design, implementation and development of assistive device for hearing impaired. Sign language is one of the oldest and most natural form of language for communication, but since most people does not know sign language. Finding interpreters for different group of people is difficult. To treat the deaf people as one among the society, a user-friendly assistive device is necessary. This paper provides the details about the implementation of standalone interpreter using transfer learning for finger spelling-based American Sign Language using Raspberry Pi. A Graphical User Interface (GUI) is created and tested for establishing a two-way communication to convert text into sign language. A language barrier is created using an assign language structure that is different from normal text. Hence, the users will now depend on the vision-based communication that will be able to bring normal people, deaf and mute people on the same grounds of interaction. There is a need to build a possible sign language translator, which can take communication in sign language and translate them into written and oral language. Such a translator that uses neural networks for finger spelling-based American Sign Language would greatly lower the barrier for many deaf and mute individuals to be able to better communicate with others in day-to-day interactions. In proposed method, hand gesture is first passed through a filter, and after the filter has applied,

P. Nikita (✉) · B. H. Shraddha · V. Mane · N. Shashidhar · P. Vishal · N. C. Iyer
KLE Technological University, Hubballi, India
e-mail: nikhita_patil@kletech.ac.in

B. H. Shraddha
e-mail: shraddha_h@kletech.ac.in

V. Mane
e-mail: mane@kletech.ac.in

N. Shashidhar
e-mail: shashisn@kletech.ac

P. Vishal
e-mail: vishalbps@kletech.ac.in

N. C. Iyer
e-mail: nalinic@kletech.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_53

the gesture is passed through a classifier which predicts the class of the hand gestures. This method provides 95.7% accuracy for the 26 alphabets and 0, 1, 2 numerical.

Keywords American Sign Language · GUI · Translator · Interpreters

1 Introduction

One of the five senses necessary for human survival is hearing. Language acquisition, learning to speak, and general knowledge all require the use of the hearing sense. Hearing, above all, allows people to communicate and socialize with others. Hearing loss affects deaf persons in numerous ways; they are unable to communicate or mingle properly with the general public. Furthermore, deaf people's safety is jeopardized in many situations because they are unable to make the correct decision based on the brain's analysis of noises, particularly worrisome ones. Hearing loss is treated differently depending on the severity of the condition. The only disability dumb and deaf people have is related to communication and they cannot use spoken languages hence the only way for them to communicate is through sign language and American sign language is a predominant language. Hands are the essential part of body which will be used by deaf and dumb people to communicate their thoughts with other people by using different gestures. Gestures are represented in non-verbal form where in there would be exchange of messages and these gestures will be understood with vision. This non-verbal communication of deaf and dumb people is called sign language. Sign language is often considered as a visual language. The three major components involved in this language are as follows [4].

Finger spelling: The finger spelling method also known as the manual alphabet is a manual form of communication.

World level sign vocabulary: This method involves collection of large amount of data set. The experimentation will be carried out using multiple deep learning methods to identify the signs depending on the words and performs its evaluation. This method is applicable to large-scale scenarios.

Non-manual features: It is the combination of eye gaze, position and movement of head and upper body. This basically creates a model that predicts the finger spelling-based hand gestures to produce a full word. Figure 1 shows the different gestures that are trained for the proposed methodology and also reflects the American Sign Language symbols.

American Sign Language is the main sign language. Since the only obstacle for deaf mute people is communication, they cannot use spoken language, so the only way to communicate is through verbal gestures [5–8].

Fig. 1 American sign language symbols

2 Literature Survey

Through the recent years there has been tremendous research carried on the hand gesture recognition which do not ensure reliable results.

- Mahamud and Zishan A hearing aid for the deaf and hard of hearing. "In this study, an assistive gadget for hearing impaired people is proposed, allowing them to communicate freely with anyone who comes close to them". The disabled person must wear a wristwatch-style device and carry an Android phone. This device receives voice and continuously converts it into legible text, which is displayed on the LCD screen. When someone wants to hear something from them, all they have to do is open the application on their phone and go through his watch [1].
- Dhanjal and Singh "Tools and Techniques of Assistive Technology for Hearing Impaired People". Education is the most crucial aspect of any human life since it allows us to advance in life by allowing us to transmit information through speech, writing, or other means. Hearing loss or deafness makes it harder to communicate and comprehend others. Hearing impaired people can communicate with normal people with less stress with assistive technology. Assistive Listening Devices (ALDs), Augmentative and Alternative Communication (AACs), and Alert systems are three types of assistive devices [2]. Various tools and strategies based on parameters like as alarm systems, speech augmentation, SL recognition and learning apps, telecommunication systems, and voice to text are described in this study article. The majority of hearing aids on the market are prototypes and not yet commercially available.
- Ren et al. [3]. "Wireless Assistive Sensor Networks for the Deaf". The purpose of this study is to develop a wireless sensor network-based assistive system (WASN) to assist the deaf or hearing impaired in being more aware of their surroundings. There's also a quick rundown of the latest assistive technology. For two common application scenarios, smart home and smart school playground, the system architecture, components, and specifications are described. To implement the system,

a node platform is being built. Finally, simulation is used to assess the system's performance, which is followed by a feasibility and availability analysis [3].

3 Proposed Methodology

This section discusses about the different design alternatives and the proposed design methodology.

(i) Hand gesture recognition based on computer vision (Fig. 2):

This method uses computer vision technology to recognize the user's gestures that uses one or two hands to perform specific gestures in front of the camera. It will be connected to the frame of the system, including gestures that use one or two hands in front of the camera. These kinds of systems include various methods of extracting features and classifying gestures as per the application requirement [9].

(ii) Real-time recognition of gestures through finger segmentation:

The block diagram of real-time recognition of gestures through finger segmentation is shown in the Fig. 3. The process is as follows. Initially perform hand recognition by subtracting the background part of it. The result of hand recognition is converted into a binary image. The principle of segmentation is applied to the fingers and palms to facilitate the finger recognition. Thus, gestures are recognized by a simple rule classifier in these kind of systems.

(iii) AI hand tracking and gesture recognition:

Gesture recognition provides computer with real-time data for executing user commands. The motion sensor in this device can track and interpret gestures and use

Fig. 2 Hand gesture recognition based on computer vision

Fig. 3 Block diagram of real-time recognition of gestures through finger segmentation

Fig. 4 Block diagram of hand tracking and gesture recognition using AI

it as the main source of input data. Most of the motion control solutions involve the combination of 3-D cameras, depth sensors, infrared cameras and machine learning systems. The machine learning algorithm learns from the marked depth image of the hand so that it can recognize the position of hand and fingers (Fig. 4).

Gesture recognition consists of three basic levels. They are as follows.

Detection: This device uses a camera to detect hand or body movements. With the help of machine learning algorithms, the image is segmented to determine its edge and position of the hand. **Tracking**: This device tracks the movement frame by frame, record each and every movement and provides the accurate input for data analysis [6]. **Recognition**: This system tries to find the patterns based on the collected data. When the system finds the match and interprets the gesture, it will perform the action associated with the gesture. The feature extraction and classification in the following scheme realizes the recognition function. This completes the discussion of three different design alternatives of assistive systems. The discussion of the proposed design is as follows.

The functional block diagram is shown in the Fig. 5. The functions are expressed as a set of elementary blocks connecting the inputs and outputs. The role of web camera is used to recognize the gestures. They are further applied to the preprocessing block for converting the images to the standard sizes. The system is developed in such a way that converts the gestures into its respective text. Further the images are stored

Fig. 5 Functional block diagram of system

in the database. The system uses the appropriate database to convert the image to text. Thus, it can be concluded that this system provides the output data in the form of text which in turn helps to reduce the communication gap between deaf and dumb people.

4 Implementation

With the help of the proposed design, the system architecture is developed and the specification of each component is discussed in this section. It also includes the detailed description of the algorithm and flowchart of the proposed design. **Data Collection**: It is the method of collecting data and information of the variables which will be targeted in a conventional system. It provides the authorization to respond to the questions which will be relevant and further estimates the outcomes. It is often considered as a component of research in all the kind of study fields (including technical and non-technical). The methods may diverge by discipline, the prominence on ensuring precise and sincere collection remains the same (Fig. 6).

The complete data set consists of 14,500 images which are of 60 × 60 pixels. The total number of classes that are available in this data set are 29 [each comprises with 500 images], 26 for the letters A–Z and 0,1,2 numbers. This data is the sole of the user gesturing in ASL, with the images taken from the laptop's webcam. These photos were then cropped, rescaled, and labeled for use. Note the difficulty of distinguishing fingers in the letter E. Five different test sets of images were taken with a webcam under different lighting conditions, backgrounds, and use of dominant/non-dominant hand. These images were then cropped and pre-processed. **Data Augmentation**: Data expansion during data analysis is a technique of expanding the amount of data by adding a slightly modified copy of existing data or newly created synthetic data from existing data. It acts as a regularizer and helps to reduce over fitting when training

Fig. 6 Data collection

machine learning models. **Transfer learning**: This system uses transfer learning methods to build, train, and test models. Transfer learning is a machine learning technique in which a model trained on one task is redirected to a second related task. Focus on storing the knowledge gained from solving one problem and applying it to another related problem. In practice, few people train a convolution network from scratch (randomly initialized) because they rarely get enough data sets. Net weights, such as fixed feature initialization or extractors, help to solve most of the discussed problems. Training very deep networks is expensive. It takes weeks to train the most complex models on hundreds of expensive GPU-powered machines. The learning method/hyper parameter of deep learning is a black art, and there is not much theory for us to guide. In transfer learning, the data set is trained by using the underlying network and the underlying problem, and then reuse the learned attributes or transfer them to the second target network to train the task and target data set. If the characteristics are general, meaning suitable for basic tasks and objective tasks, not specific to the main task. **Flow of Algorithm**: The flow of the algorithm and Gesture reorganization is shown in the Fig. 7. **Layer 1**: The Gaussian blur filter is applied and some threshold value will be set to the image which is captured by Open CV to obtain the processed image after feature extraction. This processed image is passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for forming the word. Space between the words is considered using the blank symbol.

Layer 2: This layer recognizes different fonts, displays similar results during recognition. It also tries to classify collections using a classifier.

Flowchart:

The overall design flow of the system is represented in the form of flowchart shown in Fig. 8. The flowchart shows the complete architecture of the system. The device is split into phases—schooling section and checking out section. In the training period, the gestures will be captured by the camera using implicit camera. The captured

Fig. 7 Flow of algorithm and gesture reorganization

Fig. 8 Flow chart of the designed algorithm

images will be identified by using the image processing techniques such as image enhancement and segmentation, process of color filtering, skin segmentation and reduction of noise techniques. The images are further processed to extract the vital features using count value. They will be reserved in a file folder and then compared with the test images. In test period, the user makes gestures in front of the camera, the gestures are captured as image by the camera using virtual mouse. Images are further processed and their features will be extracted. The test images will be compared with the stored images. If the gestures are recognized, the output is produced in the form of text. Otherwise an error message is generated. **Training and Testing**: The training data convert input images (RGB) into grayscale and apply Gaussian blur to remove unnecessary noise. Adaptive threshold is applied to extract the hand from the background and resize the images to 64X64. After applying all the above operations, the pre-processed input image is loaded into the model for training and testing. The prediction layer estimates how likely the image will fall under one of the classes. So that the output is normalized between 0 and 1 and such that the sum of each value in each class summates to be 1. This is achieved using the soft max function.

The result of the predicted level will initially be a bit far from the actual value. In this way the network will be trained with the labeled data.

The images related to the testing have been shown in the Figs. 9, 10 and 11

Fig. 9 Training data from 0 to Z

Fig. 10 RGB testing images from 0 to O

Fig. 11 RGB testing images from P to Z

Fig. 12 Demonstrating the various letters and numbers

5 Results and Discussion

The discussion about the obtained results is as follows. Figure 12 shows the demonstration results showcasing various letters 7 numbers. The prediction window has two frames, one is the actual frame of 300 × 300 in dimension to capture the user input and another frame is of 64 × 64 which is a Gaussian filter window. The inputs of the user are captured and converted to gray images by applying Gaussian filter and these images are used for prediction. As shown in the pictures, the user's hand textures, background colors and other related parameters are changing and the prediction model is able to recognize any changes at the user's end and provide us the proper prediction results.

6 Conclusion

This paper finally concludes about the implementation of an automatic sign recognition system in real-time through migration learning. Implementation using basic methods is more effective than complex methods. The time constraints and difficulties involved in creating records from scratch is present. Until then, it is best to have a data set that can be used. In our live demonstration, some letters are more difficult

to classify, such as “a” and “i”, because they have a small difference (“I” puts the little finger up). The classification effect is very good. With pictures, there are many possibilities for future work. Possible extensions of the project include extending the gesture recognition system to all ASL letters and other non-letter gestures. The scope of the project can also be extended to various other applications, such as the use of gestures to control robot navigation.

References

1. M.S. Mahamud, M.S.R. Zishan, Watch IT: an assistive device for deaf and hearing impaired. In *2017 4th International Conference on Advances in Electrical Engineering (ICAEE)* (2017). https://doi.org/10.1109/icaee.2017.8255418
2. A.S. Dhanjal, W. Singh, Tools and techniques of assistive technology for hearing impaired people. In: *2019 International Confer- ence on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (2019). https://doi.org/10.1109/comitcon.2019.886245.4
3. H. Ren, M. Meng, X. Chen, Wireless assistive sensor networks for the deaf. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2006). https://doi.org/10.1109/iros.2006.282354
4. M. Wald, ”Captioning for deaf and hard of hearing people by editing automatic speech recognition in real time”, In: *Proceedings of 10th International Conference on Computers Helping People with Special Needs ICCHP*, LNCS 4061, pp. 683–690 (2006)
5. S. Ahmed, “Electronic speaking system for speech impaired people: Speak up”, in *2015 International Conference on Electrical Engineering and Information Communication Technology(ICEEICT)*, pp. 1–4 (2015)
6. Mohammed Waleed Kalous, Machine recognition of Auslan signs using Power Gloves: towards large-lexicon recognition of sign language
7. M.M. Zaki, S.I. Shaheen, Sign language recognition using a combination of new vision based features. Patt. Recogn. Lett. **32**(4), 572–577 (2011)
8. P. Nikita et al., Active learning in electronic measurements and instrumentation course through hands-on. J. Eng. Educ. Transform., [S.l.], Jan. 2016. ISSN 2394–1707. Available at: http://www.journaleet.in/index.php/jeet/article/view/85543. https://doi.org/10.16920/jeet/2016/v0i0/85543
9. B.H. Shraddha, et al., Model based learning of linear integrated circuit course: a practical approach. J. Eng. Educ. Transform., [S.l.], Jan. 2016. ISSN 2394–1707. Available at: http://journaleet.in/index.php/jeet/article/view/85549. https://doi.org/10.16920/jeet/2016/v0i0/85549

Internet of Things Security: A Blockchain Perspective

Mohammad Luqman and Arman Rasool Faridi

Abstract Since the arrival of computers in 1950s, technology has leaped over many generations at a rapid pace. Computers have found their use in every human activity, namely scientific experiments, industries, transportation, medicine, healthcare, home management, etc. Internet of things (IoT) is one such computing paradigm that aims to assist and automate various trivial and non-trivial aspects of these activities. IoT generates a huge amount of data at a fast rate. The data can be trivial, personal, or sensitive to someone or an organization. Hence, securing IoT data and its aggregation is of utmost importance. The paper gives a brief introduction of IoT in terms of its security aspects to understand the various facets of malicious attacks possible on it. Then, how blockchain, an ever-popular security suite, can be used to provide security to IoT in its various integrations is presented. Finally, some future research issues are discussed towards blockchain-IoT future.

Keywords Internet of things · Blockchain · Security · Internet of vehicles · Internet of medical things · Consensus algorithms

1 Introduction

Today's modern society is a data driven society. Personalized services are facilitating the ease of life in people's life round the clock. People commuting to home wants the shortest path with the least traffic in their route planning system; healthcare workers require accurate and timely reports of a patient body conditions or homemaker could use their smartphone to control AC or other devices in a smart home. This fine-grained insertion of computers in human life was made possible due to two major reasons that are the small form factor of microcontrollers and universal connectivity of embedded devices to other unassociated devices with the help of Internet. Internet

M. Luqman (✉) · A. R. Faridi
Aligarh Muslim University, Aligarh, Uttar Pradesh 202002, India
e-mail: luqman.geeky@gmail.com

A. R. Faridi
e-mail: ar.faridi.cs@amu.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_54

of things (IoT) is one such technology which is driving our data driven society. It is a universal network of things (smart objects) where things can get instructions from either human, from other things or from prebuilt commands.

Before we discuss various ways through which blockchain can secure IoT, a brief introduction of IoT and blockchain is needed to understand their significance in real-world applications. This paper aims to provide readers with the following details:

1. Introduction of IoT and its security requirements.
2. Acute analysis of blockchain and its functioning.
3. Discuss how various IoT networks are vulnerable to internal/external attacks.
4. Discuss the roles in which a blockchain can be used to provide security services to IoT and its similar network.

2 Internet of Things

IoT enables the soft touch of computers with the human life. The data collected are not stored by the IoT devices by themselves but instead are send to fog or cloud computing network where the data are stored and statistically analyzed for behaviors, routines, or patterns to enhance the quality of service provided to its user. Each device in an IoT network is directly connected to Internet thus has a unique address to identify itself over the network. This allows user's comfort to access their IoT device anywhere and anytime. IoT device provides and supports lots of applications in different areas such as transportation, industries, battlefields, healthcare, smart transportation systems, smart cities, smart traffic systems, e-governance, smart retails, smart farming, e-gaming, smart home, smart equipment, and automated security and surveillance systems. A survey by International Data Corporation (IDC) [1] indicates that more than 41.6 billion devices will operate by 2025 which will generate approx. 79.4 zettabytes of data. A survey by Kagan [2] estimates that 28% homes in USA will be smart by 2021 compared to 12.5% in 2016. Business insider intelligence [3] predicts 12 million sensors will be used in agriculture by 2023. Forrester [4] predicts that more than 250 million smart home devices will be installed in US homes by 2021.

IoT device being, generally, humancentric generates and transmits data which can be private, secretive, or possess critical value that create security and privacy concerns. Hence, securing the IoT data generated and transmitted as well as need of a secure centralized storage via fog/cloud computing to permanently store it are critically important. IoT security majorly consists of three security layers: hardware security, network security, and data security. **Hardware security** deals with securing the IoT device from tampering, destroying, depackaging, side-channel attacks, and other attacks to sensors and chips. **Network security** concerns with preventing crypt-analysis of embedded security cipher and the secure data transfer from transport layer. **Data security** concerns with secure execution of OS, application, and software on the IoT device, to provide secure support and middleware services to other

Fig. 1 Security overview of IoT

devices on the network and preventing execution of virus or similar rogue application. These attacks on the three layers of IoT are shown in Fig. 1 above. Blockchain is an upcoming solution which can guarantee security and privacy of IoT.

3 Blockchain

Blockchain is a digital ledger which is distributed, transparent, incorruptible, and equally accessible to its participants. It can store transactions which are immutable and verifiable. Blockchain network allows applications to store transactions in secure and peer-to-peer (P2P) manner. Blockchain forms the foundation of Bitcoin which is the first known cryptocurrency. Although blockchain was formally introduced with Bitcoin in a paper published by Satoshi Nakamoto [5] in 2008, but earlier attempts were made as well. David Chaum [6] in his dissertation proposed a protocol similar to blockchain in 1982. Stuart Haber and W. Scott Stornetta [7] in 1991 introduced a cryptographic system to securely timestamp documents. Stuart Haber, W. Scott Stornetta, and Dave Bayer [8] in 1992 improved upon the previous work on timestamping documents by using Merkle trees, hashcash [9] was formally introduced in 2002 paper as CPU cost function which was to be used in PoW. Blockchain benefits from the drawbacks of centralized system such as no single point of failure, amount

of trustworthiness of the centralized system, no monopoly on transaction cost, and speed or security. Some of the major uses of blockchain are in smart contracts, cryptocurrencies, digital assets, and distributed storage. Blockchain was developed keeping in mind the Byzantine's general problem [10]. Byzantine's general problem is a unique characteristic of a distributed system which indicates that some nodes in the distributed system may fault, and it is not possible to reliably determined its truthfulness. Blockchain guarantees valid state and permanent records even when adversaries try to disrupt the whole system, by properly consenting every transaction with each participants involved using some consensus algorithm. Some of the famous consensus algorithms are

- *Proof of work (PoW)* is the most famous algorithm. It was introduced with Bitcoin. People, group, or network which participate in creating blocks for Bitcoin are called miners. Miners participate in mining blocks to get rewards in form of Bitcoin. PoW algorithm relies on energy spent in mining any block which should not be less than the worth of the block produced. PoW trivially guarantees Byzantine's general problem. Bitcoin used hashcash to create a highly difficult mathematical problem which uses the previous block information among other things and challenges the miners to solve to receive rewards. First miner to solve the problem can add the block to the network and he/she will be rewarded with one Bitcoin. To prevent rapid mining of blocks by adding more compute units, Bitcoin network will automatically increase the difficulty of mining the new blocks by increasing the mining difficulty of the whole network.
- *Proof of stake (PoS)* is another famous consensus algorithm which is currently adopted by Solana, Cardano, Nxt etc., cryptocurrencies. PoS is deemed as alternative to the drawbacks of the PoW. The first cryptocurrency to use PoS was PPCoin [11]. PoS uses age of coin to determine the stake rights of a participant. Higher the age of a coin, more rights are given to its owner as well as some rewards. PoS encourages its user to hold the coins for longer time. Using coin age as important parameter in creating consensus, resource wastage is automatically reduced. Attackers trying to attack network needs to hold sufficient number of coins as well as coin age to successfully attack it.
- *Practical Byzantine fault tolerance (PBFT)* [12] is an improved and efficient version of Byzantine fault tolerance algorithm. Byzantine fault tolerance algorithm was used in distributed computing which provides better results but require exponential time complexity. In 1999, when PBFT was introduced, then the wider adoption of algorithm was started. PBFT decreases algorithmic complexity from exponential to polynomial which was a major improvement. It involves five step to achieve consensus that are request, pre-prepare, prepare, commit, reply.
- *Delegated Proof of Stake (DPoS)* [13] is an improved version of PoS, which was developed by Daniel Larimer in the year 2014. Bitshares is a cryptocurrency that uses DPoS. It utilizes less resources to operate the network and is more efficient compared to PoS and PoW algorithms. This algorithm selects witness based on the stakes hold by it for every node. Accounting rights are given to those witness which have participated in the network and are in top N position. The number N is

Fig. 2 Block structure of a blockchain

determined by the voting stakeholders when at least 50% of the voting stakeholders believe there is enough decentralization in the network.

The basic unit of a blockchain is a block which consists of version number, the previous hash value, current hash value, Merkle tree root, timestamp, and a nonce. Merkle tree or hash tree is a tree data structure, generally a binary tree, whose non-leaf nodes are hashes of its child nodes label while leaf nodes are hashes of data blocks. Nonce is a random value whose requirement is, it should be lower than or equal to the hash target set by the network. Figure 2 shows the basic structure of a block.

4 Applications of Blockchain in IoT and Similar Networks

4.1 IoT with Blockchain

IoT devices regularly perform trivial tasks on behalf of is owner, sense data, and transmit it to store in the cloud. But IoT devices are incapable to properly manage data or provide data filtration and reporting service. Some of the benefits which will come by integrating blockchain with IoT [14] are as follows:

- *Auditability*: Every transaction such as ordering, payments, delivery receipt, etc., can be stored in blockchain which can be used as proofs of service. Blockchain permanently stores records which is immutable and incorruptible and hence can be audited later for obtaining proofs.
- *Reliability*: Blockchain stores data in immutable and distributed form which guarantees against malicious data tampering or data loss. Moreover, blockchain can enable sensor data traceability and accountability which improves overall reliability of the network.
- *Identity*: A single blockchain can easily provide identity service in a IoT network [15]. It can easily identify devices by the data transmitted by them and provide

trusted, distrusted authorization, and authentication system in network. Data stored in blockchain can easily be traced back to its source node.

- *Autonomy*: Blockchain can greatly help in *hardware as a service* and smart autonomous assets. It will optimize the hardware usage in situations such as when a hardware is constantly drawing power but it is not performing any computational work. Using blockchain and smart contracts, devices can self-decide and self-manage its interaction and usage with other devices without the need of any backend solution such as cloud or fog computing.
- *Service Markets*: Blockchain can greatly enhance in the development of various new services and data marketplace. Services and payments can be made by the device itself securely even in trustless environment by using blockchain.
- *Secure code update/deployment*: Blockchain can provide secure and immutable data storage which can be used to provide regular code updates or deployment securely. Manufactures can reliably send updates and be ensured that the code update process will be successfully processed.

4.2 Internet of Vehicles (IoV) with Blockchain

IoT engulfs existing vehicular ad hoc network (VANET) into the new concept of Internet of vehicles (IoV) [16] in which nodes transmit data to everyone in the network, i.e., vehicle to vehicle, vehicle to traffic system, vehicle to road, vehicle to city, etc. Sensors embedded in vehicles generate critical information which is processed to extract useful information such as fuel efficiency, maintenance prediction, smart braking and speed control, and smart navigation. As a result, IoV generates lots of information which can cause unauthorized access, DoS attack, information and transmission security, etc. Blockchain can be used to provide following features to IoV:

- *Traffic safety and efficiency*: Blockchain can reliably transmit real-time vehicular information to its peer as well as to the centralized storage using cryptographic methods [17].
- *Micro-payment gateway service*: Using blockchain and smart contracts, payments of e-toll, e-parking, taxes, insurances, etc., can be made autonomously. Blockchain can be used to reliably made these transactions and hence allows fast and efficient IoV operation [18].
- *Vehicle insurance* and *audits*: IoV can collect car usage information and driver's driving tendencies to calculate premium for insurance. Blockchain can immutably and efficiently store these data to provide reliable audit results when queried by the insurance company [19].
- *Improved security*: Blockchain provides immutable and distributed storage. It can be used to reliably provide code updates, secure sensor data transmission, tolerance to data theft and hacking which results in efficient and resilient system [20].

4.3 Internet of Medical Things (IoMT) with Blockchain

IoMT [21] plays vital role now in healthcare. Sensors gives real-time reports of various body conditions as well predications, analysis, and reporting of medical data for future use. Medical data are very sensitive and private for any individual. Proper handling and transmission of medical data are required so that any data leakage or hacking can be prevented. Incorrect data or missing information can prove lethal for patients. So, IoT data need to be securely stored, managed, and disseminated. Blockchain can be used to secure IoMT and its data. Below are some benefits of using blockchain with IoMT:

- *Medical data management*: Smart contracts can restrict data access/share permissions along entities like doctor, hospital, research organizations, etc.
- *Medical data storage*: Blockchain can be used to provide secure data storage to store monitoring data of patient's conditions, activities or to monitor the development of a chronic disease over time. It will also enhance availability of data from malicious attacks.
- *Enable P2P applications*: Though healthcare data are generally stored in centralized databases, blockchain can guarantee secure decentralized storage service enabling P2P applications and systems that will improve the overall system.
- *Ease of data transformation*: Currently, data transformations are done by third parties in hidden and costly manner. Invested parties cannot know the process involved in data transformation which reduces safety, cost involved, regulations, and policies compliance. Blockchain with smart contracts can enable data transformation in transparent and cost-effective manner.

4.4 Internet of Battlefield Things (IoBT) with Blockchain

IoBT [22] is the amalgam of IoT with military applications, war machines, or human war-fighters. Future wars will be fought by machines on behalf of their human operators. Planning war strategies, extracting enemy secret information, enemy key assets surveillance, optimized soldier deployment, precise defensive planning, and performing accurate surgical strikes on enemy troops are some of the applications where IoBT will be very useful for militaries around the globe. Some of the benefits of using blockchain in securing IoBT are as follows:

- *Secrecy* and *authentication*: Blockchain stores data with keeping attributes like data integrity, data confidentiality, and data authentication. Once a command is issued by army leadership, it will not be possible to alter it by any external entities or it is not possible to be accessed by unauthorized persons and is only visible to properly authenticated parties.

- *Decentralization of data security*: Since data will be stored in a decentralized nature, it will be very difficult for enemies to attack or modify the data in comparison with centralized systems which will help in clear and coordinated attacks.

4.5 Green IoT with Blockchain

Green IoT [23] aims to drive toward eco-friendly sustainable development of IoT and its assisting technologies. IoT is exploding in its number due to affordable access to high-speed Internet and cheap production of microcontrollers. Current explosion of IoT devices is neither sustainable nor environment friendly. Hence, green IoT focusses on the reduction of energy consumption by IoT devices, reducing carbon footprint of IoT devices and using renewable energy sources in place of fossil fuels in powering IoT devices.

- *Green energy*: IoT can be used in smart grids to conserve energy and reduce wastage of energy [24]. Blockchain can enable secure transmission of critical information regarding power generation, power transmission and distribution, and power consumptions. Energy trading, electric vehicles, microgrid operations, and cyber-physical security are some areas of smart grids which can be hugely benefitted by the combination of blockchain with IoT.
- *Green Computing*: Cloud computing caters to our large demands of computing power and storage system by individuals or corporations [25]. But power consumptions of cloud computing or other similar types of computing are also creating major issue toward green goals. Green computing aims to introduce eco-friendliness along with sustainability to different computing types by efficiently using computing resources, minimizing energy uses, and reducing carbon emissions.
- *Green finance and e-governance*: Green finances enable efficient data availability, increased data sharing, fast access, and reduces cost to store or transmit data which overall optimize financial strategies that are built upon these data [26]. Cryptocurrencies and smart contracts can be used as a greener alternative to current conventional approach. Cryptocurrencies can exchange currencies or do international money transfer in real-time with no middlemen involved which saves additional energy, cost, and time. Smart contracts will reduce wastage by automating purchase orders, order monitoring or service deliveries, etc. E-governance can highly benefit from blockchain-IoT integration by eliminating excess paperwork, reduced public service delivery time, and minimizing manual interventions.

4.6 Future Research Issues

IoT has a huge role to play in future whether in civilian issues (Internet of vehicles, Internet of energy, etc.) or military issues (Internet of drones, Internet of battlefield things, etc.). IoT will continue to provide interactive smart services but the research enabling IoT lacks properly secured data communication, efficient data aggregation, access controls, etc. Thus, it is the need of the hour to increase the research focus of researchers on securing IoT. Some of the future research issues that should be worked upon immediately are as follows:

- *Lack of datasets*: There is a general lack of IoT dataset. IoT devices produce heterogeneous data which are often controlled by their device manufactures. Open public datasets are required to further analyze and optimize IoT device activities.
- *Merging artificial intelligence (AI) with IoT and blockchain*: Artificial intelligence (AI) has opened new avenues in automating services to consumers. But research in the field of AI with blockchain and IoT are very lacking. AI can easily detect fraudulent activities, prevent hacking and tampering activities, make sound decisions on behalf of its users, provide automation services, predict possible disease from patient history [27], etc., into the IoT network and blockchain.
- *Prediction of IoT device security*: Hospitals, assisted driving, battlefield surveillance, etc., are just some areas where physical security of IoT device is highly critical to the human lives. IoT devices can be destroyed whether by malicious intent or due to constant usage over time. Either case requires that security of IoT device can be made well in advance to prevent an unknown accident/loss.
- *Optimization of blockchains and blockchain-based platforms*: Lots of work has been done to improve the efficiency of blockchain but still enterprises do not wish to run their own blockchains due to high-energy costs and operational cost. For wide adoption of blockchain by enterprises and to create an optimized solution for IoT, research is further needed in optimizing blockchains in terms of energy efficiency, computational needs, and memory requirements.
- *Efficient* and *lightweight blockchain-based security solutions*: Blockchains are too computationally heavy to be run by the IoT device themselves. Hence, IoT devices cannot secure themselves from tampering, abusing, using anti-forensic techniques, etc., from a malicious third-party by using blockchains by themselves. Hence, blockchain solutions that can help in determining the security status and the information leakage during hacking process to investigators are needed to catch the culprit involved and prevent such activities in future.
- *Scalability* and *cost-efficiency*: Blockchain implementation should be scalable and cost-effective for huge IoT networks. IoT networks are dynamic and rapidly changing. Consensus algorithms should also be elastic enough to accept rapid changes to IoT network. Some of the current consensus algorithms are scalable compared to others but they lack in terms of security provided.
- *Uniform* and *interoperable security system/protocols*: IoT devices are heterogeneous in nature which causes issues in standards development and interoperability. A defined set of protocols or rules should be developed in conjunction

with blockchain which should declare a strict framework that dictates software and hardware compliance. Framework should be multi-layered like OSI. Layer-wise classification will define roles to specific layers and determines proper security level and transmission protocols before starting any data transmission within devices.

5 Conclusion

IoT and blockchain have highly steered the modern world toward smart and data driven future. Gone are the days when human potentials are destroyed in performing trivial, repetitive tasks. IoT devices are characterized by their ability to perform redundant tasks easily, and for intelligent tasks, it automatically senses some data using sensors and performs the task based on the sensed data. But IoT device cannot guarantee the security of sensed data or its transmission to centralized storage or other nodes. Blockchain, a distributed ledger with cryptography at its helm aims to provide security services to IoT. Blockchain with IoT can guarantee security, auditability, and incorruptibility while performing autonomous tasks. The IoT, blockchain, and AI combined will become the building blocks of the futuristic smart cities, smart healthcare, smart transportation, etc.

References

1. "The Growth in Connected IoT Devices Is Expected to Generate 79.4ZB of Data in 2025, According to a New IDC Forecast."
2. "Smart Homes In The U.S. Becoming More Common, But Still Face Challenges | S&P Global Market Intelligence."
3. "Smart Farming in 2020: IoT Sensors & Precision Agriculture - Business Insider."
4. "Future Of Retail 2020: Retailers Test Consumer-Facing And Operational IoT Solutions."
5. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System."
6. D. Lee, "Computer Systems Established, Maintained and Trusted by Mutually Suspicious Groups," (1982)
7. S. Haber, W.S. Stornetta, How to time-stamp a digital document. J. Cryptol. **3**(2), 99–111 (1991)
8. D. Bayer, S. Haber, W. S. Stornetta, "Improving the efficiency and reliability of digital time-stamping," in *Sequences II*, Springer New York, pp. 329–334 (1993)
9. A. Back, "Hashcash-A Denial of Service Counter-Measure," 2002.
10. L. Lamport, R. Shostak, M. Pease, The Byzantine generals problem. ACM Trans. Program. Lang. Syst. (TOPLAS) **4**, 382–401 (1982)
11. S. King, S. Nadal, "PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake," (2012)
12. M. Castro, B. Liskov, "Practical Byzantine Fault Tolerance," (1999)
13. "Delegated proof of stake - Bitcoin Wiki."
14. M.A. Khan, K. Salah, IoT security: review, blockchain solutions, and open challenges. Futur. Gener. Comput. Syst. **82**, 395–411 (2018)
15. X. Zhu, Y. Badr, "Identity management systems for the internet of things: a survey towards blockchain solutions," Sensors (Basel, Switzerland) **18**(12) (2018)

16. L. Zhang et al., Blockchain based secure data sharing system for internet of vehicles: a position paper. Vehicular Commun. **16**, 85–93 (Apr. 2019)
17. R.A. Sanchez, Acosta Rodriguez, W.M. Prieto, "Control of vehicular traffic through the IoV," Contemp. Eng. Sci., **10**(33), 1643–1650 (2017)
18. A. Ensor, S. Schefer-Wenzl, I. Miladinovic, "Blockchains for IoT payments: a survey," (2019)
19. R. Jabbar, M. Kharbeche, K. Al-Khalifa, M. Krichen, A.K. Barkaoui, Blockchain for the internet of vehicles: a decentralized IoT solution for vehicles communication using ethereum. Sensors (Switzerland) **20**(14), 1–27 (2020)
20. Y.R. Chen, J.R. Sha, Z.H. Zhou, "IOV privacy protection system based on double-layered chains," Wireless Commun. Mobile Comput., **2019** (2019)
21. J.M. Roman-Belmonte, H. de la Corte-Rodriguez, E.C. Rodriguez-Merchan, "How blockchain technology can change medicine," *Postgraduate Medicine*, vol. 130(4). Taylor and Francis Inc., pp. 420–427 (2018)
22. A. Kott, A. Swami, B.J. West, The internet of battle things. Computer **49**(12), 70–75 (2016)
23. F.K. Shaikh, S. Zeadally, E. Exposito, Enabling technologies for green internet of things. IEEE Syst. J. **11**(2), 983–994 (2017)
24. T. Alladi, V. Chamola, J.J.P.C. Rodrigues, S.A. Kozlov, Blockchain in smart grids: a review on different use cases. Sensors (Switzerland) **19**(22), 1–25 (2019)
25. P.K. Sharma, M.Y. Chen, J.H. Park, A software defined fog node based distributed blockchain cloud architecture for IoT. IEEE Access **6**, 115–124 (2018)
26. K. Anwar, J. Siddiqui, S. Saquib Sohail, "Machine learning techniques for book recommendation: an overview," SSRN Electron. J., (2019)
27. M. Atif, J. Siddiqui, F. Talib, "An overview of diabetes mellitus prediction through machine learning approaches," in *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom),* pp. 1145–1150 (2019)

Fuzzy C-Means Clustering of Network for Multi Mobile Agent Itinerary Planning

Nidhi and Shuchita Upadhyaya

Abstract Mobile agent (MA) works potentially efficiently in reducing network bandwidth consumption for distributed computing. In an MA-based system, a small-sized processing code is transmitted through the network rather than the raw data. Using strong migration capability, MA's processing code is being executed at each targeted node. Subsequently, useful and significant information is delivered to the intended user. Despite its benefit, MA has its issues also. Finding the appropriate number of MAs to be dispatched, the set of targeted nodes and their sequence of migration are the major issues related to multi-mobile agent itinerary planning (MIP). This paper tries to find a suitable number of MAs by considering the size of the data payload of a single MA and the load to be carried out by the MAs from the whole network. Moreover, this paper gives an idea of partitioning the given network into load balanced and nonoverlapping clusters using a fuzzy c-means clustering algorithm.

Keywords Clustering · Itinerary planning · Mobile agent

1 Introduction

A mobile agent is an autonomous and intelligent application program that is used to perform any data processing task on behalf of a remote user. The processing element (PE) can deploy one or more mobile agents to complete the assigned task. MA migrates to each targeted node locally and processes the data to gain some significant information [1–4]. After processing the data, aggregation of processed data will be performed. Thus, only significant accumulated information is transmitted to PE. On behalf of this retrieved information, a good decision can be made by the remote user. In other words, instead of transmission of a large amount of raw data; MA (a small-sized source code) can be transferred to each targeted node to process the data toward the completion of the assigned task [4–7]. This progressively processed and aggregated data is carried out by the MA in its data payload. Further, MA returns to

Nidhi (✉) · S. Upadhyaya
Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, India
e-mail: nidhikashyap@kuk.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_55

deliver this retrieved information to the PE to unload itself. Hence, only small-sized source code and retrieved significant information are getting transmitted through the network. Therefore, it will save large consumption of network bandwidth.

Besides, MA has one more exceptional characteristic of strong mobility that it can work even after disconnection also. During processing, it does not need any network connection. After processing when it will get connected it will be migrated to the next node and resume itself. In conclusion, MA needs an Internet connection only at the time of migration. In contrast, it does not need an Internet connection during the processing of data.

Moreover, when a single MA becomes insufficient for the given network then multiple mobile agents can be dispatched by the PE. Each MA has to migrate to the targeted set of nodes toward the completion of its task. Their route of migration to the assigned set of nodes is called itinerary planning. Whenever a single MA has been deployed to perform the task, the designed itinerary is called single mobile agent itinerary planning (SIP) [1, 8, 9]. Large delay and reliability are the most prominent issues related to SIP. In contrast, whenever more than one MA is dispatched, then a planned itinerary is called multi-mobile agent itinerary planning (MIP), in which, all mobile agents have their own set of nodes, and relative to each set they have separate itinerary planning. MIP came into existence to overcome the aforementioned issues of SIP. Despite its benefits, MIP leads to several issues also. Respective research has been done to resolve these specific issues. This paper suggests an algorithm to partition the network into k number of disjoint domains. Moreover, by considering the threshold value of a single MA, the appropriate number of MAs for the respective domain can be determined.

2 Related Work

Several algorithms have been proposed by the researchers to cluster the network for multi-mobile agent itinerary planning. They dealt with this identified problem in their way by keeping mobile agent routing into consideration. Kuila [10–13] suggested algorithms are using k-means and x-means clustering algorithms. These algorithms are very useful for large-scale problems. After the choice of cluster heads (CHs), nodes are associated with the nearest CH. Further, a single MA is deployed to each cluster to perform the assigned task.

Mpitziopoulos [4, 6, 9, 14, 15] algorithms use a tree-based approach to distribute their network into clusters. A dedicated single MA is responsible for each stemmed branch. Tree-based clustering is not useful for time-critical problems. Chou [8, 16–18] algorithms use Genetic algorithm (GA)-based evolutionary technique. Due to the initialization of the population on a random basis, GA-based MIP is not useful for time-based applications. [9] Second Near Optimal Itinerary Design (SNOID), [2] Tree-Based Itinerary Design (TBID) distributes the network into concentric elliptical or circular zones. However, not all sensor nodes need to be scattered in any geographical shape.

Chen [1] Central Location-based MIP (CL-MIP), [3] Energy Efficient MIP (EMIP) partitions their network by considering visiting center location (VCL) as the center of, clusters to be made. VCL can be taken any node which is situated at the center of the dense area. Source Grouping-based MIP (SG-MIP), [5] Directional Source Grouping-based MIP (DSG-MIP), [19] Angle Gap-based MIP (AG-MIP) distributes their network into directional sector zones by considering an angle θ. However, these algorithms are limited in determining the angle θ.

A lot of research has been dealt with the issue of partitioning the network through heuristic approaches, optimization algorithms, or evolutionary techniques. Most weaknesses and the robustness of aforesaid algorithms have been discussed. To get the suitable partitioning of the network into domains, soft computing is the best technique. Using a fuzzy C-means clustering algorithm, an individual sensor node can be associated with the appropriate domain. This algorithm states that each node has some degree of belongingness to each domain center (DC). But the node will be associated with whom it has a higher degree of membership. The proposed algorithm uses this technique to make clustering into the number of disjoint and load balanced domains by keeping the expected estimated assigned load in advance into consideration. Through the suggested algorithm, a node is associated with the DC with maximum membership value. In addition, if a single node has the same membership value with more DCs, then it will be connected with the DC having the least expected estimated assigned load so that each domain would get load balanced.

3 Proposed Method

In the proposed clustering algorithm, a network having n number of nodes is to be partitioned into k number of domains (clusters). Each domain has a single domain center (cluster head). For this, suitable nodes have to be chosen to represent domain centers (DCs). The remaining nodes of the network are to be associated with appropriate DC to form disjoint domains.

While clustering, cluster overlapping and load balancing are the most important issues on which research can be carried out. In the proposed algorithm both issues have been resolved. In crisp or traditional partitioning technologies, each node has the characteristic value of either the node belonging to any particular domain or not. However, using fuzzy logic, the degree of belongingness of each node relative to each domain center can be determined. This degree of belongingness is called degree of membership (μ). Afterward, the ith node is assigned to jth DC by considering, $\mu_{i,j} \in [0, 1]$, which has been described through Eq. 1. Where $\mu_{i,j}$ can be determined through Eq. 2. This value is inversely proportional to the distance value. If the distance is small, then it tends to greater membership value and vice versa.

$$\mu i, j = \begin{bmatrix} 1 & \text{if } i = j \\ \frac{d_{i,j}^{\frac{-2}{m-1}}}{\sum_{l=1}^{k} d_{1,j}^{\frac{-2}{m-1}}} & \text{Otherwise} \end{bmatrix} \forall i \in N, l \neq j \tag{1}$$

$$\mu i, j = \frac{d_{i,j}^{\frac{-2}{m-1}}}{\sum_{l=1}^{k} d_{i,j}^{\frac{-2}{m-1}}} \tag{2}$$

where k is the number of domains in which the whole network is to be partitioned.

'm' is the weighing exponent which controls the fuzziness or crispness of the clustering. This 'm' decides the boundaries of the domain as either softer or harder.

Equation 3, specifies that $\mu_{i,j}$ of each source node corresponding to every DC is always 1. It ensures that each source node is assigned to exactly one domain center.

$$\sum_{i=1}^{k} \mu_{i,j} = 1, \ \forall j \in \{1, 2, 3, \ldots\ldots, n\} \tag{3}$$

Equation 4; every node has some membership value w. r. t. each domain. In other words, it assures that each domain has some expected assigned load relative to source nodes in the network, otherwise it needs to be dropped out.

$$\sum_{i=1}^{n} \mu_{i,j} > 0, \ \forall j \in \{1, 2, 3 \ldots\ldots, k\} \tag{4}$$

A brief description of the clustering algorithm as applied to the network for determining candidate DCs and assignment of the rest of the nodes to appropriate DC is given below.

3.1 Problem Statement

Given:

1. $N = \{N\, i \in i = 1, 2, \ldots\ldots, n\}$—Indexed set of n nodes.
2. $D = \left\{d_{i,j}, i = 1, 2, \ldots\ldots, n;\ j = 1, 2, \ldots\ldots, n; i \neq j\right\}$—Table 1: distance table, where d_{ij} is the Euclidean distance between node I and node j and is specified as integer.
3. Maximum number of domains is defined as k where $k < n$.

Determine:

Selection of k nodes to represent as a set DC (candidate domain centers) and then partitioning the network into k equals loaded domains with each domain having a single DC (element of set DC).

To be determined:

Exhaustive partitions of the set of nodes (N_i) into c number of domains.

Table 1 Euclidean distance matrix for each node

$D_{i,j}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	3	6	3	2	8	8	6	8	11	11	12	12	10	12
2	3	0	3	2	3	7	8	6	8	12	12	12	12	10	12
3	6	3	0	3	6	4	6	6	7	10	10	10	10	8	10
4	3	2	3	0	3	5	6	4	6	10	10	10	9	8	10
5	2	3	6	3	0	8	6	4	6	10	10	10	10	8	10
6	8	7	4	5	8	0	5	8	9	5	5	5	7	9	11
7	8	8	6	6	6	5	0	3	4	6	4	4	4	4	6
8	6	6	6	4	4	8	3	0	2	9	7	6	6	4	6
9	8	8	7	6	6	9	4	2	0	10	8	6	4	2	4
10	11	12	10	10	10	5	6	9	10	0	2	4	8	10	12
11	11	12	10	10	10	5	4	7	8	2	0	2	6	8	10
12	12	12	10	10	10	5	4	6	6	4	2	0	4	6	8
13	12	12	10	9	10	7	4	6	4	8	6	4	0	2	4
14	10	10	8	8	8	9	4	4	2	10	8	6	2	0	2
15	12	12	10	10	10	11	6	6	4	12	10	8	4	2	0

FCM Clustering Algorithm

Some abbreviations which are used in the given algorithm:

N_i—Set of nodes.
n_i—element of set N_i.
DC— Set of domain centers.
dc_j—element of set DC_j.

1. Given: Network with n number of nodes. $N_i = \{1, 2, 3, \ldots\ldots n\}$;
2. Determine membership values of each sensor node *w. r. t.* rest $n-1$ nodes;

   /* Step 3 is evaluated to choose k number of candidate domain centers.
3. Choose top k number of nodes from the original set (N_i) of nodes such that $\sum_{j=1}^{n} \mu_{i,j} : i \neq j$; is maximum; denoted by set DC_j, where $j = 1$ to k, set $N_i = N_i - DC_j$;

   /* Steps 4 and 5 will be evaluated to associate each node to the relative candidate domain center.
4. Determine membership value of each node n_i relative to each dc_j, i.e., $\mu_{i,j} : \forall i \in N_i, \forall j \in DC_j$;
5. Associate each $i \in N_i$, to relative $j \in DC_j$ such that n_i will be associated to dc_j with whom it has maximum membership value. If for any $i \in N_i, \exists j \in DC_j : \mu_{i,j}$ is equal then n_i will be associated to dc_j, the whose expected estimated load is least;

Fig. 1 Illustration network of fifteen nodes

6. Afterward, if there exists any dc_j, whose load to be carried out is less than the threshold value.[1] Then this node, as well the associated nodes need to be connected to some other appropriate dc_j.

4 Instance Formulation of the Network with FCM Clustering Algorithm

For instance, a network of 15 nodes has been depicted in Fig. 1. Furthermore, distance of each node relative to other node is given by Table 1.

The motive of this algorithm is to make domains around DCs. Furthermore, these DCs should be situated at the center of the densest places in the network. From the below, DCs using Eq. 2 can be determined. Uppermost k number of nodes, where $k = 4$, nodes having maximum $\sum_{j=1}^{n} \mu_{i,j} \forall i \in N_i$.

In particular, using above Table 2 selected candidate DCs are DC_4, DC_7, DC_8 and DC_9 with their respective $\mu_{i,j}$ are 1.14, 1.11, 1.12and 1.12. In below Table 3, membership value of each node w. r. t. each candidate DC, using Eq. 2 has been determined.

[1] Threshold value of a MA is the minimum number of nodes for which a single MA may be deployed. For the illustrated network, this threshold value has to be taken 2 nodes.

Table 2 Membership values of each node w. r. t. other nodes in the network

$D_{i,j}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0.14	0.07	0.14	0.21	0.05	0.05	0.07	0.05	0.04	0.04	0.03	0.03	0.04	0.03
2	0.13	0	0.13	0.19	0.13	0.06	0.05	0.06	0.05	0.03	0.03	0.03	0.03	0.04	0.03
3	0.07	0.14	0	0.14	0.07	0.11	0.07	0.07	0.06	0.04	0.04	0.04	0.04	0.05	0.04
4	0.11	0.17	0.11	0	0.11	0.07	0.06	0.09	0.06	0.03	0.03	0.03	0.04	0.04	0.03
5	0.19	0.13	0.06	0.13	0	0.05	0.06	0.09	0.06	0.04	0.04	0.04	0.04	0.05	0.04
6	0.06	0.06	0.11	0.09	0.06	0	0.09	0.06	0.05	0.09	0.09	0.09	0.06	0.05	0.04
7	0.04	0.04	0.06	0.06	0.06	0.07	0	0.12	0.09	0.06	0.09	0.09	0.09	0.09	0.06
8	0.06	0.06	0.06	0.08	0.08	0.04	0.11	0	0.17	0.04	0.05	0.06	0.06	0.08	0.06
9	0.04	0.04	0.05	0.06	0.06	0.04	0.08	0.17	0	0.03	0.04	0.06	0.08	0.17	0.08
10	0.04	0.04	0.05	0.05	0.05	0.09	0.08	0.05	0.05	0	0.24	0.12	0.06	0.05	0.04
11	0.04	0.03	0.04	0.04	0.04	0.08	0.1	0.06	0.05	0.19	0	0.19	0.06	0.05	0.04
12	0.03	0.03	0.04	0.04	0.04	0.08	0.1	0.07	0.07	0.1	0.2	0	0.1	0.07	0.05
13	0.03	0.03	0.04	0.04	0.04	0.06	0.1	0.06	0.1	0.05	0.06	0.1	0	0.19	0.1
14	0.03	0.03	0.04	0.04	0.04	0.04	0.08	0.08	0.16	0.03	0.04	0.05	0.16	0	0.16
15	0.04	0.04	0.05	0.05	0.05	0.04	0.08	0.08	0.11	0.04	0.05	0.06	0.11	0.23	0

Table 3 Membership values of each node with respect to the domain centers

Node N_i	Membership value w.r.t. DC_4	Membership value w.r.t. DC_7	Membership value w.r.t. DC_8	Membership value w.r.t. DC_9
1	0.4	0.2	0.2	0.2
2	0.5	0.1	0.2	0.1
3	0.4	0.2	0.2	0.2
5	0.4	0.2	0.3	0.2
6	0.3	0.3	0.2	0.2
10	0.2	0.3	0.2	0.2
11	0.2	0.4	0.2	0.2
12	0.1	0.4	0.2	0.2
13	0.1	0.3	0.2	0.3
14	0.1	0.2	0.2	0.4
15	0.1	0.2	0.2	0.4
$\sum \mu_{i,j}$	3.0	2.9	2.5	2.6

The expected assigned load for DC_4, DC_7, DC_8 and DC_9 are 3.0, 2.9, 2.5 and 2.6, respectively. Then, nodes have been associated with appropriate DC. The node has been connected to the DC with whom it has maximum membership value. Further, if any node had the same membership value with 2 or more DCs then the node has been connected to the least loaded DC. In the below Table 4, the association of all

Table 4 Assignment of successive nodes to the domain centers

Assignment of successive nodes to the domain centers				
Nodes with possible DCs	DC_4 (assigned load (3), associated nodes)	DC_7 (assigned load (2.9), associated nodes)	DC_8 (assigned load (2.5), associated nodes)	DC_9 (assigned load (2.6), associated nodes)
1	3, {1}	2.7,{}	2.2, {}	2.5, {}
2	3, {1, 2}	2.6, {}	2.1, {}	2.3, {}
3	3, {1, 2, 3}	2.4, {}	1.9, {}	2.1, {}
5	3, {1, 2, 3, 5}	2.2, {}	1.6, {}	2, {}
6	2.7, {1, 2, 3, 5}	2.2, {6}	1.4, {}	1.8, {}
10	2.5, {1, 2, 3, 5}	2.2, {6, 10}	1.2, {}	1.6, {}
11	2.3, {1, 2, 3, 5}	2.2, {6, 10, 11}	0.9, {}	1.4, {}
12	2.2, {1, 2, 3, 5}	2.2, {6, 10, 11, 12}	0.7, {}	1.1, {}
13	2.0, {1, 2, 3, 5}	1.9, {6, 10, 11, 12}	0.5, {}	1.1, {13}
14	1.9, {1, 2, 3, 5}	1.7, {6, 10, 11, 12}	0.2, {}	1.1, {13, 14}
15	1.8, {1, 2, 3, 5}	1.4, {6, 10, 11, 12}	0, {}	1.1, {13, 14, 15}

nodes has been depicted. There is a single candidate domain center, DC_8 having an estimated expected load of 2.5 but none of the nodes has been connected to it. The load carried out by this node is zero, which is less than the threshold value. In particular, DC_8 needs to be dropped out as candidate DC. Also, it needs to be connected to some other DCs. Membership value of N_8 w. r. t. DC_4, DC_7 and DC_9 can be determined again using the Eq. 2. Besides, its respective membership values are 0.2, 0.3 and 0.5. Hence, N_8 will be connected to DC_9.

Finally, set of nodes assigned to DC_4, DC_7 and DC_9 are {1, 2, 3, 5}, {6, 10, 11, 12} and {8, 13, 14, 15} with their respective loads 1.8, 1.4 and 1.6. Hence, all nodes are mostly equally loaded containing same number of nodes. Therefore, network of 15 nodes has been partitioned into $c = 3$, number of disjoint domains {1, 2, 3, 4, 5}, {6, 7, 10, 11, 12} and {8, 9, 13, 14, 15} which are significantly load balanced.

5 Conclusion

This paper has presented an algorithm for partitioning the network into a required number of domains for multiple mobile agent itinerary planning. This paper presented the idea that the domain whose data is to be carried out by the MA is less than the threshold payload required to drop out. Therefore, by considering the threshold value adaptively for a single mobile agent, the decision of deploying an appropriate number of mobile agents for each domain can be taken. Using fuzzy c-means clustering each node is associated with the domain for which it has the largest membership

value and is assigned to exactly one domain. Thus, each domain becomes non-overlapped. Hence, MA visits each targeted node once. Therefore, it incurs less network bandwidth and reduces energy consumption. Also, it tries to balance each domain. Consequently, the proposed algorithm is a disjoint and balanced clustering algorithm so that an appropriate number of MAs can be deployed to an MA-based system.

References

1. M Chen, S Gonzalez, Y Zhang, V Leung, "Multi-agent itinerary planning for wireless sensor networks." In: International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, (Berlin, Heidelberg, 2009)
2. D Gavalas, C Konstantopoulos, B Mamalis, G Pantziou, "New techniques for incremental data fusion in distributed sensor networks," in *Proceedings of the 11th Panhellenic Conference on Informatics (PCI'2007)* (2007)
3. J Wang, Y Zhang, Z Cheng, X Zhu, "EMIP: energy-efficient itinerary planning for multiple mobile agents in wireless sensor network." Telecommun. Syst., **62**(1), 93–100 (2016)
4. A Mpitziopoulos, D Gavalas, C Konstantopoulos, "Deriving efficient mobile agent routes in wireless sensor networks with NOID algorithm," in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications* (2007)
5. M Chen, S Gonzalez-Valenzuela, VCM Leung, "Directional source grouping for multi-agent itinerary planning in wireless sensor networks," in *International Conference on Information and Communication Technology Convergence (ICTC)* (2012)
6. D Gavalas, IE Venetis, C Konstantopoulos, "Energy-efficient multiple itinerary planning for mobile agents-based data aggregation in WSNs," Telecommun. Syst., **63**(4), 531–545 (2016)
7. K Lingaraj, RV Biradar, VC Patil, "OMMIP: an optimized multiple mobile agents itinerary planning for wireless sensor networks." J. Inf. Optim. Sci., **38**(6), 1067–1076 (2017)
8. YC Chou, M Nakajima, "A clonal selection algorithm for energy-efficient mobile agent itinerary planning in wireless sensor networks." Mobile Netw. Appl., **23**(5), 1233–1246 (2018)
9. C Konstantopoulos, A Mpitziopoulos, D Gavalas, "Effective determination of mobile agent itineraries for data aggregation on sensor networks." In *IEEE Transactions on Knowledge and Data Engineering* (2009)
10. P Kuila, SK Gupta, PK Jana, "A novel evolutionary approach for load balanced clustering problem for wireless sensor networks." Swarm Evol. Comput., **12**, 48–56 (2013)
11. I Aloui, O Kazar, L Kahloul, "A new itinerary planning approach among multiple mobile agents in wireless sensor networks (WSN) to reduce energy consumption." Int. J. Commun. Netw. Inf. Secur. (IJCNIS) **7**(2), 116 (2015)
12. I Aloui, O Kazar, L Kahloul, A Aissaoui, "A new" data size" based algorithm for itinerary planning among mobile agents in wireless sensor networks." In *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies* (2016)
13. HQ Qadori, ZA Zulkarnain, ZM Hanapi, S Subramaniam, "A spawn mobile agent itinerary planning approach for energy-efficient data gathering in wireless sensor networks." Sensors **17**(6), 1280–1296 (2017)
14. B Liu, J Cao, J Yin, W Yu, B Liu, X Fu, "Disjoint multi mobile agent itinerary planning for big data analytics." J. Wireless Commun. Netw. **1**, 1–12 (2016)
15. M Chen, W Cai, S Gonzalez, V Leung, "Balanced itinerary planning for multiple mobile agents in wireless sensor networks." In *International Conference on Ad Hoc Networks*. (Berlin, Heidelberg, 2010)
16. Q Wu, NSV Rao, J Barhen, SS Iyengar, "On computing mobile agent routes for data fusion in distributed sensor networks." In *IEEE Transactions on Knowledge and Data Engineering* (2004)

17. R Rajagopalan, CK Mohan, P Varshney, "Multi-objective mobile agent routing in wireless sensor networks." In *IEEE Congress on Evolutionary Computation* (2005)
18. W Cai, M Chen, T Hara, L Shu, T Kwon, "A genetic algorithm approach to multi-agent itinerary planning in wireless sensor networks." Mobile Netw. Appl. **16(**6), 782–793 (2011)
19. W Cai, M Chen, X Wang, "Angle gap (AG) based grouping algorithm for multi-mobile agents itinerary planning in wireless sensor networks." (2009)
20. JV De Oliveira, W Pedrycz, Advances in Fuzzy Clustering and its Applications. (Chichester, Wiley, 2007)
21. Pratyay Kuila, Prasanta K. Jana, Clustering and Routing Algorithms For Wireless Sensor Networks: Energy Efficient Approaches. (London: Taylor and Francis Group, 2018)
22. Bo Yuan George, J. Klir, Fuzzy Sets and Fuzzy Logic Theory and Applications (New Delhi, Prentice Hall of India, 2005)

Acceptance of Eco-Friendly Substitute of White Skim Coat (Wall Putty) in Mumbai

Mandar Anil Chitre, Shivoham Singh, and Manuj Joshi

Abstract Wall putty available in the market is white in colour and made from perishable raw materials. It is possible to have a substitute product, black in colour and made from using waste powders from quarry and construction. Replacement of the current white putty with this substitute black putty is technically possible. Our research is on the acceptance of this product. On conducting a survey in Mumbai and surrounding suburbs, we asked questions on the awareness, acceptance of self and perceived acceptance of the industry. The results are compiled and analyze. The resistance shown by the various influencers of the industry is significantly on the higher side. However, as researchers we feel that there is certainly a possibility to make this change and develop the market in the direction of usage of black putty which is actually an eco-friendly product using waste materials. Another important benefit of using this product is that it actually reduces pollution. The waste product used as raw material is very fine, and these particles create air pollution. We have also studied and worked out the cost of this substitute product which is actually cheaper than the current white putty. The research has further scope to be carried out in other areas of the country and interacting with more population.

Keywords Building materials · Wall putty · Waste management · Cement mortars · Wall finishes

1 Introduction

As defined by Wikipedia; putty is a material with high plasticity, similar in texture to clay or dough, typically used in domestic construction and repair as a sealant or filler. Although some types of putty (typically those using linseed oil) slowly polymerize and become stiff, many putties can be reworked indefinitely, in contrast to other

M. A. Chitre · S. Singh · M. Joshi (✉)
Pacific Academy of Higher Education and Research Society, Udaipur, Rajasthan, India
e-mail: manujjoshi@gmail.com

M. A. Chitre
e-mail: mandar.chitre@batonconsultants.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_56

Fig. 1 Layers of a wall as per SCP

types of filler which typically set solid relatively rapidly. It is a material which is used regularly in construction of walls, ceilings, and other surfaces. After plastering, the wall is not smooth finished or not on "zero-zero" level. If painting is done directly after plaster, it will have two effects.

1. The finish of the paint will be very poor because of the undulations in the wall.
2. The consumption of paint and primer will be very high and un-economical for construction.

Wall putty is the material which solves this problem. This fine paste fills up the voids and makes the wall as smooth as a mirror. Thus, the primer + paint consumption is reduced, and a good surface radiates an excellent finish.

As illustrated in Fig. 1, the wall has layers. Each layer has its own significance, and the wall putty is applied after the plaster and before the primer.

The wall putty is a very important material in construction. It is used on both internal and external walls. This paste can also be used as a filler material in those applications which do not demand water proofing or strength. This is because the wall putty is merely a filler paste. It is not a water proofing material which can hold water from passing or prevent it from penetration. The putty itself does not have a strong compressive strength and therefore will not last if applied to accept any load which is compressive, flexural, or shear. Thus, to conclude, we will say that the wall putty is purely a filler paste which is used to make the wall "zero-zero" removing undulations and make it ready to accept the next layers.

Colour of Wall Putty:

In India, since the last two decades, wall putty has been white in colour. This has created a precedence that putty should always be white. In all other tropical countries where putty is used, (the name is skim coat) the colour is dark.

Fig. 2 Wall putty—raw materials

Table 1 Binder, filler, and enhancer

The binder: cement	The filler: dolomite	The enhancer: polymers
Is responsible for binding all material. It reacts with water and releases heat of hydration. The effect is that it becomes hard after drying	It gives the body to the product. This is an inert material which does not react and provides the mass required	These are chemicals which enhance the properties of the product, like adhesion, workability, and water retention

The Wall Putty Raw Materials:

There are three main raw materials of wall putty. Each component has its own role to play (Fig. 2 and Table 1).

White Cement—As defined by Wikipedia, a cement is a binder, a substance used for construction that sets, hardens, and adheres to other materials to bind them together.

The cement used for decorative purpose is white. The normal cement which is available and used in bulk is grey in colour. Cement is made from limestone. White cement availability is limited, because it is restricted to only those mines which produce limestone which is free from Fe_2O_3 and a few other minerals which impart the dark shade. A few mines in India have limestone which is suitable for producing white cement. Thus, white cement is limited, expensive, and to be used for decorative purpose.

Dolomite—It is a mineral available in abundance across regions of Rajasthan, MP, Chattisgarh, Maharashtra, AP, and Tamil Nadu. This yellowish rock becomes white in colour when it is fine grinded to a size of 300 mesh or 50 microns or 10–3 mm. This powder can be thoroughly mixed with white cement to achieve homogeneity of over 99%.

Polymers—Chemicals such as MHEC and RDP are added in very small proportions. These are the enhancers which react with water and cement to produce properties which help the product perform. Adhesion, workability, water retention, etc.,

Fig. 3 MHEC and RDP

are very important properties for the wall putty, and these polymers enable them. They are white in colour (Fig. 3).

2 Background

In context to our paper, we have considered six papers/thesis/articles, which were reviewed and the summary of it is as follows.

The first review is of researcher Kantharia Mohan, who has done a thesis on usage of polymer and nanomaterials in cement mortars. The researcher states the various admixtures which are added to the mortars to enhance the required properties. He also gives importance of durability and strength of the mortars. Our product is also a mortar and usage of such admixtures are considered for the same. Our research includes usage of waste materials and creates a substitute. This review was useful to us on the admixture and properties usage [1].

The second review conducted is on the usage of clay tile waste as a partial replacement of cement in the masonry mortar. Researcher Jiji Antony gives importance to usage of waste materials and its feasibility. This is a good review for us as we are also using waste materials [2].

The third review is on the use of lightweight filler materials. The objective is to improve the insulation properties. The researchers have shown the effects of the dosage of different components to optimize the performance of the final product [3].

The fourth review is about usage of activated recycled powder from construction and demolition waste. The research done by four researchers talks about use of hybrid recycled powder (HRP) collected from construction and demolition waste (C&DW). This has partial cement which is un-hydrated or not reacted. This is an important reference for us because we are also recommending usage of waste material which is inert and not reactive. This paper uses waste which is partially active [4].

The fifth review is on the new product as exterior wall putty. The product that we are proposing is for both internal as well as external. Researchers have highlighted the requirements of the exterior putty which are more and need to be tougher than the internal putty. Usage of certain pozzolanic materials enhance the properties

Fig. 4 2/3rd particles of Mumbai's pollutant are road dust and construction [7]

and achieve better results in workability, water resistance, high-bonding strength, and dynamic crack resistance. Article from industry magazine Drycotec Diaries published in January 2020, an expert of the industry Mr. Anantha Krishnan Subraveti mentions that colour of the putty does not matter and he does not consider it a customer complaint. This reference is important to us because we are proposing a black colour putty [5, 6].

Research Gap: No such study is conducted to identify the market potential.

Problem: About the current scenario in Mumbai (Fig. 4).

Pollution is rampantly increasing in Mumbai. Figure 3 by Times of India shows that 2/3rd particles of Mumbai's pollutant is road dust and construction. Manufactured sand is the major culprit. Construction quarries which produce rocks and sand generate a lot of dust. This dust is a fine powder of less than 50 microns and consists of inert rock. Its usage in construction is very limited because it contains deleterious material which affects the strength of the concrete and is harmful to the structural loading.

Consumption of White Putty:

Mumbai consumes approximately 20–25,000 metric tonnes of wall putty every month. This product is produced by mining of dolomite and white cement. Then, mixing all the components together after accurately weighing each one.

The Solution:

It is easily possible that we use this quarry dust (rock particles below 50 microns) in the wall putty. However, this wall putty will certainly not be white in colour, because a large component is this filler which will be black or grey in colour.

Table 2 Black putty is cheaper than the current white putty

White putty 40 kg bag—raw materials				
Material description	Function	Unit cost / kgs	Quantity (kgs)	Total cost
White cement	Binder	22	8	176
Dolomite	Filler	3	32	96
Chemicals	Enhancer	300	0.3	90
Total raw material cost				362.00
Black putty 40 kg bag—raw materials				
Material description	Function	Unit cost/kgs	Quantity (kgs)	Total cost
Grey cement	Binder	6.5	8	52
Quarry dust	Filler	0.3	32	9.6
Chemicals	Enhancer	300	0.3	90
Total raw material cost				151.60

Commercial Solution:

The product we have designed has the following combination and is a proven substitute of current wall putty which is white in colour.

Table 2 indicates that the black putty is cheaper than the current white putty. Manufacturers can produce 2.3 kgs of this dark grey putty in the price of 1 kg of white putty. This will help in reducing a small part of the cost of construction and eventually help the final buyer. We have tried and tested this product and are perfectly replaceable with the white putty.

Market References:

The products which are not white are currently available in the market and we wish to provide two such examples as below (Fig. 5).

One such, e.g. reference provided is of Dalmia magic putty. A product which is not white but slightly light red or pink in colour. This product is actually still struggling to perform in the market (Fig. 6).

Fig. 5 Dalmia magic putty colour light red or pink

Fig. 6 GoMix wall putty colour grey

The market has a few products in the grey colour zone but they are used in the semi-filler applications to cover up the errors of RCC or any other reason. Not on mass scale.

3 Methodology

The main objective of our research is to find the acceptance of eco-friendly product of black putty in Mumbai market.

To find if the market will accept this product, we decide to conduct the market research in Mumbai. This research would reveal if the market is in a position to accept this product or not.

Process:

Area sampling is done by us. The sample details are as in Fig. 7.

Data Collection—Interviews (scheduled).

The Wall Putty is sold to a distributor who operates in two categories, new & repair. The end user does not have any role to play. It is the buyer & influencer who control the market.

Fig. 7 Sample details

Considering the market of Mumbai, we have selected a list of 30 retailers and applicators (painters) who do business in the Mumbai and surrounding areas. Different areas were selected for considering the retailers for the sample.

The following questions were asked:

Q1. Are you aware that the wall putty is not white and is dark grey when used outside India in markets like Dubai and Singapore?
Q2. Are you aware that this product is not white and has the same features, quality, and functioning as the white putty?
Q3. If this product is offered to you, will you keep in your retail outlet and sell it to your customers?
Q4. Do you think the other retailers in the market will accept and start selling this product?
Q5. What according to you will be the reasons of success or failure of this product?

The short interview captured the reactions of the sample lot. The questions selected are direct to the point and adequate to gain the objective of this research. This paper is limited to studying the acceptance and reasons for the same before proceeding to the next level.

The territory distribution for collection of data was as below, coverage of the city is complete and the response generated would truly resemble the opinion of all the influencers of the city (Fig. 8).

Limitations:

Due to the pandemic situation and lockdowns, all interview work has been done over phone, and the size of sample has been limited to 30.

4 Results

The data compiled and analyzed give the following results (Fig. 9):

The awareness levels in the sample size were negative. Overall, they are not aware about such a product which exist in other markets (Fig. 10).

Since there is no awareness of the product. When explained most of them were not sure if it would work. Therefore, the confidence level of this product seems to be weak (Fig. 11).

In regards to their own acceptance of the product. Majority was on the negative side. Therefore, we can surely imply that the product will not be easily acceptable to the influencers (Fig. 12).

A strong perception in the selected sample about others not accepting this product. The confidence level felt by us as interviewers indicate that the problem is big. The product will definitely struggle to get an acceptance in the market (Fig. 13).

Maximum respondent said that it is the mindset which would be the major reason for rejection. The level of confidence displayed by them was very strong.

Locations	
Ambernath	Khopoli
Ambernath	Kurla
Andheri	Manpada
Andheri	Mira Road
Bhandup	Nalasopara
Byculla	Panvel
Chembur	Sanpada
Digha	Santa Cruz
Dombivali	Sion
Ghatkoper	Palghar
Ghatla	Thane
Jogeshwari	Turbhe
Kalyan	Ulhasnagar
Kausa	Vasai
Kharghar	Wadala

Fig. 8 Territory distribution

Fig. 9 Product awareness

5 Conclusion and Future Scope

It was really surprising to understand that the reason for rejection was not the colour. They all accepted that it is the mindset which would need to change. The respondents

Fig. 10 Product benefits

Fig. 11 Product acceptance—self

agree that the colour would be an issue but they also understand that primer will be able to cover it. This particular observation makes us believe that there is scope for this product to get established in the market and it has a future.

To Conclude:

- Acceptance of this product in the market is possible but challenging
- Work will have to be done by the manufacturer wants to launch this product
- Focus on education and concept selling will enable the change

Scope for Future Work:

- This research was done in Mumbai and suburbs
- Other parts of the country can be targeted for future research

Fig. 12 Product acceptance—perception of others

Fig. 13 Product acceptance—reason of decision

- The researcher could get a different report in the rural areas where price sensitivity is higher.
- The product could be introduced in rural areas and then gradually moved for acceptance in semi-urban and urban markets.

References

1. M. Kantharia, *Strength Investigation of Cement Mortar Using Polymer And Nano Materials*, Amity University M.P. (2020), http://hdl.handle.net/10603/309937
2. J. Antony, Investigation of clay tile waste as partial replacement of cement in masonry mortar and building blocks. Cochin University of Science and Technology (2019), http://hdl.handle.net/10603/309913
3. M. Kupinski, K. Stobieniecka, K. Skowera, Influence of Lightweight Fillers on the performance of cement based skim coat. Struct. Environ. **12**, 5–11 (2020). https://doi.org/10.30540/sae-2020-001
4. T. Meng, Y. Hong, K. Ying, Z. Wang, Comparison of technical properties of cement pastes with different activated recycled powder from construction and demolition waste (2021), https://doi.org/10.1016/j.cemconcomp.2021.104065
5. G.L. Zhang, L.W. Mo, J.Y. Sun, J.B. Chen, J.Z. Liu, Preparation and experimental study on a new type of exterior wall putty. Adv. Mater. Res. **671–674**, 1914–1917 (2013). https://doi.org/10.4028/www.scientific.net/amr.671-674.1914Researchers
6. A.K. Subraveti, *Issues and Possible Reasons in the Application of Wall Putty, Drycotec Diaries* (2020), www.drycotec.com/magazines
7. https://timesofindia.indiatimes.com/city/mumbai/road-dust-makes-up-2/3rds-of-particles-in-mumbais-air/articleshow/81300796.cmsvisited on 07 Jun 2021

Economics of Immutability Preserving Streaming Healthcare Data Storage Using Aggregation in Blockchain Technology

Sachin Gupta and Babita Yadav

Abstract The contemporary technology landscape is big data oriented and driven by analytics. The paradigm shift fueled by rapid data generation and higher storage capabilities has resulted in a massive transition across data storage technologies. Traditional database management systems being used in record-based transactional applications are paving the way for new rapid engagement-based data stores leading the newer applications. This transition is a result of a sharp growth trajectory witnessed across many parallel technology landscapes including cloud for storage and processing, Internet of things (IoT) for rapid data generation and transmission, support for unstructured, semi-structured and structured data by social media platforms, and polyglot persistence supported by NoSQL, to name a few. The emerging applications thrive on insights acquired by data analytics, and immutability of data storage is expected to be one of the key factors in growth of forward leaning enterprises. There has been a constant rise in application development requiring a full history of transactions, which is both trustworthy and traceable. Such applications mandate the append-only nature of the data storage to support analytics and trust. Immutability is also the underlying premise of popular cloud native storage systems. This paper explores blockchain technology as a solution for privacy and disclosure compliance in stored healthcare data. Blockchain as a data structure is primed to store only small amounts of data to maintain the properties of immutability, tamper proofing, security, and transparency in applications and is not suitable for storing big data. However, healthcare blockchain implementations have several possibilities for storage management, and the paper presents a comparative analysis of the aggregation-based storage with IoT-based streaming e-healthcare application data for the use case. The in-depth analysis for the potential blockchain storage in the paper includes the cost factors, in addition to immutability and privacy preservation, and the results show that costs may be saved up to an order of magnitude 300 with aggregation.

S. Gupta (✉) · B. Yadav
SoET, MVN University, Aurangabad, India
e-mail: sachin.gupta@mvn.edu.in

B. Yadav
e-mail: babita.yadav@mvn.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_57

Keywords Healthcare · Compliance data · Immutability · Privacy · Storage economics · Blockchain

1 Introduction

There has been an exponential growth of social and economic activities in online mode. Not only has the routine average consumption of data per user shot up in the past decade, similar amounts of content and data are also being created per user as a by-product. While most user generated data and content can be considered "harmless" to share, there exist several domains of individual data that unquestionably classify as private data. Trust as a commodity is becoming increasingly difficult to maintain with social media rapidly pushing us in the era of data exploitation. Governments are waking up to this reality, and we have witnessed some landmark data privacy legislation happening across the globe in the recent decades. As per the UNCTAD data [1] analyzed for 194 countries, it has been reported that 128 countries have created stringent laws for data and privacy protection of their citizens. Non adherence of the data privacy laws is punishable with heavy fines [2].

The rules of engagement with databases are changing as the technology spectrum is shifting from pure storage to analytical engines. The long drawn "AI winter" [3] ended almost a decade back with a symbiotic increase in data volumes as well as the computation capacity. The new generation data stores are thus not designed for storage alone, they are preferred to be analytical to handle a variety of big data including social media posts, Internet of things (IoT), and immutability is desirable in synchronization the velocity requirements and with the cloud native nature of applications. Analyzing the data for business intelligence and useful patterns requires high fidelity data, with a wide time range of data generation, and the previously used mutable data storage with data overwriting is no longer a good choice in many scenarios. The stock market trends data for instance requires the best of big data storage [4, 5] and immutability for analytics. It would be similarly wise to store the disaster and its management datasets on immutable storage for in-depth analytics and prediction [6].

The recent developments in blockchain have made it possible to ensure the desirable traits of both immutability as well as verifiable data security compliance as discussed in the above sections. This paper presents the economics of immutability preservation by using on-chain blockchain storage and a comparative cost analysis in case of off-chain storage methods with immutable storage and protected via the blockchain technology. The resting heart rate data with 9000 data points from the personal fitbit app health profile [7] of the author have been used for the paper. The data are IoT-based with analytics desired, personal healthcare data governed by data privacy laws, and are a good representative use case for assessing the storage cost impact of using blockchain technology on immutable and GDPR privacy preserving off-chain and on-chain storage methods.

The remainder of the paper has been structured as follows. Section 2 discusses the research and development in the domains of immutability and private data disclosure compliance through a brief literature survey, while Sect. 3 focuses on explanation of the methodology used in the paper for analyzing the storage costs with immutability on blockchain. In Sect. 4, we have presented the comparative cost analysis on on-chain solutions with and without aggregation as the results from Sect. 3, and a concluding summary of the paper along with future research directions has been presented in Sect. 5.

2 Related Work

This paper deals with both privacy preservation laws and immutability of storage for trust-related issues and the proposed solution being blockchain, it is imperative to understand the research landscape of all these topics. The topics have been discussed by various researchers across several applications, but we did not come across any research work unifying these topics comprehensively in the same paper. The available literature across all the related topics is reviewed and presented ahead with topic wise associations as they appear.

2.1 Immutability of Storage

The advancements in cloud-based storage brought the age long myth about immutability being propagated by vendors of storage devices to the fore. The term was being used incorrectly for the difficulty in deletion or modification of a storage-based file by a user. The precise immutability definition as per storage and compliance requires that the data termed as immutable shall not be modified or deleted for contract-bound duration of time by neither men nor software [18]. Immutability is the key USP of a blockchain-based storage [19], and all applications focusing on immutability of data are potential candidates for a shift to blockchain in some form or the other. It is imperative to note that immutability is not a silver bullet. The organizations have to consider what happens when a user actually wants to get his personal medical records deleted. Or the impossible situation of keeping every piece of personal medical data due to it being immutable even after a person dies. Similarly, there might be a requirement to maintain the compliance and regulation data as mutable to allow for possible customer relocation leading to address change or likewise. A split mutable/immutable approach as proposed in some cloud-based solutions like Fluree [20] may help maintain data with different requirements across different media. The pros and cons of on-chain and off-chain storage in blockchains have been comprehensively covered in [32], but we consider only the on-chain storage as a special case in this paper.

2.2 Healthcare Data Privacy Laws

Trust in healthcare analytics has been defined [8] as "a dynamic and always emerging attribute of relationships among people that evolves with time." Bulk of healthcare data is originating from IoT-based medical sensors especially in remote monitoring and telemedicine, and several researchers have proposed trust solutions for such applications. Privacy preservation using a blockchain including a forensic framework for IoT discussed in [21], and a trust framework for collaborative IoT [22] gave a general approach to trust establishment using blockchain. Trust in a healthcare blockchain with a design has also been discussed in [23]. Personal data processing using blockchain to maintain privacy concerns has been adequately covered in [24], while the specific discussion on general data protection regulation (GDPR) [29] compliance has been discussed in [25]. An implementation for a blockchain-based data-sharing platform with GDPR compliance is available in [26] while a Health Insurance Portability and Accountability Act (HIPAA) [28] compliant off-chain data store has been implemented in [27].

2.3 Blockchain in Healthcare

Ever since its origin with the Bitcoin cryptocurrency envisioned as a financial application in the now virally sensational breakthrough research paper [9] of Satoshi Nakamoto in 2008, blockchain has been making waves in every technology domain due to its distributed and immutable ledger property. Estonia is among the early adopters of blockchain for storing healthcare data [10] in the form of medical records of patients. Since then, there has been a significant interest of the healthcare community in this technology. The electronic medical records (EMRs) for the patients with blockchain as a consensual access control mechanism have been discussed by MedRec [11] and HDG [12] while blockchain as an immutable storage mechanism has been discussed in Medibchain [13]. There has been a significant debate on using permissioned vs permissionless blockchains in healthcare owing to the sensitive nature of the patients' medical records and has been discussed in [14, 15]. Almost the entire literature survey focuses on the inability of blockchain to handle copious amounts of data [14] and suggest off-chain solutions but blockchain scalability with large documents still remains a concern [16, 17].

3 Methodology

3.1 Data Description

The potential blockchain use cases for data storage based on decentralized trust relationships can be attributed to applications that work with information transfer and compliance. The applicability of blockchain in this age of internet is universal, and the best representative industries include supply chain management, healthcare, and insurance to name a few. In this paper, the author has used personal fitness data from a popular fitbit application for representative calculation for storage in streaming data on a blockchain. Healthcare as an industry has been specifically chosen here to cover data having both modification and exchange requirements, while working on crucial patient integrity maintenance and law mandated compliance management. We have chosen HIPAA from among the universal laws for representing the legal use case requirements for this paper, as GDPR is a generalized law but for prevention of the misuse or disclosure of the patients' healthcare data is thoroughly covered in HIPAA.

The data include 9000 data points from an IoT wearable fitbit versa device generating sensor-based data for resting heart rate of a user. The data may have multiple stakeholders including the patient, care provider, physician, and the insurance company and is a suitably complex representation of healthcare data under HIPAA regulation. The healthcare data from millions of fitbit devices being used by the customers are currently being stored on company servers with a potential need of load balancing [30].

For the purpose of illustration, we have assumed that the storage system being used by the company is a legacy system with a mutable overwriting-based relational database. The basic problem associated with such systems in case of compliance-based applications is the misinformation due to lack of historical data. With an immutable distributed ledger, it would be easier to provide data ownership to the concerned individual along with traceability of all the previous medical history including diagnosis, consultations, and treatments along with medicine records. The patient's confidence in such a system shall be high on trust and the data thus generated can be used effectively in individual and community AI and ML applications for drug discovery [31] and treatment-related studies. A comparative storage cost analysis using an on-chain blockchain storage with application of aggregation has been done on the acquired fitbit data in the following sections.

3.2 On-Chain Storage Costs

The understanding of storage costs on the blockchain can be considered for public vs private/hybrid global blockchain systems in use being Bitcoin and Ethereum, the latter has been used for cost calculation in our paper. The basis of calculations

comes from the Ethereum yellow paper [33] where it is mentioned that the basic unit of payment in the Ethereum network is gas. Each transaction, whether transfer or execution of smart contract, is made with gas which fuels the network. It is basically a fee paid to the miners for maintaining the transaction computations going on.

Any transaction cost may be represented with a simple to calculate formula given by Eq. 1 as shown below, where gas price is the amount of ETH paid per unit of gas and fluctuates on the basis of what the transaction drivers *intend to* pay the miner for this transaction confirmation, and the gas limit is fairly standardized on type of operation. The service is premium, if you pay more gas fees, transaction confirms faster.

$$\text{gas price} * \text{gas limit} = \text{total cost} \tag{1}$$

The gas is measured in wei, which is the smallest unit of ETH. Typically measured in gwei, or gigawei where 1,000,000,000 Gwei is equal to 1 ETH. From the yellow paper, it is also given that each transaction consumes at least 21,000 gas. Similarly, the cost of storing a 256 bit word to storage has been given a cost of 20,000 gas. We have used the Eth Gas Station [34] for the cost of saving the resting heart rate data. The Fitabase [35] research library adds some unique perspective to resting heart data storage with possibilities of heart rate data being generated at variable sampling with readings every 15 s. We assume a single 256 bit word to store the resting heart rate data in each case.

We have calculated the storage costs on 9000 points, and the cost calculations have been shown under different categories of aggregation as per below mentioned cases:

- Case 1 is with no aggregation, with one point stored per transaction as in streaming data.
- Case 2 considers a grouping aggregation with 300 points stored together during a single transaction
- Case 3 considers a temporal aggregation on the device whereby the aggregated data per 15 min is sent to be stored on the blockchain.
- Case 4 considers a daily aggregation of the streaming healthcare data with all points across active 14 h with 3360 data points were aggregated together and stored during one transaction leading to 9000 points being generated in 2.68 days and stored on immutable blockchain storage, while the device keeps mutable storage records of up to one week.

In each case, as per the fitbit data, each data point is read by the sensor every 15 s, which can be considered on a minute basis (4 points) or an hourly basis (240 points) while the device is being actively worn. We also assume an 8 h sleep cycle of the user and maximum 14 h data availability in a day.

4 Results

The results from the study have been shared in Table 1. The table has been split into three subsections where the first subsection considers the average gas cost to be 73 Gwei (as per the ETH Gas Station data available on 11th October 2021 at 5 PM IST), and the second and third subsections record the data values with maximum cost at (105 Gwei) and minimum cost (69 Gwei).

The comparative cost analysis with the four case scenarios has been plotted in Fig. 1 which shows the cost of storage across all cases across the average, high, and low gas costs.

5 Conclusions and Future Research

Healthcare data may come in various forms, shapes, and sizes. The streaming IoT-based healthcare data are not amenable to be stored on a blockchain owing to heavy expenses of generating a single transaction per data point but with a clever aggregation mechanism as shown in the paper, a partially mutable storage capable of holding a few days data on the IoT sensor device and a partial immutable storage on the blockchain can be used to reduce the costs by approximately 300% as shown in the paper. Paying the highest gas costs invariably speeds up the mean time to confirm the transaction on blockchain in seconds, but the speed/cost tradeoff between the average and the lowest cost is not as significant and thus its always advisable to pay the average gas fee on Ethereum rather the lowest which may make the transaction wait for very long periods.

It would be interesting to have similar calculations as a future work for storage costs in the case of off-chain data storage methods of blockchain in the presence of gas costs associated with execution of smart contracts and to analyze whether the costs of blockchain storage may be offset by storing the data off-chain as the hash and smart contracts are executed on-chain.

Table 1 Ethereum on-chain storage costs

Data stored	Gas used	MTTC (Sec)	Tx fees (ETH)	TX fees (Dollars)	Total cost (9000 points)	Remarks
1	41,000	624	0.002993	$10.31	$92,798.01	Streaming 9000 points storage
300	6,021,000	2270	0.439533	$1,514.19	$45,425.74	300 data points stored with each transaction
60	41,000	624	0.002993	$10.31	$1,546.63	Data aggregated every 15 min
3360	41,000	624	0.002993	$10.31	$27.63	Daily storage (3360 points aggregated over 14 h)
Tx costs @105 Gwei (highest)						
1	41,000	28	0.004305	$14.83	$133,476.48	Streaming 9000 points storage
300	6,021,000	1413	0.632205	$2,177.95	$65,338.39	300 data points stored with each transaction
60	41,000	28	0.004305	$14.83	$2,224.61	Data aggregated every 15 min
3360	41,000	28	0.004305	$14.83	$39.75	Daily storage (aggregated over 14 h)
Tx costs @69 Gwei (lowest)						
1	41,000	1075	0.002829	$9.75	$87,713.19	Streaming 9000 points storage
300	6,021,000	22,464	0.415449	$1,431.22	$42,936.65	300 data points stored with each transaction
60	41,000	1075	0.002829	$9.75	$1,461.89	Data aggregated every 15 min
3360	41,000	1075	0.002829	$9.75	$26.12	Daily storage (aggregated over 14 h)

Fig. 1 Storage costs for 9000 data points in Eth with variable gas fees

References

1. "Data Protection and Privacy Legislation Worldwide," *UNCTAD*. [Online]. Available: https://unctad.org/page/data-protection-and-privacy-legislation-worldwide. Accessed 07 Oct 2021
2. P. Yifat Perry, "Meeting Data Compliance with a Wave of New Privacy Regulations: GDPR, CCPA, PIPEDA, POPI, LGPD, HIPAA, PCI-DSS, and More", Cloud.netapp.com (2021). [Online]. Available:https://cloud.netapp.com/blog/data-compliance-regulations-hipaa-gdpr-and-pci-dss. Accessed 07 Oct 2021
3. "AI winter - Wikipedia", En.wikipedia.org (2021). [Online]. Available: https://en.wikipedia.org/wiki/AI_winter. Accessed 08 Oct 2021
4. S. Kalra, S. Gupta, J.S. Prasad, predicting trends of stock market using SVM: a big data analytics approach. In: *Data Science and Analytics. REDSET 2019. Communications in Computer and Information Science* ed. by U. Batra, N. Roy, B. Panda, vol. 1229 (Springer, Singapore, 2020). https://doi.org/10.1007/978-981-15-5827-6_4
5. S. Kalra, Dr. S. Gupta, "Performance evaluation of machine learning classifiers for stock market prediction in big data environment". Published—J. Mech. Continua Math. Sci. Web Sci., **14**(5), (2019). https://doi.org/10.26782/jmcms.2019.10.00022
6. S. Gupta, B. Gupta, Performance modeling and evaluation of transportation systems using analytical recursive decomposition algorithm for cyclone mitigation. J. Inf. Optim. Sci. **40**(5), 1131–1141 (2019). https://doi.org/10.1080/02522667.2019.1638003
7. "Fitbit Official Site for Activity Trackers and More", Fitbit.com, 2021. [Online]. Available: https://www.fitbit.com/global/in/home. Accessed 08 Oct 2021
8. P. Holub, F. Kohlmayer, F. Prasser, M.T. Mayrhofer, I. Schlünder, G.M. Martin et al., Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-health. Biopreserv Biobank. **16**, 97–105 (2018)
9. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (2008). [Online]. Available: https://bitcoin.org/bitcoin.pdf. Accessed 08 Oct 2021
10. T. Heston, A case study in blockchain healthcare innovation. Int J Curr Res. **9**, 60587–60588 (2017)
11. A. Ekblaw, A. Azaria, J.D. Halamka, A. Lippman, "A case study for blockchain in healthcare:"MedRec" prototype for electronic health records and medical research data", 13 (2016)

12. X. Yue, H. Wang, D. Jin, M. Li, W. Jiang, Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst. **40**, 1–8 (2016)
13. A. Al Omar, M.S. Rahman, A. Basu, S. Kiyomoto, "Medibchain: a blockchain based privacy preserving platform for healthcare data", In: *Proceedings of the International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage* (Guangzhou, China, Springer, 2017)
14. O. Choudhury, H. Sarker, N. Rudolph, M. Foreman, N. Fay, M. Dhuliawala et al., Enforcing human subject regulations using blockchain and smart contracts. Blockchain in Healthcare Today. **1**, 1–14 (2018)
15. A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, F. Wang, "Secure and trustable electronic medical records sharing using blockchain" (Washington, D.C., United States, American Medical Informatics Association, 2017)
16. Q. Xia, E. Sifah, A. Smahi, S. Amofa, X. Zhang, BBDS: blockchain based data sharing for electronic medical records in cloud environments. Information **8**, 44 (2017)
17. P. Mamoshina, L. Ojomoko, Y. Yanovich, A. Ostrovski, A. Botezatu, P. Prikhodko et al., Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget **9**, 5665–5690 (2018)
18. B. Tolson, Litigation Hold and Data Immutability (2021). [online] Archive360.com. Available at: https://www.archive360.com/blog/litigation-hold-and-data-immutability-why-a-litigation-hold-does-not-meet-the-regulatory-definition-of-immutability. Accessed 9 Oct 2021
19. Network World, How data storage will shift to blockchain (2021). [online] Available at: https://www.networkworld.com/article/3390722/how-data-storage-will-shift-to-blockchain.html. Accessed 9 Oct 2021
20. Fluree, Fluree | The Web3 Data Platform Fluree (2021). [online] Available at: https://flur.ee/. Accessed 9 Oct 2021
21. D.-P. Le, H. Meng, L. Su, S. L. Yeo, V. Thing, "BIFF: a blockchain-based iot forensics framework with identity privacy," in *Proc. TENCON IEEE Region Conference*, pp. 2372–2377 (2018)
22. B. Tang, H. Kang, J. Fan, Q. Li, R. Sandhu, "IoT passport: a blockchain-based trust framework for collaborative Internet-of-Things," in *Proc. 24th ACM Symposium Access Control Models Technology*, pp. 83–92 (2019)
23. A. Dwivedi, G. Srivastava, S. Dhar, R. Singh, A decentralized privacy-preserving healthcare blockchain for IoT. Sensors **19**(2), 326 (2019)
24. K. Rantos, G. Drosatos, K. Demertzis, C. Ilioudis, A. Papanikolaou, "Blockchain-based consents management for personal data processing in the IoT ecosystem," in *Proceedings 15th International Joint Conference e-Bus Telecommunications*, pp. 572–577 (2018)
25. M. Rhahla, T. Abdellatif, R. Attia, W. Berrayana, "A GDPR controller for IoT systems: application to e-Health," in *Proceedings IEEE 28th International Conference Enabling Technology Infrastructure Collaborative Enterprises (WET-ICE)*, pp. 170–173 (2019)
26. B. Faber, G.C. Michelet, N. Weidmann, R.R. Mukkamala, R. Vatrapu, BPDIMS: a blockchain-based personal data and identity management system. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (Honolulu, Hawaii, United States, Hawaii International Conference on System Sciences (HICSS), 2019)
27. K. Griggs, O. Ossipova, C. Kohlios, A. Baccarini, E. Howson, T. Hayajneh, Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. J. Med. Syst. **42**, 1–7 (2018)
28. UNITED STATES, The Health Insurance Portability and Accountability Act (HIPAA). [Washington, D.C.], U.S. Dept. of Labor, Employee Benefits Security Administration (2004). http://purl.fdlp.gov/GPO/gpo10291
29. EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1

30. M. Yadav, S. Gupta, Hybrid meta-heuristic VM load balancing optimization approach. J. Inf. Optim. Sci. **41**(2), 577–586 (2020). https://doi.org/10.1080/02522667.2020.1733190
31. J. Vamathevan, D. Clark, P. Czodrowski et al., Applications of machine learning in drug discovery and development. Nat. Rev. Drug. Discov. **18**, 463–477 (2019). https://doi.org/10.1038/s41573-019-0024-5
32. X. Xu et al., "A Taxonomy of Blockchain-Based Systems for Architecture Design", In *Proceedings of 2017 IEEE International Conference on Software Architecture, ICSA 2017*, pp. 243–252 (2017)
33. G. Wood, Ethereum: a secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper **151**(2014), 1–32 (2014)
34. Eth Gas Station [Online] Available: https://legacy.ethgasstation.info/calculatorTxV.php. Accessed 9 Oct 2021
35. "Fitbit Research Library", Fitabase.com. [Online]. Available: https://www.fitabase.com/research-library/. Accessed 9 Oct 2021

A Comparative Cost Analysis of Organizational Network Security Test Lab Setup on Cloud Versus Dedicated Virtual Machine

Sachin Gupta, Bhoomi Gupta, and Atul Rana

Abstract The global network infrastructure spectrum is witnessing its fastest growth since the last decade with concurrent rise in cloud computing, Internet of things (IoT), and edge computing. There has been a multitude of heterogeneous networking devices spanning different configurations and using a variety of access methods. A parallel evolution of the network infrastructure security is happening with increasing attempts to exploit the security vulnerabilities in mission critical cyber-assets of organizations. Several organizations invest heavily in security research using lengthy and cryptic mathematical models while ignoring the practical network implementation situation and focus only on the monetary implications of the attack and defense. Attack tree has evolved as a convenient and cost effective way of plotting the network in which an attack may take place and can also help organizations understand the way it can be defended. Attack trees combined with the MITRE ATT&CK framework are widely used for crown jewels risk assessment globally. However, the major challenge for information security experts using the attack tree methodology lies in manually creating the attack tree and plotting all the crown jewels and perimeter network so that it can be defended from attackers. We propose a test lab setup for simulation and attack tree generation, which can be used in conjunction with the MITRE ATT&CK framework and allow us to create and assess various attack scenarios while providing flexibility in subnet configuration and movement, addition or removal of networking devices. The lab can be cloud hosted with a popular cloud hosting on Microsoft Azure or may be created on a VM within a dedicated high-resource machine to be used as a portable testbed. The results indicate that both services have their own pros and cons based on the hours of usage, and the dedicated resource VM testbed may

S. Gupta (✉) · A. Rana
SoET, MVN University, Haryana, India
e-mail: sachin.gupta@mvn.edu.in

B. Gupta
MAIT, GGSIPU, Delhi, India
e-mail: bhoomigupta@mait.ac.in

A. Rana
Research Scholar, MVN University, Haryana, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_58

perform better in a low-risk potential small network while the cloud-based approach is useful for the scalable organizations with high-threat potential.

Keywords MITRE framework · Attack tree · Test lab · Network security

1 Introduction

The term cybercrime, cybercriminals, and cyberattack are not new for us in today's world, but with increased use of remote working, automation, and increased device connectivity, the term cyberattack has reached new horizons of all time high.

Businesses have gone online and organizations trying to provide all the information as readily available to its customers in a self-serving manner has also allowed cybercriminals to infiltrate the organization even more than ever. If we look at the last 5 year data of cybercrimes, it has been on a constant growth trajectory and has also triggered safety costs as per a report published by cybersecurity ventures [1] which shows that spending on cybersecurity all around the globe was even more than $ 1 trillion (2017–2021). The attack surface of cybercrime is so high that it seems very difficult to understand how we can even protect our identity and assets. Cybercrime is causing unpreceded damage to both industry as well as the personal lives of individuals. It has also been observed that by end of 2016, a business becoming a victim of a cyberattack was 21 s but this has been reduced to 14 s by year 2019 and is also expected to reduce further to 11 s in 2021. Phishing, ransomware, and malware are few of the fastest growing cybercrime mechanisms. This has caused extremely damaging network infections and big downtime with a lot of reputation damage. If we look at the data of damage caused only by ransomware which was $5 billion in 2017 which is already 15 times higher if compared to year 2015 have also increased further to $11.5 billion in 2019 and expected to grow $20 billion in 2021 as per the report published by Herjavec group [1].

Over a decade, cloud computing opened a new horizon for businesses and has generated exciting opportunities for organization in terms of flexibility, cost, high availability and scalability with reduced costs. Many companies have made a move from traditional data centers to cloud environments to help them get better returns with lower costs. The on-demand access to network resources has made it very scalable. According to a recent Gartner press report, [14] published the demand of cloud computing has been rising every year and it is expected to be reaching up to $308 billion by end of year 2021 which is far higher when compared to the previous years. With the increased use of cloud computing, it brings new risks also with it which are even higher than that of traditional data center environments.

Security operation centers in organizations find it difficult nowadays to keep pace with the volume of cyberattacks and volume of events received every day. Most of the security tools in the market focus on alerting when an event has happened rather than making a predictive guess and showing the insight about the weakness and strengths within the organizations. Another survey shows that most of the organizations are

finding it difficult to understand the network as it is huge and that is one of the biggest reasons for defending it against adversaries.

Attack tree (which was first introduced by Bruce Schneier in 1999) is a convenient way of plotting the network in which an attack may take place and can also help organizations understand the way it can be defended. There are lots of research papers published on attack trees which show lots of academic ways with the help of mathematical manners to understand metrics related to probability, cost, and impact, but this research paper will focus more on how we can create an attack tree and how organizations can be benefited from such an attack tree.

Attack trees are diagrams that represent assets of the organization that can be a target to an attacker, it also provides a mechanism to not only model threats against organization but also helps in identifying a countermeasure for that threat. It is similar to a red team blue team exercise in one attack tree. Security analysts can define a way of accessing and breaching the defense while other teams can suggest a countermeasure against that attack, i.e., placing a detective or preventive tool against it. It provides graphical methods of how a security of an organization looks like and represents possible attacks against those security measures. Attack tree may consist of a few root nodes which represent a final target of an attacker but also several child nodes that help an attacker reach its target. There are also few connectivity's that help an attacker to reach to the next child node, and these are attack methods.

Attack tree can become a very big and complex representation based on organization size and also depends on some specific attack. Attack tree can also prove beneficial in defining the strategy on how to protect assets from an attack and understand the scenario to mitigate the threat to the root node. In practical situations, this may not always be possible to eliminate the threat with 100 percent accuracy but we can always make it more difficult for an attacker to reach the root node. We can try to increase the child nodes between an attacker and the root node but this comes with a cost and that can be a cost of placing a new security tool or hiring a new security analyst to administer and monitor the tool.

2 Related Work

Attack trees were introduced for the first time in 1999 by Bruce Schneier [2] to understand the attacker, their goals, likelihood of an attack, and the need of threat modeling. The concept of root node, leaf node, and also Boolean nodes (AND, OR) is also introduced. The concept of attack tree [9, 10] and attack graph is used to identify threats based on vulnerabilities in the organization with possible impact that they can cause. The attack tree and graph were very good in explaining "what if" situation of an attack while they lacked in the graphical virtualization as it is very complex to build and maintain for long run which has been explained by Steven Noel [13] managing attack graph complexity paper and also precision on manipulation of the attack tree construction by study of transformation and attack tree [11, 12] with attributes and projection of the tree.

There was another paper published in 2005 [3] which reached to another side of attackers and this time it is insider attackers within the organization and the paper proposed a framework to detect malicious activities from an authorized insider within the organization. This concept was to generate an alarm when it assumes that a malicious activity is about to trigger which was a shortcoming with the intrusion detection systems which work on the signature bases and trigger when some malicious operation has already been performed. The concept was to use an additional framework on top of the intrusion detection system and build an attack prediction system.

In a research paper published in 2006 [4], it was researched to provide foundation to the attack tree which was done to answer a few questions that were not provided previously. This research paper described semantics in terms of attack, attack suits, valuation, and projection in a manner that is formal in nature and also provided algebraic conditions that provide better results. This paper also looked at the fault that were proposed in the previous paper [5] where author preferred attack net which is a formalization of threat analysis based on petri net more than the attack tree and lacking in semantics formalization which was later improved by another research paper published [6] that successfully applies attacks trees to multi-model Internet attacks. The paper proposed a new approach and it proposed a new word called bundle which is a connection from a node to multiset node, and this bundle should be executed to execute an attack which is different from Schneier's proposed attack tree.

In a research paper published in 2007 [7] which highlighted the limitation of attack trees being time consuming to generate for a complex system and proposed a recommendation that automation can help in this shortcoming. The paper proposed combining attack libraries with analyst provided specifications can make the automation easier. This paper also talks about a protection tree which provides a methodical approach of mitigating weakness highlighted by the attack tree. The few important words like probability, cost, and impact also were used, and metrics were created to draw a conclusion but according to this paper, it is a bit hard to guess.

In [13], Barbara Kordy (2010) has proposed a new tree structure named as attack defense tree which uses an interactive approach that supports semantic approach that helps in prompt response in an event of attack. A new research paper which was published in 2015 [8] proposed a new idea of the first foundation of the SAND attack tree which is the extension of an attack tree. Here, in this approach, a new attack tree is defined with a new sequential operator and then a new graph which is a series–parallel graph has been defined on which the semantics of the new attack tree is based.

In a research paper, [8] published by Jhawar, Kordy, Mauw, Radomirović, and Trujillo-Rasua, 2015 first formal foundation of SAND tree was proposed which was an extension on attack tree by addition of sequential operator SAND, which helps in modeling of ordered events. This paper is a further extension to multiset semantics to SAND attack trees with series–parallel graphs, and it helps in interpreting the tree using both commutative and sequential conjunctive refinement.

3 Methodology

Attack trees are a very interesting and innovative way to understand the current status of an organization with regards to its defense mechanism as well as its exposure and coverage level for an attacker. Attack trees can also be used based on application or as a whole for an organization, and they are a good way to document and realize the threat possessed by an organization. There are, however, a few common questions related to the understanding of an attack tree with reference to an organization, and they include

- What is the best way to model an attack tree?
- How can we attach a framework with regards to tactics and techniques used by attackers?
- How can attack trees help in understanding the risk for an organization?
- Are attack trees a solution going forward with regards to complexity and exponential growth?

In this paper, we have created the template for a test bed to model the attack tree of a virtual organization and to assess it with MITRE ATT&CK framework which is a well-known attack framework based on tactics and techniques based on real-world observations. The configuration thus created has been used for comparative cost analysis of the attack tree modeling using the physical network, a cloud-based approach, or a VM-based setup on a dedicated high-resource machine.

3.1 Cloud-Based Testbed

We have used the Azure cloud environment to estimate the costs for setting up the lab [15]. This will require a subscription to Azure environment and few resource groups within the subscription to place resources according to the device type, e.g., all servers in one resource group with balancing [20]. Also we can use Microsoft security tools to create a security operation center in this test environment. The main aim of creation of this lab is to simulate the same kind of environment that is used in the organizations and then create an active attack defense environment for testing the attack tree.

The estimated costs for a security-based network in Azure are shown in Table 1.

3.2 Dedicated Machine VM-Based Testbed

The dedicated machine with a testbed virtual network can be created on a machine with following hardware configuration:

Table 1 Estimated cost for an Azure-based testbed for security

Microsoft azure estimate		
Service type	Description	Estimated monthly cost ($)
Storage accounts	1 K GB with 10 K operations	21.84
Virtual machine scale sets	Licensed 3.5Ghz 730 h win machines	102.20
Azure DNS	Public domain name service	0.00
Network watcher	1 gigabyte logs	5.80
Azure firewall	Single firewall 730 h	912.50
Azure bastion	5 gigabyte 730 h monthly usage	138.70
Azure front door	1000,000 custom requests	138.00
Azure firewall manager	Free with firewall	0.00
StorSimple	Storage size dependent	0.00
Azure active directory external identities	Active directory service	0.00
Azure sentinel	Daily logs up to 100 numbers	8880.00
VPN gateway	128 tunnels p2s	0.00
Azure monitor	Log analytics: 0 GB daily logs ingested, 1 month data retention; Application Insights: 0 GB daily logs ingested, 3 months data retention, 0 multi-step Web tests; 1 VM monitored X 1 metric monitored per VM, 1 log alert at 5 min frequency, 0 additional events, 0 additional emails, 0 additional push notifications, 0 additional Web hooks (in millions)	1.60
Support	**Support**	0.00
	Total	**10,200.64**
	Annualized	**1,20,000.00 (Approx)**

Motherboard supporting 6 GB/s or above, 8 core processor, 256 GB RAM, 2 TB SSD, and 16 GB HDD or above. This shall allow for setting up a virtual network with high availability and optimized as per the plan indicated below:

- Install a BASE machine and create a "full clone" for each machine (disk consuming), i.e., around 20 GB per machine + the applications space.
- Install a BASE machine and create a "linked clone."

The second mentioned technique will use a base 20 GB image used by all the VM's, i.e., every further machine saves 20 GB of disk space each in this technique.

For e.g., base machine = 20 GB.

Machine 1 = 0 GB of linked machine + VRAM of its own + application data.

Table 2 Estimated cost for a dedicated VM-based testbed for security

Dedicated machine estimate		
Part	Description	Estimated cost ($)
Motherboard	MSI MAG B460M 6 Gb/s Intel Motherboard	190
Processor	Intel Core i9-11900KF 8-Core 3.5 GHz	600
RAM	256 GB (8 × 32 GB) DDR4	1660
SSD	2 TB 3,500 MB/s SSD	250
HDD	16 TB HDD 6 Gb/s 512e/4Kn	339
Firewall	NETGATE 5100 PFSENSE + SECURITY GATEWAY [17]	699
Squid	Squid for proxy	0
Server	Windows server 2012 (DNS, DHCP)	429
Client machines	Win 7, 8, 10	300 (per machine) 3000 (10 machines)
Client machines	Kali Linux, Ubuntu	0
AD	Windows server 2016 (active directory)	1069
IDS	Security onion for security monitoring tool	0
Honeypot	Tpot for honeypot services [18]	0
Log	Sysmon log system activity to the Windows event log	0
SIEM	Splunk/Sentinel/Graylog for SIEM solution	0 or 2000 (Annualized)
	Total	**9537**

All machines using this technique access the same portion of the HDD simultaneously. We should not use this technique if we are running an HDD which can lead to bad sectors. An SSD is highly recommended for such access. RAM disk out performs, we need to allocate physical RAM for it though.

The cost estimates available as on date using Newegg.com [16] and individual vendors software for the recommended hardware configuration are listed in Table 2:

The software can be run on VMware 16 Workstation Pro, available at [19] for $199. Adding miscellaneous costs, we may approximate the total system cost at a $10,000 onetime investment.

4 Results

The comparative analysis of the details presented in Sect. 3 cannot be considered on the presented costs alone. The hidden cost factors in both scenarios exist and are summarized in Table 3 for an easy reference. Keeping the target variable fixed at simulation of the attack trees on the testbed machines, the comparable variable parameters have been discussed.

Table 3 Comparative analysis of testbeds

Type of testbed	Scalability	Human monitor expertise	Cost of setup	Dynamic configuration	Ease of experiment	Annualized charges ($)
Cloud (Azure) testbed	High	Low (built-in analytics)	Fixed + Pay as you use	Easy	Low	1,20,000
Dedicated VM testbed	Low	High (analysis needed)	Fixed one time cost	Difficult	High	10,000*

* The prices used may vary as per market

If we keep on increasing the size of the VM testbed, the comparative cost difference keeps decreasing as the licensing costs go up. The number of machines vs licensing costs has been depicted in Fig. 1, but the scalability is an issue here as explained below:

The 256 GB RAM limitation imposes an upper limit on the number of VM client machines even with cloning methods. The cost dynamics for the testbed considering a 2 GB clone machine size and a 256 GB RAM availability shall reach $30,000 for licensing with a $300 individual license for a maximum of 100 machines occupying 200 GB RAM leaving 56 GB operational to run servers, network switches, and host operating system utilities.

Each component in Table 3 is explained below for an easy reference:

Fig. 1 Cost variation for dedicated VM with increasing users

- Scalability: The cloud-based testbed shall provide an easy scalability across network configuration, while the dedicated machine is limited by the RAM capacity and SSD size.
- Monitor Expertise: The testing person has to be highly skilled across several technology domains including virtualization and networking besides security in case of a dedicated VM testbed, while the built-in analytics dashboards with Azure make the task easier in case of cloud-based testbeds.
- The cost of setup for Azure is recurring and billed monthly while the dedicated VM cost is onetime investment.
- The services running on Azure cloud testbed can be dynamically integrated with new requirements while it shall remain difficult to change the configuration of a running dedicated VM testbed.
- The ease of experiment in case of a cloud-based testbed shall remain limited to the permissible frequency of use cases provisioned by the tier while an unlimited number of trials can be handled on the dedicated VM testbed.
- The cloud-based model is safe from disasters as discussed in [21].

5 Future Work and Conclusions

The comparative cost analysis indicates that establishing a testbed for up to 100 client machines is still a lot cheaper than the annualized cost of a security testbed established on the cloud with basic free tier usage of most components. A cloud-based VM setup with a bare metal cloud service provider and custom software can lower the costs in comparison with both approaches and needs to be studied further. It would also be possible to research the risks attached to an organization by setting scenario-based experiments on the testbeds using the attack tree which may help us model the risk associated and corresponding mitigation strategies. Gamification of network security assessment through the testbeds is another open area of research.

References

1. Herjavecgroup.com Official Annual Cyber Crime Report 2018. [online] Available at: https://www.herjavecgroup.com/wp-content/uploads/2018/12/CV-HG-2019-Official-Annual-Cybercrime-Report.pdf. Accessed 10 Oct 2021
2. B. Schneier, "Attack trees: modeling security threats," Dr. Dobb's Journal, December (1999)
3. I. Ray, N. Poolsapassit, Using attack trees to identify malicious attacks from authorized insiders. In Proceedings of the 10th European Conference on Research in Computer Security (ESORICS'05) (Springer, Berlin, Heidelberg, 2005), pp. 231–246. https://doi.org/10.1007/11555827_14
4. S. Mauw, M. Oostdijk, Foundations of attack trees. Lecture Notes Comput. Sci. **3935**, 186–198 (2006). https://doi.org/10.1007/11734727_17
5. J. Stefan, M. Schumacher, Collaborative attack modeling. In Proc. SAC 2002 (ACM, 2002), pp. 253–259

6. T. Tidwell, R. Larson, K. Fitch, J. Hale, Modeling internet attacks, in Proceedings of the 2001 IEEE Workshop on Information Assurance and Security (2001)
7. K. S. Edge, G.C. Dalton II, R.A., Raines, R.F., Mills, "*Using Attack and Protection Trees to Analyze Threats and Defenses to Homeland Security*" MILCOM 2007 (2007), pp. 1–7
8. R. Jhawar, B. Kordy, S. Mauw, S. Radomirovic, R. Trujillo-Rasua, Attack trees with sequential conjunction. IFIP Adv. Inf. Commun. Technol. **455** (2015). https://doi.org/10.1007/978-3-319-18467-8_23
9. S. Noel, S. Jajodia, "Managing attack graph complexity through visual hierarchical aggregation", in Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security (New York, USA, 2004), pp. 109–118
10. O. Sheyner, Scenario Graphs and Attack Graphs. Ph.D. thesis, Carnegie Mellon University (2004)
11. S. Noel, S. Jajodia, "Managing attack graph complexity through visual hierarchical aggregation", in Proceedings of the workshop on Visualization and data mining for computer security (New York, USA, 2004), pp. 109–118
12. S. Mauw, M. Oostdijk, Foundations of attack trees, in *ICISC 2005*. ed. by D.H. Won, S. Kim. LNCS 3935. (Springer, Heidelberg, 2005), pp. 186–198
13. B. Kordy, S. Mauw, S. Radomirovic, P. Schweitzer, "Foundations of Attack–Defense Trees," In: LNCS (Springer, Heidelberg, 2010). Available at http://satoss.uni.lu/members/barbara/papers/adt.pdf
14. Gartner, Gartner Forecasts Worldwide Public Cloud End-User Spending to Grow 18% in 2021 (2020). [online] Available at: https://www.gartner.com/en/newsroom/press-releases/2020-11-17-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-grow-18-percent-in-2021. Accessed 16 Oct 2021]
15. All prices shown are in US Dollar ($). This is a summary estimate, not a quote. For up to date pricing information please visit https://azure.microsoft.com/pricing/calculator/. This estimate was created at 10/16/2021 6:09:58 AM UTC
16. Newegg.com. 2021. Newegg—Shopping Upgraded. [online] Available at: https://www.newegg.com/. Accessed 16 Oct 2021
17. Netgate. 2021. Netgate 5100 pfSense+ Security Gateway. [online] Available at: <https://shop.netgate.com/products/5100-pfsense> [Accessed 16 October 2021].
18. GitHub, GitHub—telekom-security/tpotce: T-Pot - The All In One Honeypot Platform (2021). [online] Available at: <https://github.com/telekom-security/tpotce>. Accessed 16 Oct 2021
19. V. Pro, VMware Workstation 16 Pro. [online] Store-us.vmware.com (2021). Available at: https://store-us.vmware.com/vmware-workstation-16-pro-5424176500.html. Accessed 16 Oct 2021
20. M. Yadav, S. Gupta, Hybrid meta-heuristic VM load balancing optimization approach. J. Inf. Optim Sci. **41**(2), 577–586 (2020). https://doi.org/10.1080/02522667.2020.1733190
21. S. Gupta, B. Gupta, Performance modeling and evaluation of transportation systems using analytical recursive decomposition algorithm for cyclone mitigation. J. Inf. Optim. Sci. **40**(5), 1131–1141 (2019). https://doi.org/10.1080/02522667.2019.1638003

Insights into the Black Box Machine Learning Models Through Explainability and Interpretability

Sachin Gupta and Bhoomi Gupta

Abstract Artificial intelligence (AI) and machine learning (ML) technologies are considered to be the Holy Grail for the researchers across the world. The applications of AI and ML are proving disruptive across the global technological spectrum, and there is practically no area which has been left untouched by these technologies right from computer science to manufacturing, healthcare, insurance, credit ratings, cybersecurity, and many more. It would not be an exaggeration to say that it is the next big thing after the advent of the Internet and potentially holds a similar impact in touching the lives of human beings. Whilst most researchers using machine learning in research across diverse domains do not need to look beyond the model abstraction for their work, the need for understanding what is happening beneath the surface is sometimes necessary. This becomes especially important in the cases where the predictions are too good to be apparently true, and the researcher running the model is not sure about its validity as the logic for prediction is obscure. The process of feature engineering brings in more accuracy to predictions, but in the absence of intuitive background information regarding the features, the task gets more challenging. The scientific reasoning has been driven by logic through ages, and the scientist community remains sceptical of the results unless they can extract useful insights from the black box ML models. The paper applies five popular explainability algorithms being used by the research community to demystify the abstract nature of ML black box models and compare the relative clarity of the insights being provided individually by each from a practitioner's perspective using the publicly available UCI wine quality dataset.

Keywords Machine learning · Black box models · Feature engineering · Interpretability · Explainability

S. Gupta (✉)
MVN University, Aurangabad, India
e-mail: sachin.gupta@mvn.edu.in

B. Gupta
Maharaja Agrasen Institute of Technology, Delhi, India
e-mail: bhoomigupta@mait.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_59

1 Introduction

Data are the new oil, and a similar analogy can be thought of as the advances in machine learning are lubricating the fast paced wheels of innovation and technology. There are a multitude of ML models being used by the scientific community, and with the explosion of data and ever increasing computational power, the accuracy of the models is also getting near perfect. This remarkable increase in accuracy is certainly a desirable attribute of ML models, but it is not the only most important thing [1]. There has been a long standing debate between those advocating the computational efficiency of neural network-based deep learning at the cost of human interpretability versus the researchers who favour the traceability of regression methods for an enhanced experience of explainable artificial intelligence (XAI) [2, 3]. The famous explainable machine learning challenge of 2018 which was co-hosted by FICO, Google, Berkeley, Oxford, UC Irvine, MIT, and Imperial invited the participants to write a black box ML model and explain its working but the winning team flipped the challenge with writing a solution that was a globally interpretable white box and worked with the same accuracy [4]. The challenge has since then brought forward a tremendous amount of research in explainability of the machine learning models. Whilst the black box models still hold a position of dominance in the decision-making process of the ML world for low-stake decisions involving no risk to human lives, it is no longer the only option. With increasing penetration of ML in life-saving applications including critical healthcare, robotic surgeries, and the pandemic predictions, it is becoming imperative to have insights into "how" the prediction is coming through and "what" portions of data it is using for the same [5].

2 Related Work

The earliest use of interpretability, though not exactly in the same terms as explainability, is due Miller [6], who defined interpretability by correlating the cause of decision-making to the degree of human understanding. In the specific context of machine learning, Kim et al. [7] described interpretability in the terms of the ease of human reasoning and traceability of any prediction made by a model, and the degree of consistency with which the reasoning could be made. The contextual definition evolved in recent times with a paper by Kim and Doshi-Velez redefining interpretability in terms of "explainability" in understandable ways to humans [8]. The model explainability as a trust issue has also been explored by Ribeiro et al. for classifier prediction [9], whilst the new European General Data Protection Regulation [10] guidelines make explainability of algorithmic decision-making mandatory as a fundamental right in the context of data. A detailed survey of the black box models being used in ML can be attributed to recent work of Guidotti et al. [11] which makes a general classification of the explanation models into three distinct categories based on the granularity of explainability. The coarse grain methods just

providing an approximate explanation of the model as a black box are classified as global [12, 13] whereas the local methods [9, 14, 15] allow better insights into the mechanism by providing outcome specific explanations over instance specific datasets. Whilst the global methods are sometimes as obscure as a black box, at times, they provide useful insights in the working of a model consistently across different datasets; on the contrary, the local models give detailed insights but the insights obtained are highly variable across different datasets and are thus unreliable at times. In the next section, we discuss noteworthy model-agnostic global and local ML explainability models including permutation importance [16], partial dependence plots [17], individual condition execution [18], local surrogate [9] and SHAP [19].

3 Methodology

The famous white wine quality dataset from UCI machine learning repository [20] has been used as a means to procure multivariate data of a sufficient quantity, and with 12 attributes and 4898 rows, the dataset is suitable for making a fair assessment of the explainability models. The dataset was also chosen for the desirable properties with no null values and having only numeric data types including 11 float type of variables/1 integer for the ease of study. The features in the dataset used are described in [20] and the chosen target variable is "quality" as recorded by the sommeliers. After cleaning the dataset, dropping the target attribute and applying the basic exploratory data analysis to the remaining 11, the following insights were discovered from the histogram. No feature engineering has been applied for the study. Only sample plots shown in Fig. 1.

From the plots, it is clearly apparent that some features need to be engineered to compensate for the skewed data. The volatile acidity as seen from Fig. 1 is a potential feature engineering candidate for instance. In order to assess model-agnostic interpretability methods, we have not used any feature engineering at this stage as well.

To further understand the correlation between different attributes, and especially with the target attribute of quality, a heatmap was also plotted as per Fig. 2, which gave additional insights into the relative importance of the attributes. The correlation of quality was observed having the highest degree of positive correlation with alcohol content with 0.44, whilst a high-negative correlation is seen with the density of the wine at −0.31. The other attributes can also be seen in context and needs a good domain knowledge to interpret other than the usual taste palette wine parameters of acidity and alcohol.

Having explored the dataset through basic exploratory analysis methods, we now split our dataset in the ratio of 80:20 for training and testing the machine learning model, in order to apply the post-hoc analysis methods. It is interesting to observe that although it would be difficult to predict "why" our ML model is giving certain answers, it is still possible to understand "how" it is coming to those conclusions.

Fig. 1 Basic EDA of the white wine quality dataset

Fig. 2 Attribute correlation heatmap

This interpretability gives us the ability to improve our work through debugging and makes it suitable for better feature engineering and therefore a better human decision-making process. It adds the essential human TRUST component to the otherwise black box model.

3.1 Feature Importance

A baseline understanding of the relative importance of the features was obtained by plotting the "feature importance" function of XGB which allows for an analysis based on feature weights as well as that based on feature coverage in samples. The native feature importance given by both methods has been shown in Fig. 3a, b.

The plots obtained by the two native XGBoost methods to plot global relative importance of features are inconclusive, and this anomaly needs to be investigated even though the feature by sample coverage seems more in synchronization with the

Fig. 3 **a**. Relative importance of the features as per feature weight. **b**. relative importance of the features as per coverage

Fig. 4 Permutation importance using Eli5 for relative importance of features

Weight	Feature
0.0968 ± 0.0058	alcohol
0.0790 ± 0.0106	volatile acidity
0.0370 ± 0.0028	free sulfur dioxide
0.0224 ± 0.0038	residual sugar
0.0214 ± 0.0045	citric acid
0.0147 ± 0.0022	density
0.0145 ± 0.0034	chlorides
0.0130 ± 0.0050	total sulfur dioxide
0.0115 ± 0.0033	sulphates
0.0114 ± 0.0024	pH
0.0108 ± 0.0043	fixed acidity

results obtained from the multivariate heatmap. The inconclusive nature of these plots also leads to opacity, and that is precisely the reason why interpretability methods may be used to demystify the "hows" of our machine learning model.

3.2 Permutation Importance

The "Explain like I'm 5" or popularly known as the permutation importance method Eli5 [16] applied to the dataset produces a weighted list of features as given in Fig. 4. Permutation importance grades the features based on how the model accuracy deteriorates when the feature is not available or its values are shuffled across the column. The method when compared with results obtained by native feature importance is comparatively better in terms of its intuitive nature.

3.3 Partial Dependence Plots

"The partial dependence plot (short PDP or PD plot) shows the marginal effect one or two features have on the predicted outcome of a machine learning model" [17]. The method does not provide a relative feature importance but indicates how our target variable gets affected by the change in value of a feature. Figure 5 shows how sulphates values between 0.5 to 0.6 give a maximum boost to quality which remains relatively stable after 0.6 with a very little but consistently increasing contribution with values.

Fig. 5 Partial dependence plot showing the effect of sulphates on quality

3.4 Individual Conditional Expectation (ICE) Plots

The PD plot is built around average behaviour and hides the individual instance effect on the contribution of a feature on the prediction. The ICE plots are instance specific may uncover heterogeneous relationships and give better insights into the average behaviour of the PD plots by plotting a number of instances. In Fig. 6, the ICE plot clearly shows a distinct jump in the quality prediction based on sulphates values between 0.5 and 0.6, and this might be a characteristic to predict a good quality white wine.

Fig. 6 Independent conditional expectation plot for sulphates contribution to quality

3.5 SHapley Additive ExPlanations (SHAP)

SHAP by Lundberg and Lee [19] explains the individual predictions based on the output as a sum of effects of each feature being introduced into a conditional expectation. The underlying game theoretic approach makes the predictions pretty close to human intuition. The Shapley value is the average marginal contribution of a feature value over all possible coalitions.

Explainability of the Shapley values is very intuitive and can be reasonably well understood with the following example. Consider n features or attributes for a model which are allowed to enter a gaming contest in a random order. The game here is called predict to win, and scoring is based upon the contribution of each feature. If we denote the Shapley value with ϕ_{ab}, it would be calculated as the "average marginal contribution" of the feature instance f_{ab} after it is added to the mix of other feature values that have entered the contest before it.

$$\phi_{\mathrm{ab}} = \sum_{\text{All Orderings}} \mathrm{val}(\{\text{features.before.}b\} \cup x_{\mathrm{ab}}) - \mathrm{val}(\{\text{features.before.}b\}).$$

Shapley values are considerably the best to predict the importance of features based on its calculation, and Fig. 7 shows the feature importance as seen from the Shapley perspective. It is interesting to have a comparison of this technique with the permutation importance, and the authors consider that Shapley is a better metric since permutation-based feature importance computes based on the degradation in model performance whilst Shapley computations account for magnitude of feature attributions.

Fig. 7 Shapley values-based feature importance

Fig. 8 High quality wine parameters through shapley

The Shapley values allow for several interactive plots like the one plotted in Fig. 8, which shows the components of a "high" quality (quality >5) wine in terms of the features along with the relative contribution of each feature to the quality score.

4 Results

The first glance at the various outputs obtained through several interpretability models throw surprising variations on the relative feature importance and have been shown in Table 1. The most interesting outcome of this tabulation is a high degree of similarity that exists between the Shapley value predictions and the average of all other methods *minus* the basic heatmap approach. This relationship between results can also be understood in context of the ensemble-based approaches of ML for an easy perspective or analogy.

Table 1 Feature ranking by different models for quality prediction

Feature	Heatmap	XGB feature by weight	XGB feature by coverage	Permutation importance (Eli5)	Average without heatmap	SHAP
Alcohol	1	4	1	1	1	1
pH	2	7	10	10	9	11
Sulphates	3	10	7	9	8	8
Free sulphur dioxide	4	2	5	3	3	3
Citric acid	5	6	6	5	5	6
Residual sugar	6	5	8	4	6	4
Fixed acidity	7	9	11	11	10	10
Total sulphur dioxide	8	7	2	8	5	9
Volatile acidity	9	1	4	2	2	2
Chlorides	10	8	9	7	8	7
Density	11	3	3	6	4	5

Table 2 Comparison of interpretability methods, with special reference to speed of execution

Method	Native feature importance	Eli5 (permutation importance)	Partial dependence plots	LIME	SHAP
Type	Global	global: black box models/local: white box models	Global	Local	Hybrid
Consistency	Inconsistent	Consistent	Consistent	Consistent	Consistent
Speed	Fast	Slow with more features	Fast	Fast	Moderate
Reliability	Low	Moderate	Moderate, increase with ICE	High	High
Model-agnostic	No	Partially	Yes	Yes	Yes

The findings indicate that although none of the individual interpretability models can exclusively guarantee a great insight into the behaviour of the ML model, and there has to be a back and forth approach of going through various methods to uncover the real insights. Table 2 lists the basic differences between models based on the literature, along with the findings of the study on the comparative speed of operation of each calculated as a function of time.

5 Conclusion and Future Work

The ML scientists always try to strike a balancing role between bias and variance for accuracy of the models, and there always exists a trade-off between the two. The study of this paper indicates that a trade-off in interpretability to accuracy also exists if we consider the interpretability of the models. There exists a direct relationship between the accuracy of a model to its complexity, which is a result of the using multiple stacked models or ensemble-based boosting, but the two are inversely proportional to interpretability with each level of a model addition making it increasingly difficult to understand its outputs.

It can also be concluded that any model with a variable number of parameters like XGBoost is harder to understand since the parameters tend to increase with the volume of training data. Parametric models on the other hand can provide easier insights with coefficient values for interpretability.

The most important takeaway from the study is that none of the interpretability models is an absolute authority, and any assumption made by looking at the output of a single interpretability technique can (and most certainly will) go absolutely wrong with its dependency not only on the concerned model, but also with the values

available in the raw data. The scope of interpretability and explainability of the ML models in the presence of hyperparameter tuning with complex models remains an interesting research area and is expected to precipitate further research in associated domains.

References

1. N. Burkart, M. F. Huber, "A survey on the explainability of supervised machine learning," *Journal of Artificial Intelligence Research*, *70* (2021). https://doi.org/10.1613/JAIR.1.12228
2. A. Datta, M. Fredrikson, K. Leino, K. Lu, S. Sen, Z. Wang, "*Machine Learning Explainability and Robustness*," (2021). https://doi.org/10.1145/3447548.3470806
3. C. Rudin, J. Radin, Why are we using black box models in ai when we don't need to? a lesson from an explainable ai competition. *Harvard Data Science Review*, *1*(2) (2019). https://doi.org/10.1162/99608f92.5a8a3a3d
4. "Explainable Machine Learning Challenge", FICO Community, Available: https://community.fico.com/s/explainable-machine-learning-challenge. [Accessed: 03-Sep-2021]
5. A. Vellido, The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, *32*(24) (2020). https://doi.org/10.1007/s00521-019-04051-w
6. T. Miller, Explanation in Artificial Intelligence: Insights from the social sciences. Artif. Intell. **267**, 1–38 (2018). https://doi.org/10.1016/j.artint.2018.07.007
7. B. Kim, R. Khanna, O.O. Koyejo, Examples are not enough, learn to criticize! Criticism for interpretability. In *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, USA, 2016), pp. 2280–2288
8. F. Doshi-Velez, B. Kim, *Towards a Rigorous Science of Interpretable Machine Learning*. arXiv 2017, arXiv:1702.08608
9. M.T. Ribeiro, S. Singh, C. Guestrin, *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (San Francisco, CA, USA, 2016), pp. 1135–1144
10. European Commission. General Data Protection Regulation (2016). Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. Accessed 03 Sep 2021
11. R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models. ACM Comput. Surv. **51**(5), 1–42 (2018)
12. M. W. Craven, J. W. Shavlik, Extracting tree-structured representations of trained networks. In Neural Information Processing Systems (Cambridge, MA: MIT Press, 1995), pp. 24–30
13. N. Frosst, G. E. Hinton, Distilling a neural network into a soft decision tree. In Proceedings of the First International Workshop on Comprehensibility and Explanation in AI and ML 2017 Colocated with 16th International Conference of the Italian Association for Arti-ficial Intelligence (AI*IA 2017). CEUR Workshop Proceedings, vol 2071 (2017)
14. B. Kim, C. Rudin, J. Shah, The Bayesian case model: a generative approach for case-based reasoning and prototype classification. In Proceedings of the 27th International Conference on Neural Information Processing Systems, vol 2 (Cambridge, MA, MIT Press, 2014), pp. 1952–1960
15. M. T. Ribeiro, S. Singh, C. Guestrin, *Anchors: High-precision model-agnostic explanations. InAAAI* (NewOrleans, Louisiana, AAAI Press, 2018), pp. 1527–1535
16. A. Fisher, C. Rudin, F. Dominici, *Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class, from the "Rashomon" Perspective*. arXiv 2018, arXiv:1801.01489
17. J.H. Friedman, Greedy function approximation: A gradient boosting machine. Ann. Stat. **29**, 1189–1232 (2001)

18. A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. J. Comput. Gr. Stat. **24**, 44–65 (2015)
19. S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, USA, 2017), pp. 4765–4774
20. P. Cortez, A. Cerdeira, F. Almeida, T. Matos, J. Reis, Modeling wine preferences by data mining from physicochemical properties. Decision Support Syst., Elsevier **47**(4), 547–553 (2009)

An Improved AODV Routing Algorithm for Detection of Wormhole and Sybil Attacks in MANET

Shreeya Mishra, Umesh Kumar, and Komal Mehta Bhagat

Abstract Wireless sensor network is a promising technology in the current scenario due to its wider area of research. With the advancement of technology, communication in mobile ad hoc networks has become easier and efficient with some vulnerability. Secured system communication is required, but with the evolution and advancements in the technology, threats also increase. During communication, various kinds of attacks may occur in the mobile ad hoc network. In this paper, wormhole attack in AODV protocol is discussed. Different parameters affect the working of protocols; here parameters end-to-end delay and PDR ratio are discussed with the main aim to detect the wormhole attack and provide an efficient solution to minimize the risk of attack.

Keywords WSN · MANET · Wormhole attack · Sybil attack · End-to-end delay

1 Introduction

A wireless sensor network consists of a large number of nodes that gather information from the environment and transmit it to the base station via wireless signals [1]. Due to unstable environment and unattended deployment, WSN is oversensitive in failures [2]. In WSN, node movement is a centralized type. In a wireless network, mobile ad hoc networks (MANET) is a self-configured and small-range instant network, where wireless devices directly communicate with each other [3]. MANET network is designed anywhere and anytime in such a way to support the mobility of users in the network. Such networks serve blocking [4]; therefore, the performance depends

S. Mishra (✉) · U. Kumar
Govt. Women Engineering College, Ajmer, India
e-mail: shreeyamishra9@gmail.com

U. Kumar
e-mail: umesh@gweca.ac.in

K. M. Bhagat
Bhagwan Parshuram Institute of Technology, Delhi, India
e-mail: komalbhagat@bpitindia.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_60

on the stability of the network architecture [5]. MANET is a decentralized type of WSN. WSN interacts with the environment, while MANET is closure to human interaction. MANET has security and routing problems because of its complexity in nodes establishment, and also nodes need to be adaptable in network topologies as an ad hoc network changes their topology quickly and unpredictably [6, 7].

AODV in MANET, where a device sends a packet to another device for communication, first looks up a route to the destination in its routing table if this exists in the data packet; otherwise, it starts the route discovery process. In this case, the device will send an RREQ message in broadcast mode. The RREQ message has the source address, a destination address, source sequence number, broadcast identifier, and the most recent sequence number between the source device and the destination device. The device will be waiting for a reset route message (RREP) message to be received. If an RREP message is not received after a fixed time, the device will wait for the RREP message to be received. The site will try again a finite number of times. If it fails again, the device concluded that there is no route to that destination. When a device receives an RREQ message for a destination, one of the following actions may occur: The device that received the RREQ message has no information about the error, so it resends the message to its neighboring devices via broadcast. It also temporarily stores the route back to the destination device [8]. The device that receives the message is the destination device or has route information to destination t. In either case, it sends an RREP message back to the source device via the route by which the device received the first RREQ message. All devices maintain a routing table with the information of each destination of interest, created from the received RREQ messages, and remain active if used by any active neighboring device. Therefore, author [9] proposed a new protocol that is WAODV (waiting AODV) in which the most stable path between source and destination is chosen by using hello packets to the intermediate node to find the most stable neighbor and thus path. TAODV [10] is the extension form of AODV. It uses trust values and thus sends three messages trust request message (TREQ), trust reply message (TREP), and trust warning message (TWARN).

In a wireless sensor network, different types of attacks can be seen. Attacks are classified as active, passive, internal or external, different protocol layers, cryptography, or non-cryptography related [11]. Active Attacks create a hole in following security issues in protocols, sensor node destruction, and data alteration, which degrades the performance of the WSN protocol ([12], [13]). Gray hole attack and black hole attack are like each other. In the gray hole with a certain probability, it drops the packets, due to this it is difficult to detect and also causes network congestion [14].

Sybil attack: In a network, present malicious nodes create multiple identities to the other node. This type of attack is common in MANET [15].

Wormhole attack: The most threatened type of attack seen in the MANET is a wormhole attack. It takes place in the form of forwarding the packets in between the malicious nodes [16]. In wormhole attack work, two malicious nodes are present: One malicious node gets the packet from the source and forwards to next node which seems to be malicious node, and it further forwards to the destination [17].

2 Literature Survey

MANET has been analyzed by the authors for the detection of attacks and protocols that have been used. AODV protocol in MANETs is designed to choose the least congested path between source and destination, but choosing the least path is not enough as it does not guarantee a stable path. Routes are only maintained when they are active that is they are of use. During routing, security threats are present. In [18], author detects the black hole attack in AODV protocol and compares the parameters like throughput and PDR with traditional AODV. Authors [19] have discussed different detection and prevention methods of wormhole attack along with merits and demerits. Authors [20] suggest a method for wormhole attack called (DELPHI) delay per hop indication. This method calculates the mean delay per path of every path by using the multipath approach. In ad hoc on-demand multipath distance vector (AOMDV), an extension of AODV is proposed by the authors [21]. In this protocol, the RTT method is used in which the RTT value from source to destination is calculated for each route.

Authors [22] proposed a secure tracking of node encounters (SECTOR). In this protocol, based on the speed of transmitted data, distance between two nodes is calculated. In [23], authors suggest a method to detect wormhole attacks by applying routing protocol IPv6 for low-power and lossy networks. To detect wormhole attack on the mechanism on packet leashes is suggested by the authors [24]. This method considers geographical-based protocol which relies on the current location and transmission time of the packet and in temporal leashes clock is tightly synchronized and it is not dependent on the GPS information. Authors [25] proposed a mechanism based on directional antennas that prevent from wormhole attack. In this method after hearing a signal from a neighbor with the assumption that the antennas on all the nodes are aligned, the neighbor list is built securely. Distance consistency-based secure localization scheme is proposed by the authors [26] to detect wormhole attack. To avoid Sybil attack, author [27] collects the information of routes by using an intelligence algorithm. In this algorithm, a Sybil attack is detected when there is an energy change in an active network.

3 Proposed Work

The most common attack in MANET is a wormhole attack in a network, and the malicious node between source and destination creates s tunnel and affects the transmission of the packets. The proposed system is the modified version of AODV. An exploration of the possible control measures to reduce the risk of exploitation of these vulnerabilities by wormhole attacks is presented. Flowchart is given in Fig. 1

The simulation has been done in NS2 simulator tool. The tool helps to study the dynamic nature of communication in networks. It helps to design a network protocol by providing a comprehensive environment; under specific conditions, it creates

Fig. 1 Flowchart of proposed algorithm

virtual scenario and helps to analyzing their performance. Here in the network, 50 nodes are taken, duration for simulation was 90 s, and other parameters are listed in Table 1. In this subsection, simulations results show the effect on data communication when the number of attackers is increased in the network. Figure 1 shows the working of proposed work.

Table 2 shows the effect of an increase in the number of attackers on end-to-end delay when compared to traditional AODV, WAODV, TAODV, and proposed routing protocols in MANETs.

Table 1 Simulation parameters

Parameters	Value
Simulator	NS2
Simulation time	90 s
Topology	800 m X, 800 m Y
No. of nodes	25,50,75,100,125
Max segment size	512 bytes
Traffic type	FTP
Routing protocol	AODV, WAODV, TAODV

Table 2 Average end-to-end delay with number of attackers in wormhole attack

Nodes	Proposed work	AODV	WAODV	TAODV
25	9.18	26.52	72.56	11.23
50	12.45	21.18	82.43	14.31
75	11.34	33.45	78.34	13.52
100	11.21	36.34	74.05	12.23
125	10.52	29.21	69.36	11.44

Increasing the number of attackers in a network has an impact on data communication, as we will see in this section's simulation findings. To illustrate this, Fig. 2 compares the end-to-end latency for the standard AODV, WAODV, TAODV, and proposed MANET routing protocols to those for an increase in the number of attackers.

Increasing the number of attackers in the network results in an increase in the average packet delivery ratios (PDR) of all the compared routing methods (Fig. 3). In other words, as the number of attackers on the network grows, so does their likelihood of joining a particular assault path. As a result, attackers are dropping more packets throughout the transmission process. When using the TAODV protocol, the attacker

Fig. 2 End-to-end delays with increase in number of attackers in the network

Fig. 3 Average PDR with increase in number of attackers in the network

Fig. 4 Average end-to-end delays performance in Sybil attack

utilizes false information during the route discovery phase to ensure that he will be a part of every route found during the route discovery process, but the proposed approach loses the least amount of data when using the WAODV attack protocol.

Figure 4 depicts the impact of increasing the number of Sybil attack nodes in a MANET's protocols. The end-to-end latency of the network increases the delay of the packet as the number of attackers grows. EED is still higher than the proposed procedures in AODV because more attackers reach their objective and create a greater average delay (Table 4).

According to Fig. 5 for AODV and suggested work routing protocols, increased network and the PDR of the compared routing methods show that the PDR of the proposed protocol is much higher than that of the AODV routing protocols (Table 5 and Fig. 5).

Table 3 Packet delivery ratio

Nodes	Proposed work	AODV	WAODV	TAODV
25	9.18	26.52	72.56	11.23
50	12.45	21.18	82.43	14.31
75	11.34	33.45	78.34	13.52
100	11.21	36.34	74.05	12.23
125	10.52	29.21	69.36	11.44

Table 4 Average end-to-end delays performance in Sybil attack

Nodes	Base paper [38]	Proposed protocol
10	0.4	0.38
20	0.5	0.43
30	0.6	0.53
40	0.8	0.63
50	0.9	0.83

Fig. 5 Effect of packet-on-packet delivery ratio in Sybil attack

Table 5 Effect of packet-on-packet delivery ratio in Sybil attack

Time	No of packets Base paper [27]	Proposed work
30	43	52
60	47	67
90	50	73
120	53	86
150	57	97

4 Conclusion

MANETs are spontaneously deployed mobile networks that do not require any preexisting infrastructure over a geographically restricted area. Without the need for centralized management or set network architecture, nodes are often both autonomous and self-organized. MANETs are prone to a specific routing error known as a wormhole attack and Sybil attack because of their decentralized structure. Malicious nodes tunnel traffic from their locations to the target node through an attack called a wormhole and malicious which creates multiples fake IDs called Sybil. A fake route with a lower hop count is created as a result of wormhole assaults. This false route allows bad nodes to deliver or drop packets depending on how they feel about the originating node. The purpose of this piece is to put an end to these kinds of attacks. When it comes to defending against wormhole attacks, it investigates the usage of an artificial immune system (AIS). The suggested method detects and bypasses wormhole nodes quickly and without degrading the network's overall performance. We compare the suggested technique to alternative solutions in terms of packet delivery ratio and end-to-end latency to see how well it performs. In a simulation, the proposed technique outperforms by 8% in the term of packet delivery ratio by previous schemes in preventing a wormhole attack and Sybil attack.

References

1. J. Tian, et al., "Study on wireless sensor networks." 2010 International Conference on Intelligent System Design and Engineering Application, vol. 2 (IEEE, 2010)
2. N.N. Datta, K. Gopinath, A survey of routing algorithms for wireless sensor networks. J. Indian Inst. Sci. **86**(6), 569 (2006)
3. T.H. Hadi, MANET and WSN: what makes them different? Int. J. Comput. Netw. Wirel. Commun. (IJCNWC) **7**(6), 23–26 (2017)
4. Dr. S. S. Aarti, "Tyagi,"Study of manet: Characteristics, challenges, application and security attacks". Int. J. Adv. Res. Comput. Sci. Softw. Eng. **3**(5), 252–257 (2013)
5. S. A. K. Al-Omari, P. Sumari, "An overview of mobile ad hoc networks for the existing protocols and applications." arXiv preprint arXiv:1003.3565 (2010)
6. A.O. Bang, P.L. Ramteke, MANET: history, challenges and applications. Int. J. Appl. Innovation Eng. Manage. (IJAIEM) **2**(9), 249–251 (2013)
7. G. Wang, G. Wang, An energy-Aware geographic routing protocol for mobile ad hoc networks. Int. J. Softw. Inf. **4**(2), 183–196 (2010)
8. P.K. Maurya et al., An overview of AODV routing protocol. Int. J. Modern Eng. Res. (IJMER) **2**(3), 728–732 (2012)
9. T. Jamal, S.A. Butt, Malicious node analysis in MANETS. Int. J. Inf. Technol. **11**(4), 859–867 (2019)
10. X. Li, M.R. Lyu, J. Liu, "A trust model based routing protocol for secure ad hoc networks." 2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720) vol 2 (IEEE, 2004)
11. R. Hunt, "Network security: the principles of threats, attacks and intrusions, part1 and part 2." (2004)
12. M. Karthigha, L. Latha, K. Sripriyan, "A comprehensive survey of routing attacks in wireless mobile ad hoc networks." 2020 International Conference on Inventive Computation Technologies (ICICT) (IEEE, 2020)
13. L.K. Bysani, A. K. Turuk, "A survey on selective forwarding attack in wireless sensor networks." 2011 International Conference on Devices and Communications (ICDeCom) (IEEE, 2011)
14. M. Sathish et al., "Detection of single and collaborative black hole attack in MANET." 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (IEEE, 2016)
15. SHIATS, Allahabad. "A survey on comparison of secure routing protocols in wireless sensor networks." Int. J. **5**(3) (2016)
16. N. Choudhary, L. Tharani, "A survey of routing attacks in mobile ad hoc network." IEEE Security in Wireless Mobile AD Hoc and Sensor Networks, pp. 85–91 (2007)
17. B. Kannhavong et al., A survey of routing attacks in mobile ad hoc networks. IEEE Wireless Commun. **14**(5), 85–91 (2007)
18. A. Sharma et al., Prevention of black hole attack in AODV routing algorithm of MANET using trust based computing. Int. J. Comput. Sci. Inf. Technol. **5**(4), 5021–5025 (2014)
19. M.K. Verma, R.K. Dwivedi, "A survey on wormhole attack detection and prevention techniques in wireless sensor networks." 2020 International Conference on Electrical and Electronics Engineering (ICE3) (IEEE, 2020)
20. H.S. Chiu, K.S. Lui, "DelPHI: wormhole detection mechanism for ad hoc wireless networks." 2006 1st international symposium on Wireless pervasive computing (IEEE, 2006)
21. P. Amish, V.B. Vaghela, Detection and prevention of wormhole attack in wireless sensor network using AOMDV protocol. Procedia Comput. Sci. **79**, 700–707 (2016)
22. S. Čapkun, L. Buttyán, J.P. Hubaux, "SECTOR: secure tracking of node encounters in multi-hop wireless networks." Proceedings of the 1st ACM workshop on Security of ad hoc and sensor networks (2003)
23. G.-H. Lai, Detection of wormhole attacks on IPv6 mobility-based wireless sensor network. EURASIP J. Wirel. Commun. Netw. **2016**(1), 1–11 (2016)

24. Y.-C. Hu, A. Perrig, D.B. Johnson, Wormhole attacks in wireless networks. IEEE J. Sel. Areas Commun. **24**(2), 370–380 (2006)
25. L. Hu, D. Evans, "Using directional antennas to prevent wormhole attacks." NDSS. **4** (2004)
26. H. Chen et al., A secure localization approach against wormhole attacks using distance consistency. EURASIP J. Wireless Commun. Netw. **8**, 2010 (2010)
27. W. Jlassi et al., "A novel dynamic authentication method for sybil attack detection and prevention in wireless sensor networks." AINA (1) (2021)

An Ensemble Model (Simple Average) for Malaria Cases in North India

Kumar Shashvat, Arshpreet Kaur, Ranjan, and Vartika

Abstract Malaria is an infectious disease borne due to mosquitoes that attacks humans and other animals' bodies. Malaria is a part of the plasmodium group caused by single-celled microorganisms. This study proposes the use of ensemble model using the three regression algorithms that are linear regression, support vector machine (SVM), and auto-Arima techniques and comparing their results. Predictions of plasmodium virus cases are made with the use of linear regression, support vector machine, and auto-Arima algorithms. The accuracy of prediction is measured by calculating the explained variance score, mean squared error rate, and root mean squared error rate. Our aim is to get better prediction results compared to the individual algorithms by combining the results of these individual models. The proposed work determines the accuracy of linear regression, support vector machine, and auto-Arima and ensembles together to find the trend of prediction using simple Average. A comparison of performance among the three regression techniques indicated the SVM model performs the best and has small RMSE and MAE values. But, by introducing the technique of ensemble modeling using simple average, combining the prediction of these three algorithms results in the lowest RMSE and MAE values.

Keywords Malaria · Support vector machine · Linear regression

1 Introduction

For centuries, malaria has become a problematic disease in India. Malaria is a mosquito-borne disease that causes infection and affects both humans and animals [1–6]. It took ten to fifteen days to show the symptoms after an infected mosquito bit the human body. It should be treated properly; otherwise, the effects of the malaria will

K. Shashvat (✉) · A. Kaur · Ranjan · Vartika
Department of Computer Science and Engineering (ACED), Alliance University, Bangalore, Karnataka, India
e-mail: shashvat.sharma@alliance.edu.in

A. Kaur
e-mail: arshpreet.kaur@alliance.edu.in; arshpreet@nitdelhi.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_61

recur months later. Malaria is a part of the plasmodium group caused by single-celled microorganisms [7–14].

The diseases can be eliminated by using bugs spray and mosquito nets to prevent them from biting humans. It can also be controlled by taking measures by spraying insecticides and draining collected water and having a proper drainage system [2, 15–20]. In the 1930s, there was no single person in the country who was not infected by the disease. Due to malaria, the loss in economy in man-days was at Rs.10,000 million per year in 1935 [20–22].

The disease has majorly affected the tropical and subtropical region of India. Malaria is majorly associated with poverty and due to this economic development have negative effects. Malaria is categorized into severe or uncomplicated by World Health Organization (WHO) [22–24]. By the year 2020, there is only one vaccine that reduces the chances of malaria by 40% in children. From 2010 to 2014 that rate of malaria decreased, but from 2015 to 2017, the cases of malaria reached 231 million. Malaria is a disease associated not only with poverty, but the affirmation has suggested that it has a major hindrance to economic development [25, 26]. The Malaria Eradication Research Agenda (malERA) drive was an instructive process to rectify which parts of research and development (R&D) must be marked for worldwide elimination of malaria. India is considered traditionally native for both plasmodium vivax and plasmodium *falciparum* malaria and has a past of successes [27, 28]. Over 75% cases of malaria come from India of the total cases in Southeast Asia. In 2016, over 50% of India (698 million) was in danger of getting infected by malaria as mentioned by the World Malaria Report, which also stated that India reckon 6% of total malaria cases, 6% deaths, and 51% of the total malaria cases. In 2017, India ranked 46th with a death ratio of 1.97 per 100,000 individuals. The health of the disease is a complex issue in India developing from the geo-ecological diversity of the country [26, 27]. 80% of cases in India is caused by high indigenous areas. 95% of the population of the country lives in areas where the chances of getting infected is high [27–29].

This paper uses various techniques like support vector regression, linear regression, and Arima models to come up with a comparative analysis and see the effect and trend of the study. To improve the accuracy of our model, we combine the results from these techniques using a simple ensemble model. The aim of this paper is to analyze the prediction modeling results by overcoming the prediction error of different models. The performance of the model is evaluated in terms of mean average error, mean square error, explained variance score, and root mean square error. Based on the error rates and scores, the proposed model performs better than any of the individual techniques.

Month	2013	2014	2015	2016	2017
1	1	1	0	0	2
2	0	1	0	0	1
3	0	2	1	2	0
4	0	3	2	4	6
5	5	6	0	5	14
6	19	10	9	27	19
7	28	15	8	27	29
8	27	26	33	54	40
9	17	28	37	23	30
10	9	5	19	6	7
11	2	5	5	2	3
12	4	3	4	0	0

Fig. 1 Malaria dataset

2 Materials and Methods

We undertook detailed experimental and evaluator steps to test the effectiveness of the proposed model. For our experiment, we use a malaria dataset. The dataset is having raw data and stationary data.

2.1 Dataset

The dataset holds the monthly data from January to December for five years that is from 2013 to 2017 and the respective number of malaria positive. This dataset contains raw and stationary data with a total number of 60 inputs (Fig. 1).

The dataset is divided such that 85% of the dataset is used to train the model, and 15% is used to test the dataset to get the performance of the model. The dataset contains two columns month and PV where PV denotes the no. of cases in each month from year 2013 to 2017.

2.2 Model Architecture

The flowchart in Fig. 2 represents the whole architecture of the model which consists of the use of three regression algorithms that is linear regression, support vector machine, and auto-Arima model. The model uses the dataset of malaria which contains the total number of positive cases of malaria from the year 2013 to 2017 for each month that is January to December. The proposed model contains the use of linear regression, support vector machine, and auto-Arima algorithms. The accuracy of prediction is measured through calculating the explained variance score, root mean squared error, and mean squared error. After applying all these models, we took the average of the prediction from all the three regression models and ensemble them together to get the better prediction and less RMSE value. Ensemble modeling is a technique which combines the results of different models to improve the overall

Fig. 2 Model architecture

performance. The performance of our model is evaluated in terms of mean average error, mean square error, root mean square error, and explained variance score.

2.3 Models Applied

(A) **Linear Regression**: The term linearity in linear regression refers to the relationship which is linear between two variables. Linear regression is used to determine the value of a dependent variable (y) based on a given independent variable (x). So, regression techniques are used to find linear relation between independent and dependent variables. Figure 3 represents the independent variable (x) on the x-axis and dependent variable (y) on the y-axis, which gives a straight linear line relationship between x and y that best fits the data points. The equation

$$y = a + b * x,$$

Fig. 3 Linear regression plot from Google Image

Fig. 4 Support vector machine plot from Google Image

where y is the dependent variable, x is the independent variable, and b is the slope of the line that intercept the y-axis.

(B) **Support Vector Regression**: Support vector machine, commonly known as SVM, is a supervised machine learning algorithm that is helpful for both regression problems and classification problems. Support vector machine aims to find a hyperplane in N-dimension space to classify the data points. There can be a number of hyperplanes that can be chosen to separate two classes of data points. The aim of support vector machine is to find a plane which contains the maximum number of margins. The data points should have maximum distance from the margin to provide some future data points correctly. Hyperplanes are decision boundaries that help classify the data points (Fig. 4).

(C) **Arima Model**: Autoregressive integrated moving average abbreviated as ARIMA is a model that defines a time series which is based on the past experiences that is it predicts the new values based on what it learns from previous time series values. The Arima model is used to forecast future values. The Arima model describes the correlation between data points. An Arima model is described by three terms: p, d, q where p is the order of the autoregressive term, q is the order of the moving average term, and d is the number of differencing required to make the time series stationary (Fig. 5).

(D) **Ensemble Model Using Simple Average**: Ensemble modeling is a technique of combining different types of analytical models and then predicting their results in combined score in regards to maximize the accuracy of predictive analytics. In this method, we take an average of results from all the models

Fig. 5 Arima model plot

and use it to make the final prediction. Algorithm for ensemble learning the above models:

1. Input: Independent variables set (Features) X_i where i is 1, 2, 3…n. Output: Predicting for PV.
2. Using regression on the independent variables, predicting using SVM, LR, and auto-Arima.
3. Take the simple average of the prediction values from each individual model.
4. Use the average of predictions obtained from all three models to compute the final prediction values using the ensemble modeling.
5. Comparison of individual prediction rate with ensemble model error rate.

3 Results and Discussion

Monthly indices of malaria cases in the North India Region from year 2013 to 2017 show an increment of cases from June to September. The cases were higher for these months than the other months in each year.

The model is trained on 85% of the dataset for each model and tested on 15% of data. The correctness of the model is determined by noting the RMSE, MAE, explained variance score's values. Table 1 shows the results of the three regression algorithms that are linear regression, support vector machine, and Arima model.

Table 1 Comparison among prediction models to monthly data of malaria cases in North India region from year 2013 to 2017

Methods	MAE	MSE	RMSE	Explained variance score
Linear regression	5.96	48.16	6.93	0.68
Support vector machine	4.55	36.77	6.06	0.77
Arima model	5.24	52.48	7.24	0.72
Ensemble model using simple averaging	4.06	18.09	4.25	0.89

A comparison of performance among the three regression techniques indicated the SVM model performed the best and has small RMSE and MAE values. But, by introducing the technique of ensemble modeling using simple average, combining the prediction of these three algorithms results in the lowest RMSE and MAE values. In this method, we take an average of predictions from all the models and use it to make the final prediction. Table 1 shows that ensemble model using a simple average performs the best over the given dataset.

Figure 6 shows the plot for the linear regression algorithm and for the given malaria dataset on a monthly basis, and the results for this regression algorithm comes out to be 5.96 MAE, 48.16 MSE, 6.93 RMSE, and variance score is 0.68. The results show that linear regression does not give accurate predictions. Figure 7 shows the plot for the SVM model for which MAE is 4.55, MSE is 36.77, RMSE is 6.06, and variance score is 0.77 which show that the model performs best in all these regression algorithms. Figure 8 shows a plot for the Arima model for which MAE is 5.24, MSE

Fig. 6 Prediction for linear regression model

Fig. 7 Prediction for SVM model

Fig. 8 Prediction for Arima model for malaria dataset where blue curve shows the training set, the yellow curve is the validation set, and the green curve is the prediction set for the given data

is 52.48, RMSE is 7.24, and variance score is 0.72 which perform quite well for the given dataset.

A comparison of performance among the three regression techniques indicated the SVM model performed the best and has small RMSE and MAE values. But, by introducing the technique of ensemble modeling, combining the prediction of these three algorithms results in the lowest RMSE with 4.25, MAE with 4.06, MSE with 18.09, and variance score with 0.89 values. Table 1 shows that ensemble model using a simple average performs the best over the given dataset.

Malaria is one of the most common tropical diseases in Northern parts of India having the most number of cases in the month of June–September due to the monsoon season. Malaria is caused due to the mosquito biting on humans and animals. Poor drainage and sanitation play a major role in the spread of disease.

In the proposed methodology, we contrast the data of malaria cases in the North India region from 2013 to 2017 using the three regression and time series models, and then, we ensembled them together by taking the simple average of their prediction and found that the ensemble model gave the optimal correct predictions and lowest error rates compared to other techniques. It was also noted that malaria cases were highest in the mid of every year.

4 Conclusion

In the proposed study, we contrast the data of malaria cases in the North India region from 2013 to 2017 using the three regression and time series models, and then, we ensemble them together by taking the simple average of their prediction and found that the ensemble model gave the optimal correct predictions and lowest error rates compared to other techniques. It was also noted that malaria cases were highest in the mid of every year.

However, it is to be remembered the merits and limitations of the three regression models for which the model predicts the optimal results. The linear regression makes estimation procedure simple and easy to understand and interpret the linear equation.

Linear regression is prone to overfitting when the data are linearly separable, but in reality, the data are not separable, and the straight line relationship between independent and dependent variables is not true every time. Also, it is prone to outliers. The SVM model provides an optimal solution and allows mapping over the data, but when the dataset becomes large, it requires a large amount of tuning parameters and creates complexity. Arima model is a very popular model and is suitable for all types of time series data. But, if the data are insufficient, then it will not predict accurate results. Future studies should be able to improve these limitations. Also, this model can be used to predict the results for other public health diseases like typhoid. Furthermore, the study should encourage other researchers to carry out different tests on the different time series predictions that explains alertness for the study.

References

1. Malaria, South East Asian Quinine Artesunate, Artesunate versus quinine for treatment of severe falciparum malaria: a randomised trial. The Lancet **366**(9487), 717–725 (2005)
2. M. Poostchi, K. Silamut, R.J. Maude, S. Jaeger, G. Thoma, Image analysis and machine learning for detecting malaria. Transl. Res. **194**, 36–55 (2018)
3. V. Sharma et al., Malaria outbreak prediction model using machine learning. Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET) **4**(12) (2015)
4. P. Kumar Rai, M.S. Nathawat, M. Onagh, Application of multiple linear regression model through GIS and remote sensing for malaria mapping in Varanasi District, India (2014)
5. R. Verma et al., Identification of proteins secreted by malaria parasite into erythrocyte using SVM and PSSM profiles. BMC Bioinform. **9**(1), 201 (2008)
6. Ch. Sudheer, S.K. Sohani, D. Kumar, A. Malik, B.R. Chahar, A.K. Nema, B.K. Panigrahi, R.C. Dhiman, A support vector machine-firefly algorithm based forecasting model to determine malaria transmission. Neurocomputing **129**, 279–288 (2014)
7. S. Suryawanshi, V.V. Dixit, Comparative study of Malaria parasite detection using euclidean distance classifier & SVM. Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET) **2**(11), 2994–2997 (2013)
8. R. Anokye, E. Acheampong, I. Owusu, E. Isaac Obeng, Time series analysis of malaria in Kumasi: using ARIMA models to forecast future incidence. Cogent Soc. Sci. **4**(1), 1461544 (2018)
9. M.I. Musa, Malaria disease distribution in Sudan using time series ARIMA model. Int. J. Public Health Sci. **4**(1), 7–16 (2015)
10. A.E. Jones, A.P. Morse, Application and validation of a seasonal ensemble prediction system using a dynamic malaria model. J. Clim. **23**(15), 4202–4215 (2010)
11. T. Smith, A. Ross, N. Maire, N. Chitnis, A. Studer, D. Hardy, ... M. Tanner, Ensemble modeling of the likely public health impact of a pre-erythrocytic malaria vaccine. PLoS Med **9**(1), e1001157 (2012)
12. K. Shashvat, R. Basu, A.P. Bhondekar, Application of time series methods for dengue cases in North India (Chandigarh). J. Public Health 1–9 (2019)
13. K. Shashvat et al., Comparison of time series models predicting trends in typhoid cases in northern India. Southeast Asian J. Trop. Med. Public Health **50**(2), 347–356 (2019)
14. K. Shashvat et al., A weighted ensemble model for prediction of infectious diseases. Current Pharm. Biotechnol. **20**(8), 674–678 (2019)
15. K. Shashvat, R. Basu, P.A. Bhondekar, A. Kaur, An ensemble model for forecasting infectious diseases in India. Trop. Biomed. **36**(4), 822–832 (2019)

16. O. Sagi, L. Rokach, Ensemble learning: a survey. Wiley Interdisc. Rev. Data Mining Knowl. Discov. **8**(4), e1249 (2018)
17. B. Krawczyk, L.L. Minku, J. Gama, J. Stefanowski, M. Woźniak, Ensemble learning for data stream analysis: a survey. Inf. Fusion **37**, 132–156 (2017)
18. H.M. Gomes, J.P. Barddal, F. Enembreck, A. Bifet, A survey on ensemble learning for data stream classification. ACM Comput. Surv. (CSUR) **50**(2), 1–36 (2017)
19. T. Go, J.H. Kim, H. Byeon, S.J. Lee, Machine learning-based in-line holographic sensing of unstained malaria-infected red blood cells. J. Biophotonics **11**(9), e201800101 (2018)
20. Z. Liang, A. Powell, I. Ersoy, M. Poostchi, K. Silamut, K. Palaniappan, P. Guo, M.A. Hossain, A. Sameer, R.J. Maude, J.X. Huang, CNN-based image analysis for malaria diagnosis. in *2016 IEEE International Conference on Bioinformatics and Biomedicine* (BIBM) (IEEE, 2016) pp. 493–496
21. B.N. Narayanan, R. Ali, R.C. Hardie, Performance analysis of machine learning and deep learning architectures for malaria detection on cell images. in Applications of Machine Learning, vol. 11139 (International Society for Optics and Photonics, 2019), p. 111390W
22. C. Ju, A. Bibaut, M. van der Laan, The relative performance of ensemble methods with deep convolutional neural networks for image classification. J. Appl. Stat. **45**(15), 2800–2818 (2018)
23. T. Strauss, M. Hanselmann, A. Junginger, H. Ulmer, Ensemble methods as a defense to adversarial perturbations against deep neural networks. arXiv preprint arXiv:1709.03423 (2017)
24. E.A. Amrieh, T. Hamtini, I. Aljarah, Mining educational data to predict student's academic performance using ensemble methods. Int. J. Database Theory Appl. **9**(8), 119–136 (2016)
25. P. Bühlmann, Bagging, boosting and ensemble methods. in *Handbook of Computational Statistics* (Springer, Berlin, Heidelberg, 2012), pp. 985–1022
26. T.G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Mach. Learn. **40**(2), 139–157 (2000)
27. D. Opitz, R. Maclin, Popular ensemble methods: An empirical study. J. Artif. Intell. Res. **11**, 169–198 (1999)
28. B. Yegnanarayana, *Artificial Neural Networks* (PHI Learning Pvt. Ltd., 2009)
29. A.I. McLeod, W.K. Li, Diagnostic checking ARMA time series models using squared-residual autocorrelations. J. Time Ser. Anal. **4**(4), 269–273 (1983)

IoT-Based Smart City Management Using IAS

Pranali Nitnaware, Supriya Sawwashere, Shrikant V. Sonekar, M. M. Baig, and Snehal Tembhurne

Abstract One of the most important factors in modern information systems of the Indian administrative agencies is the lack of easy access to information through organizations outside of the organization, with the name of the machine-guns in the app. In spite of the abundance of data, and they are locked into proprietary formats, or may not be available due to the lack of an API. In this report, we will focus on the architectural designs that make it easy to share data. Therefore, we focus on a data-based information, see the reference architecture, rather than focusing on the application and communication. Internet of Things (IoT) is a new technology that has emerged in the last couple of years, and it promises to be a continued explosion of data sources. With this technology, it will, in essence, a single "thing" to begin with, the production and transmission of data. For example, each and every lamp post, street lamp, trash bin, power transformer, etc., almost anything that a person can to be created in order to be constantly informed about their disease, their perceptions, and their patterns of use, etc. These data, which, in turn, can be used to lower the level of the business processes of the administration, in order to ensure fast and efficient value-added services. In this report, we will limit ourselves to the architectural design for the integration of new IoT-based systems in a Smart City infrastructure.

Keywords Internet of Things · APIs · Application · Smart City

1 Introduction

It was thought that the combination of data and information that are of critical importance for the development of a new smart app, and it is a core component of the concept of connected data. This means that there is a need to include some additional information about the data, metadata, which the data into a broader context. However, if the digital stream is enriched with additional information such as your device location, device, brand, and the display of the blocks of the intermediate calculations,

P. Nitnaware (✉) · S. Sawwashere · S. V. Sonekar · M. M. Baig · S. Tembhurne
J.D. College of Engineering and Management, Nagpur, Maharashtra, India
e-mail: pnitnaware95@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_62

Fig. 1 Main applications of the IoT

this allows for a richer and more meaningful interpretation of the data, especially for first-time users. Even though it is a technical point of view, it is not difficult to find this information, but in practice, there is no consensus on the standards and/or requirements for the provision of those people. It is a very important aspect of our research is to identify the possible data types (or other systems) that can be used for the Internet of Things. The current implementation of Smart City solutions as is often "isolated," meaning that each application has to provide an end-to-end implementation, which is self-contained and does not work with any other application. In the context of the Internet of Things (IoT) devices that can be integrated on the basis of geographical location and are evaluated with the aid of the analytical system. The sensor of services for the collection of certain data can be used in different types of projects such as monitoring, bikers, cars, and open parking areas. There are many applications in the service sector that is to make use of the Internet of Things, infrastructure, making the operation easier in terms of air pollution and noise, ease of transportation, cars, and video surveillance systems (Fig. 1).

2 Literature Review

2.1 Survey of Existing Methods

Jayti Bhatt, Jignesh Patolia, with the title "Real-Time water quality Monitoring System." In this article, it is stated that in order to ensure access to safe drinking water, the quality of the water needs to be monitored in real time, and an Internet of Things (IoT) is an approach that is based on the monitoring of the quality of the water has been proposed. In this paper, we present a project for the Internet of Things-based water quality monitoring system, which monitors water quality in real time. This system consists of a number of sensors that can measure water quality parameters such as pH, turbidity, conductivity, dissolved oxygen, and temperature. The measured values of the sensors can be processed by a microcontroller, and these

values can be remotely transmitted to the main controller, which is a Raspberry Pi with the help of the Zigbee protocol. Finally, the sensor data can be viewed via the app, Web browser, with the help of cloud-based computing.

Michael Lom, Ondrej Pribyl, Miroslav Svitek, under the title of "Industry 4.0", as part of the "Smart Cities." This article focuses on the relationship between the Smart City initiative and the concept of "Industry 4.0." The term "Smart City" has become a phenomenon in recent years, and it is very important, especially since 2008, when the world was hit by the financial crisis. The main reasons for the rise of the Smart City initiative to create a sustainable model of cities, towns, and to the maintenance of the quality of life for those who live in them. The subject of the "Smart City" is not only to be regarded as a technical discipline but in the economic, humanitarian, and legal aspects need to be touched for a while. The concept of Industry 4.0 and the Internet of Things (IoT) is going to be used for the creation of a so-called smart products. The sub-components of the product are in their own minds. Extra intelligence is being used in the manufacturing of the product, and in the post-process, a continuous monitoring of the entire life cycle of the product (intelligent processes. Another important aspect of the Industry 4.0 As a Web of services (iOS), which is true, in particular, to smart transportation and logistics, smart mobility, smart logistics, and energy and Internet access (full), which will determine how natural resources such as electricity, water, olive oil, etc.) correct gebruikt De IoT, iOS, feature, and full, which can be considered as a factor that can lead to a relationship between the Smart City and Industry 4.0—Industry 4.0 can be considered to be a part of the "Smart City."

Zhangwei Sun, Chi Harold Lee, Chatchik Bisdikian, Joel W., Branch, and Bo Yang, whose title is "QOI-Aware energy management in the Internet of Things and Sensory Environments." This article presents an efficient power management system, which provides a useful QOI experience a touch of class-based IoT environments. In contrast to previous efforts, in a transparent manner, and consistent with all of the primary protocols used, and continues to be the energy-efficiency in the long term, and without prejudice to any and reach QOI levels. In particular, the new QOI aware of the concept of the sensor is important for the job" specifically refers to the perceptual features of a touch sensor in the IoT environment, and the QOI requirements, as required by the task. A new concept of "critical transparency" is set up for each task, the selection of sensors for the tasks in the course of time. The power management solution is considered to be dynamic at run-time, to be optimal for the long-term, its traffic statistics, while reducing service time. Finally, to demonstrate the ideas and the algorithms proposed in this paper, a comprehensive sociological study, based on the use of sensor networks for the monitoring of the water levels, is presented, and the effectiveness of the proposed algorithms has been modeled [4].

3 Proposed Algorithm

When the results of a price reduction or assignment, there are multiple clustered machines, more logic programming, in order to access and aggregate the results from each machine, if necessary. In addition, units of work, which must be distributed (correlation) between the machines, so it does not affect the final results? For example, when the identification of the following sentences, a single sentence, it should not be split across multiple nodes, which may lead to recommendations to be split between nodes, and later omitted by the identification of the words in the sentence.

The map and the oppression of the system, and can vary in size, from a single machine to thousands of a cluster of computers, and in some enterprise-class business processes. In the C# programming language provides a set of thread-safe object that can be used quickly and easily to create a map of the reduction, the style, and the programs that are running. The following sections describe some of these objects and provide examples of how to implement a reliable, parallel-card-cutting processes with the help of them.

4 Research Methodology

IoT-end devices and can only respond to a specific vendor's software stack. This will kill the IoT device interactions. In the past, it was possible to create a "dummy" light bulb from a number of different providers; it is now possible to create a "smart" light bulb from just one supplier. This is mainly due to the lack of consistent data, the schema/data model for the IoT devices.

The data from the devices are stored on the system by the supplier in its own format. Therefore, the owner of the center of the city, it is not easy for you to share or re-use the information for any other purpose or to even make them for later use. Visualization/control panel, and device management app, has been closely linked to the devices. For example, it is not going to be easy to develop applications that will be passed to the same IoT devices. An analysis of the use cases of the Indian smart cities (see Appendix A), it appears that some of the sensors and, in particular, cameras, and it can be used in many different areas of application, provided that the tests are well-designed. The current approach to the implementation, it is difficult to do, unless the seller accepts the job.

One of the most important resources in the cities and towns of the data. Easy access to all the data, and will no doubt bring many benefits to the administration, in terms of the definition of "data-driven policy and decision-making." In addition, the access to the data that can drive innovation and new applications of the citizens (Fig. 2).

Since the beginning of the electronic era, many attempts have been made to convince the doctors to enhance the sounds of the heart, in the good faith belief that if the noise level cannot be increased, a greater diagnostic capability can be

Fig. 2 Block diagram

achieved. The rich sounds that can be heard by a physician, by means of a conventional stethoscope to occur at the time of the main valve are closed. In an abnormal heart rate, are there any other comments, it sounds. It had been a whisper is to be heard between the normal voices. The noise is most likely caused by a small opening of the valve, or through the hole in the septum, separates the right and left sides of the heart. Several doctors will be able to hear the same sound, but different interpret. This can lead to an incorrect diagnosis. In addition, it is the exact unit you will be able to reproduce the entire frequency range, the majority of which has been missed by a conventional stethoscope. A tool that was designed in order to use the full range of sound with a precision, and a digital stethoscope to the heart with the aid of appropriate equipment. The extracted signal is then fed to your computer and to detect cardiac abnormalities if any, measurements of physiological parameters, such as heart rate and respiratory rate, are of crucial importance in the medical field. Advances in technology allow for the different measurements for the continuous monitoring of.

Here is an easy way to do it is to measure your respiration with the aid of a displacement of the sensor. This device can be used to monitor your breathing, heart rate, using the appropriate sensor, and heart rate. The response is fast and efficient, as compared with conventional medical devices and equipment. With the help of it, the breath can be measured in a range that can be measured is in the range of 0–999 breaths per minute (Fig. 3).

5 Integrated System

Resource endpoints, providing access to a specific instance or a sub-set of data elements, assets, and will be dealt with by the data warehouses (or databases). For example, you can use these endpoints to request the status of a street lamp or the level of brightness value for the last week, in a particular street lamp. They will handle

Fig. 3 Trans-receiver section

the database queries in the background. It is a type of database that you use to store the resource values are to be specified in the metadata of the source in the catalog, and therefore, the corresponding APIs can be used in applications in order to access the information you need. This is the mechanism you can use to make a bunch for a variety of magazines. Each room is properly configured for the type of the resource. Most of the databases that provide a REST API to interact. Database technology is also likely to be adopted in the future (Figs. 4 and 5).

Fault detection of piezoelectric sensors can only be used for very fast processes under ambient conditions. In fact, there are a lot of applications, which show that the quasi-static measurements of the momentum effect; as an effect, it is located along the neutral axis (y), and the charge will be generated in the direction (x) perpendicular

Fig. 4 Working of sections

Fig. 5 Generated output

to the line of the armed forces. The amount of energy that depends on the geometrical dimensions of the respective piezoelectric element.

The automatic fuel pump control circuit is composed of a sensor, and a part was built up with the help of the op-amp IC LM324. Op-amps configured here as a comparator. Two of the hard disk drive, the copper wires are inserted in the soil, and to determine whether or not the soil is wet or dry. As the Arduino controller is used to control the entire system, as well as monitoring of the sensor, and if there are more than two of the sensors sense a dry condition, and after, the controller will shut off the engine and turn the engine off if any of the sensors are wet.

6 Patient Monitoring

The LM35 temperature sensor is to help us to get to know the patient's temperature, which has given us the knowledge of the patient's body temperature. The temperature will be, we can judge the condition of the patient, which will help you to decide properly.

A gas pump with the detection, geolocation, which is used in the data-based program that makes use of the supply chain management.

7 Result

Frequency response of a piezoelectric sensor; output voltage versus applied force.

8 Conclusion

The city should take a more active role to play and will have incentives to make it happen. For example, it is a guarantee of the minimum size, and duration of a service contract, at a lower price, of which the supplier is in very good promotions? In the city, you can detect a variety of prices, for a variety of guaranteed volume levels, and make a decision based on a realistic assessment of what can be done about it. It will also make sure that you use all the different government agencies to coordinate their actions, are much more likely than in the past. In any case, this is a pre-condition for the long-term vision of the Smart City mission.

References

1. V. Albino, U. Berardi, R.M. Dangelico, Smart cities: definitions, dimensions, performance, and initiatives. J. Urban Technol. **21** (2014)
2. L. Arrowsmith, Smart cities: business models, technologies, and existing projects. in *Information Technology Service Research Report* (IHS Technology, 2014)
3. D. Bollier, *How Smart Growth Can Stop Sprawl* (Books, Washington, 1998)
4. A. Bridgwater, Intel Mashery: how to manage an API. *Forbes* (2015)
5. V. Buscher, *Global Innovators: International Case Studies on Smart Cities, BIS Research Paper No. 135* (Ove Arup & Partners Limited, London, 2013)
6. R. Caldwell, Portland, a city of smart growth. The Masthead **54**(1), 29 (2002)
7. A. Caragliu, C. Del Bo, P. Nijkamp, Smart cities in Europe. J. Urban Technol. **2**, 65–82 (2011)
8. E. Castro-Leon, The consumerization in the IT ecosystem. IT Prof. **16**, 20–27 (2014)

9. J. Chambers, Are you ready for the internet of everything? Davos, Switzerland: World Economic Forum, January 15, 2014. https://agenda.weforum.org/2014/01/are-you-ready-for-theinternet-of-everything/. Retrieved 20 Jan 2015
10. H. Chourabi, C. Nam, S. Walker, J.R. Gil-Garcia, S. Mellouli, T.A. Pardo, H.J. Scholl, Understanding smart cities: an integrative framework. in *Proceedings of the 2012 Hawaii International Conference on System Science (HICSS-45)* (2012)

Big Data Analytics and Machine Learning Approach for Smart Agriculture System Using Edge Computing

U. Sakthi, K. Thangaraj, T. Poongothai, and M. K. Kirubakaran

Abstract With the development of the Internet of Things (IoT) and machine learning technology, a smart agriculture environment produces more agricultural land and crop-associated data for knowledge discovery systems. Machine learning decision-making algorithm is applied to discover hidden knowledge patterns from the agricultural data stored in the distributed database. Big data analytics extract useful information from the large, distributed, and complex datasets, which helps the farmer to increase crop yield and quality of the production. The edge computing node collects crop data and land environment data from the agricultural lands using a different kind of IoT sensors. The predicted smart agricultural knowledge pattern can provide needed information to the farmers and other users like an agent, agriculture officers, researchers, and producers to get more profit. Cloud and fog computing provides efficient distributed data storage for big data and execute dynamic operations to predict business intelligence facts to increase production and minimize natural resource utilization. We have compared traditional data mining techniques with the business analytical tool hybrid association rule-based decision tree (HDAT) MapReduce approach for implementing decision tree algorithm to predict and forecast the future needs of the farmer to increase the profit and reduce the resource wastage.

Keywords Precision agriculture · Machine learning · Edge computing

U. Sakthi (✉)
Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu 602105, India
e-mail: sakthiu.sse@saveetha.com

K. Thangaraj
Department of Information Technology, Sona College of Technology, Junction Main Road, Salem, Tamil Nadu 636005, India
e-mail: thangarajkesavan@sonatech.ac.in

T. Poongothai
Department of Computer Science and Engineering, St. Martin's Engineering College, Secunderabad, Telangana 500100, India

M. K. Kirubakaran
Department of Information Technology, St. Joseph's Institute of Technology, Chennai, Tamil Nadu 600119, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_63

1 Introduction

A computer-based smart agriculture system collects data from the farmer device, IoT sensors, and embedded devices and applies pattern prediction algorithm and deliver required information to the farmers and users. The size of agriculture data collected by Open Government Data (OGD) Platform India, and India Agriculture and Climate Dataset is in the order of gigabytes. The datasets are captured and gathered continuously, so the database size is increased by approximately 92 KB/min in the OGD data repository. The Global Positioning System (GPS) and IoT sensors agriculture dataset consists of crop images, weather images, soil and productivity data, which are stored in a cloud data repository for data analysis. To perform data analysis on large datasets, there is a need for high storage and a powerful computing system to provide efficient and effective data patterns to the users [1]. Big data analytics and cloud computing have been applied to store huge data and perform data analysis in the smart agricultural system [2, 3]. Edge computing is integrated with cloud computing and IoT to provide decisions with low latency time and increase the cloud network bandwidth [4].

Figure 1 shows the smart agricultural model based on IoT and cloud computing technology. Smart farming is the integration of IoT and machine learning process to reduce excessive use of natural resources and reduce the damage of crops. The proposed system provides user-friendly Web-based application to the farmers for acquiring knowledge about the farms and the + environment. Big data are applied everywhere to increase the profit in the business and to analyze the historical data with the following four parameters: volume, velocity, variety, and veracity.

The rest of the paper is organized as follows. Section 2 presents related work in a smart agricultural system using machine learning and edge computing. The architecture and functionalities of the different layers are discussed in Sect. 3. In Sect. 3, the workflow of the proposed system is summarized. Section 4 presents the machine learning algorithm used for data classification. Section 5 describes the

Fig. 1 Smart agricultural decision-making model

experimental setup, execution method, and analyzes the performance of the proposed system. Section 6 concludes the result of the precision agricultural system and future scope.

2 Related Works

Big data analytics and machine learning approach play a significant role in large dataset analysis and efficient decision-making process, which makes the smart agricultural system more profitable, safer, and efficient [5]. Many applications use big data analytics for achieving better profit and make better decisions with less time and cost [6]. Many research works have focused on the knowledge intelligent system for monitoring and controlling the agricultural land using IoT and fog computing with the cloud environment. Edge computing performs data analysis locally without transforming data to the cloud environment, which makes the precision agricultural system provide dynamic knowledge to the farmers and reduce the network latency. The computational load and storage process of the cloud environment are reduced by introducing the edge node in the IoT environment [7]. For example, the farmer can get the knowledge about the humidity of the soil, and he can decide the level of water irrigation to the crops so that the natural resource water can be utilized properly without wastage [8].

Intelligent services are provided to the farmers by applying data classification and prediction algorithms on the large datasets stored in the data repository in the edge node and the decentralized database server [9]. Precision agriculture system includes big data analytics, machine learning classification algorithm, edge computing, and IoT to increase the profit and reduce the wastage. The need for an edge computing nodes in cloud-based environments is analyzed in many research works [10]. The intelligent system does not move all dataset to the cloud server, which transfers partial data to the cloud server, which reduces the network latency and increases the network bandwidth [11]. An effective and efficient smart system is developed to overcome the challenges in the traditional agricultural land management systems. Many researchers and business applications use big data analytics to accomplish great success [12]. A smart agricultural system is a modern application technology to provide, compute, and analyze multisource agricultural data for different management as shown in Fig. 2. It supports water irrigation management, crop management, soil analysis, and fertilizer analysis. The Hadoop distributed file system (HDFS) is used to work with a large amount of data in distributed methods in the cloud network [13]. Google proposed MapReduce programming model for parallel and distributed data analysis approach for large datasets in a cloud environment. The important classification algorithm called decision tree is implemented using the MapReduce processing model to perform data analysis in a distributed manner.

Fig. 2 Functionalities of a smart agricultural system

3 Big Data Analytics in Smart Knowledge-Based Agricultural System

A smart agricultural system includes electronic IoT sensors, edge computing transmission technology, big data analytics, and machine learning. The purpose of this proposed system is to provide required agricultural land data to the farmers and other users without any delay in a precision agricultural environment as shown in Fig. 3. A smart agricultural system consists of four layers, which are the agricultural environment data collection layer, edge computing layer, big data analytics layer, and application layer. All layers are interconnected by the information or data flow. The functionalities provided by each layer are given below.

3.1 Agricultural Environment Data Collection Layer

This layer is responsible for collecting data from sources like the sensors deployed in the agricultural land, Global Positioning System (GPS), and Roadside Unit (RSU) and transfers those data to the next level edge computing layer and cloud server. The data will be in the form of photos, text, images, videos, and pictures.

3.2 Edge Computing Layer

The cloud network connected with the edge computing node via edge gateway using the protocols like Zigbee, Bluetooth, Wi-Fi, NFC, and Ethernet protocol. The edge computing node performs data analysis on the data collected from the data source and sends useful required information to the users.

Fig. 3 Layered architecture of the smart agricultural system using big data

3.3 Big Data Analytics Layer

The association rule-based decision tree algorithm is executed to generate knowledge patterns for the farmers. The large volume of data is collected from the data source and stored in the database server for data analysis. The high performance and distributed rule mining algorithm is executed and performs data management.

3.4 Application Layer

The farmer, agent, researcher, scientist, producer, and agent can access the smart agricultural application software using mobile devices connected with the Internet. The user can send a query to the knowledge base server and can receive information about the crop yield, fertilizer availability, pest control, soil details, and water irrigation status.

Fig. 4 System framework for big data analytics

4 Machine Learning Approach for the Precision Agriculture

In the big data ecosystem, machine learning technology provides powerful agricultural knowledge prediction algorithm to execute on large datasets. In Fig. 4, we have explained the steps involved in the big data analysis. In this paper, we propose the parallel and distributed association rule mining algorithm and tested with large datasets. The hybrid association rule-based decision tree algorithm is designed to analyze the distributed datasets available in the fog node and cloud computing node. In the supervised machine learning approach, decision trees are widely used prediction algorithm for large datasets. The most widely used supervised algorithm is tree-based algorithm, which gives greater accuracy, easy implementation, and consistency.

5 Performance Analysis

The performance of the proposed has been evaluated by recording the farmer query response time for 6 days. The query processing time is calculated as the sum of user query request time and query response time. Figure 5 shows the various mobile user query response time sent on different days to the cloud environment. QID number represents query identification number, and it is calculated as the time taken by the cloud application service to compute the query and send a response to the farmer user.

Fig. 5 User query response time

The response can vary based on the type and nature of the query sent by the end user. In Fig. 6, the performance of the various approaches for generating the knowledge from the large datasets is explained. The proposed hybrid association rule-based decision tree-based MapReduce approach gives more accuracy and less error than the other data classification approach. The query response time can vary based on the user request type and nature of the query.

Fig. 6 Performance comparisons of various approaches

6 Conclusion and Future Work

Integration of cloud computing technologies and machine learning approaches swiftly moves traditional agricultural systems to smart agricultural systems. Big data analytics system has been applied in a smart agricultural system to provide knowledge patterns to the farmers to make real-time decisions about farm management. The proposed precision agricultural system is used in the field of science, which leads to maximum crop yield by reducing the usage of resources. The farm management process is optimized with the use of big data and machine learning methods by increasing the crop yield and reducing the resources like fertilizer, water, and pesticide. This system provides the required information to the farmers at the right time and right place. In future, the system can be extended for finding particular crop disease analysis.

References

1. S. Wolfert, L. Ge, C. Verdouw, Big data in smart farming-a review. Agric. Syst. **153**, 69–80 (2017)
2. G.B. Kumar, An encyclopedic overview of '"big data"' analytics. Int. J. Appl. Eng. Res. **10**(3), 5681–5705 (2015)
3. M. Carolan, Publicising food: big data, precision agriculture, and co-experimental techniques of addition. Sociol Ruralis **57**(2), 135–154 (2017)
4. H.R. Zhang, Z. Li, T. Zou, Overview of agriculture big data research. Comput. Sci. **41**(S2), 387–392 (2014)
5. M. Chunqiao, P. Xiaoning, M. Yunlong, Research status and development trend of agriculture big data technology. J. Anhui Agric. Sci. **44**(34), 235–237 (2016)
6. D. Waga, K. Rabah, Environmental conditions' big data management and cloud computing analytics for sustainable agriculture. World J. Comput. Appl. Technol. **2**(3), 73–81 (2014)
7. Z.F. Sun, K.M. Du, F.X. Zheng, Perspectives of research and application of big data on smart agriculture. J. Agric. Sci. Technol. **15**(6), 63–71 (2013)
8. C.S. Nandyala, H.K. Kim, Green IoT agriculture and healthcare application (GAHA). Int. J. Smart Home **10**(4), 289–300 (2016)
9. A. Kamilaris, A. Kartakoullis, F.X. Prenafeta-Boldu, A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. **143**, 23–37 (2017)
10. A.Z. Abbasi, N. Islam, Z.A. Shaikh, A review of wireless sensors and networks' applications in agriculture. Comput. Stand Interface **36**(2), 263–270 (2014)
11. J. Lee, S.H. Kim, S.B. Lee, A study on the necessity and construction plan of the Internet of things platform for smart agriculture. J. Phys. IV **17**(11), 1313–1324 (2014)
12. H. Tian, W. Zheng, H. Li, Application status and developing trend of open field water-saving internet of things technology. Trans. Chin. Soc. Agric. Eng. **32**(21), 1–12 (2016)
13. I. Protopop, A. Shanoyan, Big data and smallholder farmers: big data applications in the agri-food supply chain in developing countries. Int. Food Agribus. Manage. Rev. **19**, 173–190 (2016)

Real-Time Object Detection System with Voice Feedback for the Blind People

Harshal Shah, Meet Amin, Krish Dadwani, Nishant Desai, and Aliasgar Chatiwala

Abstract Lots of people suffer from temporary or permanent disabilities, and one of them is blindness or the visually impaired. According to the World Health Organization (WHO), more than 1 billion people suffer from blindness. Technology may assist blind and visually impaired persons in a number of ways; objects detecting is still a difficult process. There are so many techniques available for object detection. But, the accuracy and efficiency of detection aren't good. Therefore, in this paper, authenticate object detection in real time using the YOLO-v3 and deep learning techniques. The main focus of this paper is to create a smartphone application that is cost-efficient with high accuracy.

Keywords YOLO v3 · Object detection · Machine learning · Text to speech · Deep learning · Image processing · Video processing · Blind people · Computer vision · COCO dataset

1 Introduction

According to the WHO in 2021, around, 1 billion people will suffer from high levels of visual loss. Out of those 1 billion people, 640 million had very low vision, and 39 million was totally blind. Visually impaired persons have complicated challenges while executing things that normal people accept as normal, such as reaching for something, recognizing, and identifying items in their environment. Avoiding impediments and recognizing the things in their environment are a problem for them. The goal of this project is to develop APK deep learning and computer vision technologies that are being used to assist blind persons in overcoming obstacles. The gadget should be able to categorize and identify various things in real time, as well as alert the user of the object class via audio feedback. All of the design and layout considerations are made with visually impaired persons in mind. People's comments and thoughts are considered because they are the ones who will utilize this APK.

H. Shah (✉) · M. Amin · K. Dadwani · N. Desai · A. Chatiwala
Parul University, Vadodara, Gujarat, India
e-mail: harshal.shah@paruluniversity.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_64

1.1 Deep Learning

Deep learning is a subset of machine learning in artificial intelligence (AI) that uses several layers of neural networks to describe human brain capabilities in data processing and pattern formation for important decision-making. Deep learning is an artificial intelligence (AI) function that comprises networks capable of unlabeled data from an unstructured dataset. Deep learning has grown significantly in the digital world, resulting in an avalanche of data in all kinds of formats and from every corner of the world. Big data can be gathered from a variety of online sources including e-commerce platforms, search engines, social media, and online streaming platforms and others. This huge volume of information is easily accessible and may be shared via cloud computing tools. However, the dataset utilized may well be unstructured and of such a large scale that understanding it, and extracting useful information and patterns might take decades for a human. Deep learning is an unsupervised learning technique that does not require data with specific variables and characteristics. Instead, it uses an iterative approach to get the desired outcome. When dealing with large amounts of data, deep learning typically performs admirably. Deep learning uses neural networks to automatically analyze a large dataset for patterns and tiny correlations. The learnt associations are utilized to interpret new knowledge once the model has been trained using the dataset.

1.2 Computer Vision

Computer vision is an associative subject that analyzes how computers can recognize digital pictures or videos (imagery) at a high level. Although computer vision techniques have been available since the 1960s, there has been a significant improvement in how well the software can explore this type of data due to recent advances in machine learning, as well as leaps forward in data storage technologies, much-improved computing capabilities, and cheap high-quality input devices. Any processing using visual material, such as photos, movies, icons, or pixels in general, is referred to as computer vision. However, there are a few important jobs in computer vision that serve as a framework. A model is trained on a dataset of distinct items in object categorization. The item is identified by the trained model as belonging to one or more of the trained classes. Because the model was trained on a dataset of an object, it should be possible to detect a specific instance of the object in object identification.

1.3 Object Detection

Object detection is a computer vision approach for detecting and detecting things in pictures and videos. Object detection may be used to numerical calculations in a scene, determine and monitor their specific location, and effectively label them using this type of identification and localization. Consider the following images: two books, one laptop, and one mobile phone. Object detection helps us to classify the sorts of objects we find while also locating them inside the picture. Bounding boxes are drawn around identified items using object detection. Object detection is strongly connected to other computer vision methods such as image recognition and image segmentation in that it assists in the comprehension and analysis of situations in pictures and video (Fig. 1).

2 Previous Work

There is some previous work already done in the past for helping blind people to help to visualize the surrounding objects with help of object detection techniques. There are some ideas shown below.

- Blind stick navigator
- Smart gloves for the blind
- Talking smart glass blind stick navigator.

Fig. 1 Object detection

The objective of the present invention is to propose a blind-stick for guiding the blind comprising a location sensor, a micro-controller unit, a temperature sensor, water-level sensor, a heart pulse sensor had a voice module; the present invention relates to the blind stick which includes an electronic sensing mechanism for helping visually impaired people while walking on the road [1]. They have made it smart by having an RF remote which helps to find a stick whenever it is lost.

2.1 Problem and Weaknesses of Current System

(1) Cost plays an effective role in everything; manufacturing devices include sensors, speakers, and many other tools which increase the cost of the system. Later on, such devices are not affordable for everyone.
(2) Lifetime support: None of these products gives a lifetime guarantee as some or other day the product needs service. Device can be lost or any issue can come anytime.
(3) Need to carry the stick or smart-box every time everywhere.
(4) Low battery can also create an issue at some point.

2.2 Requirements of New System

Blind people or visually impaired people face issues whenever they need to find what they want, or later they need someone's guidance to avoid obstacles or barriers; sometimes, it may happen that guidance is not available; in that case, it becomes difficult for the blind person or their care taker because all of the items and devices mentioned above are expensive and cannot be affordable. To overcome this problem, we are designing an android application that will be available for free for any user. A user should have a smart android phone, and this application will be able to help them in each and every way covering corner things also. This application will work as a guide for blind people, which includes real-time object detection nearby them and informing the user through an assistant and will help them to know the object and the distance between user and object.

3 Objective

The technology uses object detection, and the software will send a message to the user if it finds any possible impediments in the path.

Fig. 2 Object detection code

3.1 Object Detection

For object detection, the tool uses the YOLO-v3 algorithm. For the whole input images, it utilizes a unique neural network. After that the network splits the input images into multiple regions, predicting the bounding regions in the form of boxes with their probability scores (Fig. 2).

3.2 Dataset

Microsoft's Common Object in Context (COCO) dataset [2] is large-scale object detection, segmentation, and captioning dataset. COCO dataset is implemented in this research to train the YOLO model and text reader, which can detect 91 distinct classes. Features of COCO dataset are shown in below.

- 330 K images (>200 K labeled)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image.

Fig. 3 System work flow

4 Proposed System

The system is built in an android APK that recognizes a variety of objects in real time.

4.1 System Overview

A smartphone is used to capture real-time input data. The APK [3] camera is instantly accessible, and it begins recording the surrounding items, which the application informed them of. To help them learn and navigate their system, the system speaks out about every activity (Fig. 3).

4.2 Implementation

The system is designed to use a combination of technology stacks, which are explained further below. Android Studio is the primary integrated development environment (IDE) built particularly for android application development; it is utilized to develop the app. TensorFlow lite is a lightweight version of TensorFlow that was developed with the mobile operating system and embedded devices in the brain. It gives mobile users a machine learning solution with minimal latency and tiny binary size. TensorFlow includes a collection of fundamental operators that have been optimized for use on mobile devices. Custom operations in models are also supported. The video or image is processed in real time using You Only Look Once (YOLO) [4]. It predicts the observed item in the form of bounding boxes using a single convolutional network that is quicker than the regional convolutional neural network (R-CNN).

Fig. 4 Results

5 Experimental Results and Analysis

The APK [3] can identify and recognize a variety of obstacles that may be encountered when walking, as well as everyday items, and notify the user through assistance. A user must have a smart android phone, and this APK will be able to assist them in any manner possible, even inside and in non-crowded areas (Fig. 4).

6 Conclusion

The suggested solution would be extremely valuable and would improve the user's experience with their android smartphone by offering them assistance. With the user, auto-assistance will play a critical and significant role. Assistance can also be thought of as the user's third eye. Our major goal is to develop an APK [3] that allows

blind people or visually impaired people to experience their environment simply by listening to the assistance. As a result, it will aid in the prevention of such mishaps. The mobile devices may be conveniently carried, and the device's camera can be utilized to identify an object in the environment and provide an audio output.

References

1. R.R. Mahmood, M.D. Younus, E.A. Khalaf, N. Thirupathi Rao, Real Time Object Detection for Visually Impaired Person. Annals of R.S.C.B (2021)
2. Microsoft "COCO Dataset". http://cocodataset.org/
3. Android Developer. https://developer.android.com/
4. YOLO weights. https://pjreddie.com/darknet/yolo/

Bird Video Summarization Using Pre-trained CNN Model and Deep Learning

Rachit Adhvaryu, Viral Parekh, and Dipesh Kamdar

Abstract In the past few years, the forms of data have changed drastically from the text formats to images and today, most of the data are available in the video format. With this, there is a huge demand in the techniques that can provide the overall summary of the video. In this paper, we present the summary of birds that are identified from the large datasets using convolutional neural network (CNN). CNN is one of the best image processing and video processing model. The CNN model that we have used for the task is pre-trained AlexNet. The paper clearly proves that the work proposed recognizes the various kinds of birds from the inputted video with accuracy level ranging between 85 and 99%. At last, we provide the overall summary in terms of start time and end time of existence of each bird in the video.

Keywords AlexNet · Convolutional neural network (CNN) · Deep learning (DL) · Machine learning (ML)

1 Introduction

Today, the advancement in the forms of data has changed the complete scenario of research. Earlier, the data was in the form of Text. Later, the data transformed from Text to Images. Today, the data is avalialble in the form of Videos. This has forced the researchers to adopt new technology to generate optimal outputs. Moreover, with the huge datasets in the form of videos, there is an increasing demand in the techniques that can provide the overall summary of the video. Artificial intelligence has played a vital role in changing the thinking power of a humans[1]. The continuous revolution in artificial intelligence forced the humans to adopt its technologies to make their lives more smoother and comfortable.

R. Adhvaryu (✉) · V. Parekh
C. U. Shah University, Surendranagar, Gujarat, India
e-mail: rvadhvaryu@gmail.com

D. Kamdar
V. V. P. Engineering College, Rajkot, Gujarat, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_65

1.1 Artificial Intelligence

Artificial intelligence is also known as machine intelligence; that has capable of interpreting the data in the correct manner, learn that data and put it in execution by flexible adaptation [2]. In short, artificial intelligence commands the machine to try learning the data and generate the desired results. It can easily achievd by developing intelligent systems by learning problem solving techniques of human brains in various situations. The motive of artificial intelligence is to deliver intelligent systems that have an ability of reasoning, learning and problem solving.

1.2 Machine Learning

Machine learning is one of the powerful artificial intelligence application allowing automatic learning of important data from the large sources [3]. The machine learning algorithms create a mathematical model of datasets known as "training dataset", which can be further used in designing and developing intelligent systems. Machine learning also known as Predictive Analytics and it is classified in two types: (a) Supervised Learning, where training dataset contains inputs and its expected outputs determined from inputs and (b) Unsupervised Learning, where training dataset contains only inputs [3].

1.3 Deep Learning

Deep learning differs from machine learning and mainly deals with the algorithms structured from the functioning of neural networks [4]. The execution of deep learning algorithms can be described as problem understanding, dataset identification for detailed analysis, choosing appropriate deep learning algorithm, selecting proper training on large datasets, execution and at last testing it. The major drawback of deep learning is that they learn through interpretations [4] and the deep learning is mainly concerned with datasets.

1.4 Neural Networks

Neural networks identify patterns using some various sets of algorithms. They interpret sensory data through machine observations, clustering or grouping raw inputs [5]. The patterns identified by neural networks are numbers in the form of vectors and all the real-world dataset like images, sound, text or time series are required to be translated to these vector forms. All the operations take place between hidden input and output layers.

1.5 Convolutational Neural Network

A convolutional neural network (CNN) is a deep learning algorithm which expects an image as an input, assigns differetiable learnable weights and biases to different objects known as neurons [6]. For each neuron the output is generated by takeing several inputs, weighing sum over them and then pass it through some functions. CNN requires very low preprocessing as compared to other classification algorithms and methods [6]. CNN has played a vital role in video analytics and is highly recommended where the video processing is required and the ouptuts are expected very quickly as well as accurately. There are lots of convolutional neural network models available that can process the large sets of images and videos in very time span.

2 Related Work

Today, there is a continuos growth in the length and quantity of videos and thus it becomes very difficult for the users to watch each and every video to find the required information from it. Therfore, it is expected that a user must be able to see a summarized video using some kind of tools. Here, summarizing means creating a short story in terms of text, dataset or any format that can provide overall glimpse of the complete video without compromising the main content of the video [7]. Video summarization is one of the best technique to gather all the required informations form the video or its frames [7]. In the following section, we discuss various work on video summarization.

2.1 Audio Visual Video Summarization

This paper proposed to jointly exploit the audio and visual information from the video by developing an Audio Visual Recurrent Network (AVRN) [8]. Moreover, the proposed AVRN is separated into three parts: (1) the two-stream LSTM is used to encode the audio and visual feature sequentially by capturing their temporal dependency. (2) The audio visual fusion LSTM is utilized to fuse the two modalities by exploring the latent consistency between them. (3) The self-attention video encoder is adopted to capture the global dependency in the video [8]. Finally, the fused audio visual information, and the integrated temporal and global dependencies are jointly used in video summary prediction.

2.2 Video Summarization by Learning Relationships Between Action and Scene

This paper proposd a decent deep architecture approach for video summarization in untrimmed videos that parallely recognizes various classes of actions and scenes from the video segements [9]. This can be achievd through a multi-task fusion approach based on two types of attention modules to explore semantic correlations between action and scene in the videos. The networks contain the feature embedding networks and attention inference networks to describe the inferred action and scene feature representations [9].

2.3 Soccer Video Summarization Using Deep Learning

This paper presented a deep learning approach to summarize a long soccer videos by leveraging the spatiotemporal learning capability of three-dimensional convolutional neural network (3-D-CNN) and long short-term memory (LSTM)—Recurrent neural network (RNN) [10]. The proposed approach involved, (1) a step-by-step development of a Residual Network (ResNet)-based 3-D-CNN that detects a soccer actions, (2) Train 744 soccer clips from five soccer action manually, and (3) training an LSTM network on soccer features extracted by the proposed ResNet-based 3-D-CNN [10]. Thereafter, the 3-D-CNN and LSTM models are merged to identify soccer highlights. At last, to summarize a soccer match video, authors model the video input as a sequential concatenation of video segments based on their validated existence.

2.4 Event-Based Large Scale Surveillance Video Summarization

This paper proposed a new approach of large scale surveillance video summarization based on event detection. Initially, the vehicles and pedestrian paths are gathered using tracking-by-detection technique, and then the abnormal events are detected using the gathered paths. At last, a disjoint max-coverage algorithm is designed that generates a summarized sequence with maximum coverage of required events with minimum number of frames [11]. This approach contains a rich set of features as compared to traditional key frame-based approaches. Firstly, important information can be effectively extracted from the redundant data as the approach is event-centric and the required events contain almost all the vital information. Secondly, the abnormal events are efficiently identified by combining the Random Forest classifier and the features of object's paths and displayed [11].

3 Implementation

3.1 Methodology

The implementation methodology works in the following manner:

1. A large set of dataset available online is chosen as per the requirement.
2. Using deep learning algorithm, this dataset is trained for object detection.
3. At the same time, the video is processed and the frames are generated from the video.
4. In the next step, the trained dataset and the video frames are given to a pre-trained CNN model.
5. Using transfer learning, the CNN model learns about the trained dataset and using its own algorithm, the process of object detection is initiated.
6. The objects whose probability of detection is higer than the pre-defined threshold value are stored in the new dataset.
7. Lastly, the summary of video with time frames of each object is created automatically.

3.2 Work Flow

The complete implementation work flow can be easily described with the below mentioned steps:

Step 1: Take video as an input.
Step 2: Detect the frames with accuracy level above the threshold probability value.
Step 3: Record the first and the last frame for each detected object.
Step 4: Calculate the time in seconds for the first and the last frame and store start time, end time and object name in new data list.
Step 5: Treat this new data list as the final summary.

As discussed earlier, the pre-trained CNN model used in implementing video summarization is AlexNet. The network has deeper architecture, with more filters per layer, and with stacked convolutional layers. AlexNet is designed by the SuperVision group and its training on NvidiaGeforce GTX 580 GPU needs approximately 150 h [12].

AlexNet is deep neural network that is trained over millions of images of ImageNet dataset. AlexNet is 8 layers deep and can classify images into numerous objects like animals, birds, vehicles, pencils, pens, keyboards, etc. AlexNet contains a remarkable set of representations for large sets of images [13].

The working of the algorithm is as follows:

1. A video is chosen from the authenticated online repository.

2. The video is processed into frames and provided as an input to CNN frame by frame.
3. AlexNet labels each object in the frame with its probability and object name.
4. The predicated probabilities are used as a threshold value for object detection.
5. The details of each frame like start time, end time, object name and probability are stored in database.
6. At last, the start time and end time of the objects that are identified with more than 80% accuracy are displayed as a summary of video.

4 Results

We tested our approach on a video that contained more than 25 different birds with a dataset obtained form Kaggle. The length of the video was around 19 min. The main aim was to correctly identify the birds and generate video summary.

The level of accurate results of correct detection was measured on the probability threshold value as described in Table 1. The result with 0.7 and above threshold probability gives accuracy around 98.96% and the accuracy level decreases as the threshold probability value decreases. The decreasing accuracy level results into false detection which is discussed in the later part.

As discussed in Table 1, three birds namely, brambling, gold finch and water ouzel were selected randomly for its correct detection as described in Figs. 1, 2, 3 and 4 based on diffrent threshold probability values.

Table 1 Comparative analysis of threshold probability value versus correct detection

Probability	Total available frames	No. of frames	Correct detection	False detection	Accurate results in %
> = 0.7	27,400	9002	8900	102	98.87
0.6–0.7	27,400	11,188	10,526	662	94.08
0.5–0.6	27,400	14,782	12,701	2081	85.92
0.4–0.5	27,400	19,445	14,837	4608	76.3

Fig. 1 Bird detection with threshold probability value above 0.7

Fig. 2 Bird detection with threshold probability value between 0.6 and 0.7

Fig. 3 Bird detection with threshold probability value between 0.5 and 0.6

Fig. 4 Bird detection with threshold probability value between 0.4 and 0.5

The sample results of the exact time slots of the visibility of each bird in the video is described in Table 2.

It is possible in some cases that AlexNet may detect an incorrect bird with high threshold probability value as described in Fig. 5. This is due to some of the limitations of 11×11, 5×5, 3×3 matrix used by AlexNet for feature extraction and matching.

Table 2 Sample results

S. No.	Start time	End time	Bird identified
1	0.00	0.20	Brambling
2	0.46	0.52	Goldfinch
3	0.55	0.58	Brambling
4	0.57	1.08	Kite
5	1.12	1.18	Great Grey Owl
6	1.36	1.47	Goldfinch
7	1.47	2.02	Indigo Bunting
8	2.11	2.24	Water Ouzel
9	2.24	2.25	Bulbul
10	2.26	2.39	Water Ouzel
11	2.39	2.56	Kite

Fig. 5 False detection

AlexNet predicts the highlighted portion in Fig. 6. as the body structure and color simialar to bulbul and hence detects brambling as bulbul. However, such frames are very less in number and can be ignored in video summary.

Fig. 6 Reason for false detection

5 Conclusion

Creating a summary of the video saves a lot of time as the user gets exact infromation very quickly from the video. In this paper, the approach to summarize bird video with the help of pre-trained AlexNet CNN and DL, provides about 99% accurate result for identifying birds with detection probability of 0.7 and above. Moreover, 94% accuracy level of results with detection probability of 0.6 and above is achieved. Furthermore, the similar appraoch can be implemented on plant detection, vehicle detection and also summarizing special types of scenes in the movie.

References

1. C. Ujatha, U. Mudenagudi, A study on Keyframe extraction methods for video summary. in *2011 International Conference on Computational Intelligence and Communication Networks (CICN)*, pp.73–77, 7–9 Oct 2011
2. M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, A. Baskurt, Sequential deep learning for human action recognition. in *Human Behavior Understanding* (Springer, 2011), pp. 29–39
3. W. Sabbar, A. Chergui, A. Bekkhoucha, Video summarization using shot segmentation and local motion estimation. in *2012 Second International Conference on Innovative Computing Technology (INTECH)*, pp. 190–193, 18–20 Sept 2012
4. A. Frome, G. Corrado, J. Shlens, S. Bengio, J. Dean, M. Ranzato, T. Mikolov, Devise: a deep visual-semantic embedding model. in *NIPS* (2013), pp. 2121–2129
5. Z. Lu, K. Grauman, Story-driven summarization for egocentric video. in *CVPR* (2013), pp. 2714–2721
6. X. Wang, L. Zhang, L. Lin, Z. Liang, W. Zuo, Deep joint task learning for generic object extraction. in *NIPS* (2014), pp. 523–531
7. J. Wan, D. Wang, S.C.H. Hoi, P. Wu, J. Zhu, Y. Zhang, J. Li, Deep learning for content-based image retrieval: a comprehensive study. in *ACM Multimedia* (2014), pp. 157–166
8. B. Zhao, M. Gong, X. Li, Audio visual video summarization. Computer Vision Open Journal (2021)
9. J. Park, J. Lee, S. Jeon, K. Sohn, Video summarization by learning relationships between action and scene. IEEE Xplore Open Access Journals (2018)
10. R. Agyeman, R. Muhammad, G.S. Choi, *Soccer Video Summarization Using Deep Learning* (IEEE Computer Society, 2019), pp. 270–274
11. X. Song, L. Sun, J. Lei, G. Yuan, M. Song, Event-based large scale surveillance video summarization. Elsevier Journal (2015)
12. C. Lampert, Deep Learning with Tensorflow, [Online]. Available: http://cvml.ist.ac.at/courses/DLWT_W17/. Accessed 12 Feb 2019
13. Alexnet-Pretrained Convolutional Neural Network, [Online]. Available: https://in.mathworks.com/help/deeplearning/ref/alexnet.html. Accessed 13 Feb 2019

Generation of Seismic Fragility Curves for RC Highways Vulnerable to Earthquake-Induced Landslides Based on ICT

Aadityan Sridharan and Sundararaman Gopalan

Abstract In steep terrains, major highways are laid along steep slopes that are vulnerable to slope failures. More often than not, such mountainous regions tend to be seismically active. Structures located in steep slopes are vulnerable to earthquake-induced landslides and affect the integrity of the highways. We analyze two major landslide events that disrupted major highways in the past and generate fragility curves to estimate the hazard in such a scenario. We derive the curves based on landslides that damaged the highway by measuring the distance from the source of rockfalls. Our results indicate that the RC within 3–5 m from the crest of a landslide tend to be at a greater risk. In an event of higher magnitude, the fragility curves estimate lower probability of damage. This could be because the shaking itself causes enough damage than the triggered landslides. It was also observed for a distance of 10 m from a slope crest in a low-magnitude earthquake, the damage is the most. This shows that the momentum of the triggered rockfall increases due to increased distance and causes substantial damage on the RC highways. Damage caused by Taiwan rockfall event which occurred in April 2019 was assessed using these curves and was found to be moderate. Fragility curves generated in this study can be used to estimate the damage in a future scenario where there are highway constructions along slopes that are prone to rockfalls.

Keywords Building materials · Seismic fragility curves · Earthquake-induced landslides · RC bridges and PGA

A. Sridharan (✉)
Department of Physics, Amrita Vishwa Vidyapeetham, Amritapuri, India
e-mail: aadityans@am.amrita.edu

S. Gopalan
Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India
e-mail: sundar@am.amrita.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_66

1 Introduction

Highways that are constructed along major routes are an important means of transport within and among various countries. The highways that are constructed in mountainous regions are prone to seasonal road blocks due to landslides caused by heavy rainfall [1]. Certain such areas evolve into major landslide zones due to subsequent slope failures and deterioration in slope forming materials. When earthquakes occur, tens of thousands of landslides ensue and the additional effect of seismic shaking tends to further complicate the damage [2]. Reinforced concrete (RC) structures that come in the way of such major slope failures are severely damaged by the mass wastage and the incoming seismic waves. Major roads that are constructed using RC are cut off for weeks and sometimes months together [3]. The integrity of the RC highways is very important for continued access in the mountainous terrain.

Research on estimation of earthquake-induced landslide (EIL) hazard is extensive and detailed. Statistical and physical models which evaluate the spatial distribution, landslide displacement and volume, give a complete picture of hazard posed by the EILs. However, the effect of EIL on RC structures, bridges and roads have not been explored in the literature except for a few recent studies [4, 5]. The effects of seismic shaking on RC buildings is a known phenomenon. From the models that describe the damage of earthquakes on RC buildings, a similar model can be enumerated for the damage that EILs can cause to RC structures [6].

Fragility curves define the hazard resistance of a RC structure in the event of seismic shaking [7]. They denote the probability of structural damage beyond a certain state. These curves can be empirical or analytical depending on the model suitable for the problem in hand. Empirical curves are generated from past data that have recorded instances of damage to structures reported by an earthquake event. Analytical curves are based on the following factors/parameters:

1. Response spectrum of an RC structure observed through generated ground motion
2. Simulation of RC structure about to collapse
3. Through simulation of RC structures that account for uncertainty in properties of building materials [8].

Empirical curves can be used in predicting the real-time hazard for events of similar intensity in the future. Analytical curves can estimate the hazard for RC structures in steep terrain based on simulated values and response spectrum of the acceleration records [8]. Although literature point towards hazard analysis based on fragility curves for RC buildings, there have been very few attempts in obtaining a direct model that connects fragility curves to RC Highways and roads due to earthquake induced landslides [9].

In this work, we address the hazard eminent on RC highways due to EIL's by generating fragility curves that estimate the probability of damage. Various levels of hazards have been named as limit states, which will define possible damage to RC structure. The variation in strain on the RC frame for each limit state is mapped with

the earthquake intensity to generate probability of hazard due to EILs. It has to be noted that the primary damage due to seismic shaking and the combined effect of shaking with landslides is ignored for this study.

2 Methodology

Various categories of slope failures are observed when landslides are caused by earthquakes. The most damaging of them is rockfall (RF). The damage levels on the highways for rockfall are explained in Table 1 (adapted from Hu et al. [6]). These are different categories of triggered landslides based on the seismic ground shaking. Each case will have a corresponding level of damage which an be defined as the limit state (LS) mentioned in Table 2.

The limit state (LS) described in Table 1 will correspond to various intensity levels of an EIL event. Each of these intensity levels will cause a variation in strain caused on RC frames in the structures. Table 2 gives a comprehensive overview of the possible strain change in RC frames for each LS. Distances of 3, 5, and 10 m from the landslide crest are considered for damage assessment of highways and roads based on literature[10].

Table 1 Description of each limit state (LS). Each limit state is increasing intensity of EILs

Limited state	Failure stage and observed RF due to the earthquake
1	A small amount of rockfall may appear near the highway road cuts
2	There are some small collapses and rockfalls along the valley and highway road cuts
3	Medium and small collapses and rockfalls are common but are rarely large
4	Collapses and rockfalls are widespread, as well as the existence of rock slides

Table 2 Structural damage levels corresponding to each limit state for RC highways

Limited state	Structural damage	Description
1	None to slight	The highway RC material shows a linear response which is elastic. This can generate hairline cracks that are flexural or shear type by nature. Crack width is < 1.00 mm
2	Moderate	There are wider cracks that are almost 1 mm in width and causes spalling in the RC
3	Extensive	Widespread cracks of more than 1 mm is observed; the cracks are flexural/shear by nature. The highway might require longitudinal reinforcement and proper buckling
4	Complete	The RC Highway is now beyond repair, and it is not practically possible. Horizontal seismic displacements shatter the vertical columns due to shear failure. Demolition will be required

Fig. 1 The Zhongbu expressway, **a** before the RF event, **b** the RF video still from dashcam video (from Sridharan et al. [11])

2.1 Limit State and Probability of Damage

Tension crack is the point of initiation of the landslide usually situated near the crest of a slope. Average distance from a tension crack or a crest to the road in mountainous terrain is 5–60 m depending on the construction of the RC Highway as shown in Fig. 1. Figure 1 is an example of a highway from Taiwan where there was a rockfall on April 25, 2019, due to an earthquake of magnitude M_W 6.1 off-coast Taiwan, (a) shows the highway unaffected before the event and (b) is snapshot of the dashcam video showing the boulder falling off [11]. To analyze the impact of RF, we consider 3 m, 5 m, and 10 m distance from highway to crest as three cases to arrive at fragility curves [10]. The probability function that defines the damage possibility at each limit state is a lognormal distribution given by Eq. 1 [10].

$$P(\mathrm{LS} > \mathrm{LS}_j|\mathrm{IM}) = \phi\left(\frac{L_n(\mathrm{IM}) - L_n(\mathrm{IM}')}{\beta}\right) \tag{1}$$

LS_i is the individual limit state where ($i = 1, 2, 3, 4$), IM is the earthquake intensity measured in terms of peak ground acceleration (PGA) in units of g, IM′ is the median value of the same in terms of PGA measured in g, ϕ is the cumulative density function (CDF) which is usually known as standard normal CDF and β is the log-standard deviation. The value of PGA for around 1200 acceleration records were recalculated for slope crests from recording stations. The details in selection of records are explained in the next section.

2.2 Strong Motion Records and Rockfall Inventory

Various records of strong ground motion were analyzed from two major earthquakes: 1. M_w 7.9 Wenchuan, China (2008) and 2. M_w 7.8 Gorkha, Nepal (2015). Landslide inventories of both these events were used to find rockfall incidents along major highways in the respective regions [12]. Over 2000 such incidents were used to

Table 3 Median PGA values for each LS with the RC Highway at a distance of 3, 5 and 10 m from the slope crest

Distance from crest (m)	LS1 (PGA'(g)) $\beta = 0.91$	LS2 (PGA'(g)) $\beta = 0.91$	LS3 (PGA'(g)) $\beta = 0.91$	LS4 (PGA'(g)) $\beta = 0.92$
3	0.26	0.51	0.59	0.7
5	0.17	0.33	0.51	0.66
10	0.0725	0.28	0.46	0.59

arrive at IM′ which is chosen to be the median PGA value for the three different distances in each limit states. Table 3 summarizes the median PGA value for each LS, the β values were adopted from Fotopoulou et al. [9]. The records from the events were generated using the Seismoartif software-based input parameters such as hypocenter distance and PGA at outcrop.

Each PGA values that contributed towards damage to the RC Highway were considered in the study with a segregation based on distance from highway. All the failures falling under the three categories were mapped with the PGA map of the event from the United States Geological Survey (USGS) to compare the surface PGA with the PGA from rock outcrop. Since landslides are surficial phenomenon, we considered the surface PGA as opposed to the traditional value of PGA at outcrop.

The accelerograms are generated based on local seismic codes and available sample ground acceleration records. These accelrograms are part of a large set of accelerograms produced in this study based on PGA map to analyze the variation of the accelerations with respect to distance from crest. The variation in acceleration depends on the hypocentral distance and the seismic codes in Nepal and China.

3 Results and Discussion

The fragility curves were generated for each LS and different distance values of 3, 5, and 10 m as shown in Fig. 2. The variation in the probability of each LS with respect to the PGA that can damage the RC Highway is shown in the plot. It can be observed that with higher limit states, the probability of damage due to landslides are reducing. For all the four categories of LS, it is be observed that the 10 m curve gives higher probability of damage; this is consistent with general observation that RF initiated at a farther distance can cause immense damage on the RC Highway. Similarly, with increasing LS, the reason for reduction in probability of damage is because the seismic shaking will primarily contribute to maximum damage and the RF will reach a pre-damaged RC Highway, therefore not contributing to major damage of the structure [2, 5]. On close observation of LS1, the 10 m curve starts at a probability value higher than the 5 m and the 3 m curves. This could be because a slight tremor will retain the integrity of the RC Highway but can dislodge a weak

Fig. 2 Fragility curves for landslides caused by earthquakes (Wenchuan and Gorkha) for LS1, LS2, LS3, and LS4

boulder from the crest and as Sridharan et al. [11] have measured; it can cause great damage to the highway.

The LS1 and LS2 curves are showing high-probability values for low values of PGA. At a PGA value of 0.2 g in LS1 the probabilities vary from 0.4 to 0.8 over the three distances and reaches the bare minimum of 0.09 to 0.12. This signifies the effect of PGA in lower LS is profound since any slight acceleration will cause boulders/rocks to hit the highway. On close examination of LS2, we find that the 5 and 10 m curve are more similar than the 3 m curve. The 3 m curve still relates to lower probability which could be because at a short distance, the RF will not have enough momentum to cause serious damage. But in LS3 and LS4, the probability values for 3, 5, and 10 m are very similar. Albeit showing significant variation in probability values, there is not a large variation in trend among the three different distances in LS3 and LS4. As mentioned before, this could be due to damage caused by shaking being more prominent than the RF itself.

The variation in PGA from 0.05 to 0.8 g is the maximum observed range from two of the most destructive events in history [13, 14]. It has to be noted that these are not the only values of PGA recorded in these two events, but these are the values of PGA for which there was a damage caused by RF on the RC Highways. The

literature points towards assessments of fragility curves from hypothetical values of PGA ranging from 0 to 2 g [2, 5]. Although such studies can give robust estimates for a wide range of acceleration that can be observed in an engineering structure, for RF originating from shallow earthquakes the fragility curves developed in this study might be more reliable. One of the important aspects of uncertainty that could occur in estimating the fragility on-site is topographic site amplification [15]. As it is outside the scope of this work, we do not show the variations in the probability with increase in amplified PGA values [15]. In such cases, it is good to analyze the stratification of the geology at site before assessing the vulnerability of the RC Highway using the fragility curves [16].

On validating these curves with the Taiwan earthquake mentioned earlier, we found that the LS is LS2 and the estimated distance from the highway to the crest is 10 m [11]. The PGA from generated records was found to be 0.2 g; hence, from the first figure in Fig. 2, there is a probability of 0.87, and the highway will be damaged moderately in reference with Table 2. The field image from the Dashcam video shows the rockfall knocks off the railing of the highway. This was one of the many RFs on the same route; there are news reports that the highway closed for more than twelve hours due to this event. It is worth noting that the results predicted by Sridharan et al. [11] based on the bounce height and the Newmark's displacement confirm that the RF was initiated by the slight tremor due to the Mw 6.1 event. This result is reflected in the fragility curves generated in this study; the low PGA value and a medium magnitude earthquake that corresponds to the lowest LS2 has a better probability of causing damage than a LS4 event with the same PGA.

On a general note, when using these curves for assessing the hazard for a future scenario, it is important to estimate the LS that the future event could possibly belong to. An estimated PGA value can be used to generate possible damage estimates along the highways. In assessing damage on highways that are along fragile slope, geotechnical slope stability models have to be used to identify regions vulnerable to RF along the highways [17]. A quick glance on the fragility curves can give us an estimate of damage possible to a RC Highway. This shows the importance of this work for quickly estimating the probability of damage when there is a possibility of RF. Since two of the largest inventories are included in the calculation, these curves will have a wide range of applications for RF events during an earthquake. We stress that the material in the highway is RC because most of the highways that are paved along steep slopes are observed to constitute the same.

4 Conclusion

Fragility curves that correspond to the various possible scenarios of earthquakes causing RFs are generated. The trend shows that for various LS described in this study, the variation in probability of damage depends on the distance. Farther the distance, more the probability of damage. Further, for small earthquake events that cause shaking just enough to dislodge loose boulders, the damage is observed to be

more even though the intensity of shaking might be less. The fragility curves were validated for a RF event in Taiwan that was of LS2 and a PGA of 0.2 g. The predicted probability was as high as 0.87 for a RC Highway at a distance of 10 m from the point of initiation of RF. This was cross verified by the dashcam video shown in Fig. 1, and the damage seems to be pertinent. These fragility curves can be used on a RC Highway to analyze the probability of damage due to such rockfall events.

Acknowledgements The authors would like to thank the Chancellor of Amrita university, Sri. Mata Amritanandamayi Devi for being a constant source of inspiration behind the work.

References

1. L. Jacobs, O. Dewitte, J. Poesen, D. Delvaux, W. Thiery, M. Kervyn, The Rwenzori Mountains, a landslide-prone region? Landslides **13**, 519–536 (2016)
2. S.D. Fotopoulou, K.D. Pitilakis, Vulnerability assessment of reinforced concrete buildings subjected to seismically triggered slow-moving earth slides. Landslides **10**, 563–582 (2013)
3. S. Dhakal, P. Cui, C.P. Rijal, L. Su, Q. Zou, O. Mavrouli, C. Wu, Landslide characteristics and its impact on tourism for two roadside towns along the Kathmandu Kyirong Highway. J. Mt. Sci. **17**, 1840–1859 (2020)
4. A.M. El-Maissi, S.A. Argyroudis, F.M. Nazri, Seismic vulnerability assessment methodologies for roadway assets and networks: a state-of-the-art review. Sustainability. **13**, 61 (2020). https://doi.org/10.3390/su13010061
5. Y. Jafarian, A. Lashgari, M. Miraiei, Multivariate fragility functions for seismic landslide hazard assessment. J. Earthq. Eng. **25**, 579–596 (2021)
6. H. Hu, Y. Huang, Z. Chen, Seismic fragility functions for slope stability analysis with multiple vulnerability states. Environ. Earth Sci. **78**, 690 (2019)
7. J. Kiani, C. Camp, S. Pezeshk, On the application of machine learning techniques to derive seismic fragility curves. Comput. Struct. **218**, 108–122 (2019)
8. J.E. Padgett, R. DesRoches, Methodology for the development of analytical fragility curves for retrofitted bridges. Earthq. Eng. Struct. Dyn. **37**, 1157–1174 (2008)
9. S.D. Fotopoulou, K.D. Pitilakis, Probabilistic assessment of the vulnerability of reinforced concrete buildings subjected to earthquake induced landslides. Bull. Earthq. Eng. **15** (2017)
10. S.D. Fotopoulou, K.D. Pitilakis, Vulnerability assessment of reinforced concrete buildings at precarious slopes subjected to combined ground shaking and earthquake induced landslide. Soil Dyn. Earthq. Eng. **93**, 84–98 (2017). https://doi.org/10.1016/j.soildyn.2016.12.007
11. A. Sridharan, S. Gopalan, Predictive analysis of co-seismic rock fall hazard in Hualien County Taiwan. In: *Machine Learning and Information Processing, Advances in Intelligent Systems and Computing* (Springer Singapore, 2020), pp. 348–358
12. H. Tanyaş, C.J. van Westen, K.E. Allstadt, M. Anna Nowicki Jessee, T. Görüm, R.W. Jibson, J.W. Godt, H.P. Sato, R.G. Schmitt, O. Marc, N. Hovius, Presentation and analysis of a worldwide database of earthquake-induced landslide inventories. J. Geophys. Res. Earth Surf. **122**, 1991–2015 (2017). https://doi.org/10.1002/2017JF004236
13. K. Goda, T. Kiyota, R.M. Pokhrel, G. Chiaro, T. Katagiri, K. Sharma, S. Wilkinson, The 2015 Gorkha Nepal earthquake: insights from earthquake damage survey. Front. Built Environ. **1**, 1–15 (2015). https://doi.org/10.3389/fbuil.2015.00008
14. T. Gorum, X. Fan, C.J. van Westen, R.Q. Huang, Q. Xu, C. Tang, G. Wang, Distribution pattern of earthquake-induced landslides triggered by the 12 May 2008 Wenchuan earthquake. Geomorphology **133**, 152–167 (2011)

15. R. Ramkrishnan, K. Sreevalsa, T.G. Sitharam, Development of new ground motion prediction equation for the North and Central Himalayas using recorded strong motion data. J. Earthq. Eng. 1–24 (2019). https://doi.org/10.1080/13632469.2019.1605318
16. N. Vasudevan, K. Ramanathan, Geological factors contributing to landslides: case studies of a few landslides in different regions of India. IOP Conf. Ser. Earth Environ. Sci. (2016)
17. R. Ramkrishnan et al., Stabilization of seepage induced soil mass movements using sand drains. Geotech. Eng. **48**, 129–137 (2017)

Influence of Student–Teacher Interrelationship on Academic Achievements: A Logit Model

Monika Saini, Asha Choudhary, and Ashish Kumar

Abstract The main concentration in the present work is given to identify the effect of the student–teacher relationship on the academic achievement of university students. For this purpose, data is collected from 292 students of various universities situated in Jaipur, India. The survey was conducted from January 2018 to April 2018. A questionnaire of 36 questions was filled by all respondents enthusiastically, out of which six questions are demographic. Responses have been collected on a five-point Likert scale. In the sample, 55.5% male and 44.5% students are included with 28.1% day-boarding and 71.9% hosteller students. Maximum 52.7% of students belong to the third year, and a minimum of 2.1% students are included from fourth year from various branches like engineering, science, arts and management. The reliability of the questionnaire was checked by Cronbach's Alpha, and logistic regression analysis is performed. Two logistic regression models are developed concerning gender and locality of students, and odd ratios are identified. Finally, it is recommended that findings can be used to enhance the academic achievements of the students.

Keywords Student–teacher relationship · Academic achievements · Logistic regression · Odd ratios · Gender · Locality

1 Introduction

From the ancient era of human civilization, a teacher's role fluctuates from time to time among nations. In the beginning, teachers involved in formal teaching tasks including formulation of curriculum, teaching, preparing lectures, giving lessons and assessing the performance of students. It is always expected from a teacher that he unwraps the thinkers, touch the hearts and hold the hands of students. To achieve these objectives, the teacher must play many roles like historian, sleuth, visionary, symbol, potter, poet, actor and healer. Teaching is more about the humanity inside the teacher rather than the outside the technology. It is considered as a matter of heart instead of

M. Saini · A. Choudhary · A. Kumar (✉)
Department of Mathematics and Statistics, Manipal University Jaipur, Jaipur 303007, India
e-mail: ashish.kumar@jaipur.manipal.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_67

the head. It is about the characteristics of teachers like identity, integrity, intensity and individuality. It is about commitment and connections with students. Now, teachers become facilitators, role model, assessors, planners, resource-developer and information provider more than traditional teacher. In the current competitive and globalize scenario, students and managements also expect much more from the teachers. In this situation, student–teacher interrelationship plays a key role in creating a learning environment for students. Many researchers suggested some techniques to improve student–teacher relationship like remember the name of students as soon as possible, identify some personal things of each student, discuss on some current issues, provide a positive, caring, and enthusiastic environment, avoid words like threats and punishment, remains unbiased and provide supportive atmosphere. Always try to motivate and encourage the students even though they give the wrong answer to the questions, never demoralize and humiliate them. This positive student–teacher relationship can have a long-lasting impact on the academic growth of the students. Teachers with strong relationships with students are more effective in teaching, faceless behavioural difficulties and achieve higher-level academic success. The following are the key benefits of positive student–teacher relationships:

- Improve academic success
- Prevent behaviour problems in the classroom
- Improve student attitude towards classwork
- Adding growth in and outside the classroom.

Murray [1] discovered a relationship between low-inferences classroom teaching behaviours and student ratings of college teaching effectiveness. Brophy and Good [2] studied teacher behaviour and student achievement relationship. Wang et al. [3] proposed some factors which help students to learn. Astin [4] find the factors that matter in the college life of students. Rawnsley [5] discovered the relationship between homeroom learning conditions, instructor relational conduct and understudy results in auxiliary arithmetic study halls. Evans [6] completed an investigation of understudies' social foundation and educator understudy relational conduct in optional science homerooms in Australia. Brekelmans et al. [7] set up a connection between educator experience and the instructor understudy relationship in the study hall climate. Klem and Connell [8] set up connections and connecting between instructor backing and understudy's commitment and accomplishment. Brok et al. [9] built up a numerical model between relational instructor conduct and understudy results. Bork et al. [10] considered the impact of understudy instructor relational connection on conduct and subject-explicit inspiration. This investigation welcomes insights from research on instructing, learning climate, learning in explicit subjects by associating it to educator's relational conduct with understudy's subject related demeanour. Hanushek and Rivkin [11] clarified a few speculations about utilizing esteem added proportions of instructor quality. Lewis [12] introduced pathways towards improving instructing and learning in advanced education in a global setting and foundation. Arum and Roksa [13] examined the scholastically hapless in school grounds about restricted learning. Loes et al. [14] examined the impact of educator's conduct on understudy's tendency towards long-lasting learning. It is distinguished

that controlling for a battery of possible perplexing impacts, educator association was emphatically connected with gains in understudies need for discernment. Mehdipour and Balaramulu [15] researched the impact of educator's conduct on the scholarly development and accomplishments of understudies. The investigation was performed on an example of 1080 respondents taken from 13 colleges of Hyderabad, and a positive relationship has been discovered between instructor's conduct and the scholarly accomplishment of understudies. Hamre et al. [16] built up a structure for instructor adequacy through study hall communications. Hamre et al. [17] introduced some proof for general and space explicit components of instructor youngster communications that are related with preschool kids' turn of events. Chetty et al. [18] built up a model for estimating the effects of educators for assessing predisposition in instructor esteem added gauges. Hafen et al. [19] created factor structure and useful use of study hall appraisal in the scoring framework through communications. Gershenson [20] planned a model for connecting educator quality, understudy participation and understudy accomplishment. Blazar and Kraft [21] completed an investigation on the part of instructor and showing impacts on understudy's demeanour and conduct. A lot of work has been carried out in the direction of student–teacher relationship, but there is no work visualized on which influence of student–teacher relationship on academic achievements among university students. The manuscript comprises four sections. The first section is the introduction in which a detailed literature review is also given. In the second section, material and methods are discussed that are used for analysis. The third section included the results of data analysis, and the last section is the conclusion.

2 Materials and Methods

2.1 Study Design and Participants

It is a cross-sectional study accompanied by the students at various universities situated in Jaipur, the western region of India, with students studying in five faculties including management, sciences, arts, law, journalism and engineering. Respondents comprised undergraduate students at this university. All respondent's voluntarily participated in the survey. The stratified random sampling is used to select the respondents. Respondents were asked to indicate their response on a five-point Likert scale anchored to 1—"strongly agree" and 5—"strongly disagree".

2.2 Data Collection and Questionnaire Development

A questionnaire prepared by Mehdipour and Balaramulu [15] has been used to collect data from respondents. It is used due to its relevance to the objectives of the study. It

also ensures that respondents understood the written language of the questionnaire. The questionnaire comprised two sections: one contains six questions related to demographic information and another one has 30 questions related to the student–teacher relationship. The 292 respondents filled the questionnaire and returned.

2.3 Statistical Analysis

SPSS Software version 21 has been used to statistically analyse the data. The categorical data were summarized as frequency and percentage from Table 1. Bartlett [22] developed a methodology for testing the sampling adequacy of questionnaire to perform factor analysis. The reliability of the questionnaire was measured by Cronbach's alpha. It is suggested that if Cronbach's alpha values are greater than 0.6, then the questionnaire is reliable. From Table 2, it is identified that Cronbach's Alpha value is 0.871 that is sufficiently large than 0.6, so data and questionnaire are reliable for statistical analysis.

Table 1 Frequency distribution of respondents

D.V	Type	Frequency	Per cent	Valid Per cent	Cumulative Per cent
Gender	Male	162	55.5	55.5	55.5
	Female	130	44.5	44.5	100
	Total	292	100	100	
Locality	Day-boarding	82	28.1	28.1	28.1
	Hosteller	210	71.9	71.9	100
	Total	292	100	100	
Class	First	46	15.8	15.8	15.8
	Second	86	29.5	29.5	45.2
	Third	154	52.7	52.7	97.9
	Fourth	6	2.1	2.1	100
	Total	292	100	100	
Stream	Engineering	114	39	39	39
	Management	80	27.4	27.4	66.4
	Science	6	2.1	2.1	68.5
	Arts	92	31.5	31.5	100
	Total	292	100	100	

Table 2 Reliability analysis of the questionnaire

Cronbach's Alpha	No. of items
0.871	30

3 Logistic Regression

A technique frequently used to build up a relapse type model for foreseeing a dichotomous ward variable. This technique uses an "s-bend" type capacity to build up a number-related model from at least one factors, which can yield the likelihood of having a place with one gathering or the other. It falls somewhere between regression and discriminant analysis. If an event has two outcomes, then it is termed as binary logit model. A maximum likelihood estimator is used to estimate the parameters in the logit model. The model fit is measured based on likelihood function, Cox and Snell R square and Nagelkerke R square. Wals's statistics are used to test the individual estimated parameters [23–25]. Mathematically,

$$\log_e\left(\frac{p}{1-p}\right) = a_0 + a_1X_1 + a_2X_2 + \ldots\ldots + +a_kX_k$$

Or

$$p = \frac{\exp\left(\sum_{i=0}^{k} a_iX_i\right)}{1+\exp\left(\sum_{i=0}^{k} a_iX_i\right)}$$

where p = probability of success; X_i = Predictor variable i; a_i = unknown parameter to be estimated.

4 Data Analysis

Using statistical software SPSS 21, binary logistic regression analysis has been performed. The Omnibus test Table 3 gives the Chi-square value 142.031 and p-value = 0.000 < 0.05. By observing these values, it is analysed that the null hypothesis, i.e., intercept and all other estimated parameters are zero cannot be accepted.

Cox and Snell's R square and Nagelkerke's R square tell us about the proportion of variance explained by independent variables in the categorical dependent variable. Cox and Snell's R square is based on the log likelihood of the baseline model, and Nagelkerke's R square is the adjusted version of Cox and Snell's R square. From Table 4, it is observed that the model is good but not excellent. Here, it is interpreted

Table 3 Omnibus tests of model coefficients in respect of gender

		Chi-square	Df	Sig
Step 1	Step	142.031	30	0.000
	Block	142.031	30	0.000
	Model	142.031	30	0.000

Table 4 Model summary in respect of gender

Step	-2 Log likelihood	Cox and Snell R Square	Nagelkerke R Square
1	259.253[a]	0.385	0.516

[a]Estimation terminated at iteration number 5 because parameter estimates changed by less than 0.001

Table 5 Classification table in respect of gender

	Observed		Predicted		
			Gender		Percentage correct
			Male	Female	
Step 1	Gender	Male	132	30	81.5
		Female	32	98	75.4
	Overall Percentage				78.8

that 51.6% of students think that there are significant chances that the student–teacher relationship affects the academic achievement of students.

From Table 5, it is analysed that 78.8% correct classification of respondents according to gender-wise have been made in the logistic regression model.

By running logistic regression in SPSS, the number of variables that remains in the logit model is identified. The values for the parameters of logistic regression equation for predicting the dependent variable from the independent variable and odds ratios are estimated, and Wald values are used to test the significance of estimated parameters. By using significant values and a 0.05 level of significance, it is analysed that only seven variables and constant are statistically significant having a p-value less than the level of significance. The Omnibus test Table 6 gives the Chi-square value 96.320 and p-value $= 0.000 < 0.05$. By observing these values, it is analysed that the null hypothesis that constant and all other estimated parameters are zero cannot be accepted. So, the null hypothesis is rejected.

From Table 7, Here, it is interpreted that 40.4% of students in the locality-wise classification think that there are significant chances that student–teacher relationship affects the academic achievement of students.

From Table 8, it is analysed that 83.6% correct classification of respondents has been made in the logistic regression model according to the locality of residence.

Table 6 Omnibus tests of model coefficients in respect of locality

		Chi-square	Df	Sig
Step 1	Step	96.320	30	0.000
	Block	96.320	30	0.000
	Model	96.320	30	0.000

Table 7 Model Summary in respect of locality

Step	-2 Log likelihood	Cox and Snell R Square	Nagelkerke R Square
1	250.417[a]	0.281	0.404

[a]The cut value is 0.500

Table 8 Classification table in respect of locality

	Observed		Predicted		
			Locality		Percentage Correct
			Day-boarding	Hosteller	
Step 1	Locality	Day-boarding	48	34	58.5
		Hosteller	14	196	93.3
	Overall percentage				83.6

The number of variables that remains in the logit model is identified with respect to locality also. By using significant values and a 0.05 level of significance, it is analysed that only seven variables are statistically significant having a p-value less than the level of significance.

5 Conclusion

The assessment of the student–teacher relationship on academic achievement among university students has been made with the help of logistic regression. Two logistic regression models have been developed one for gender and the other for the locality of the students. The models are as follows:

$\log_e\left(\frac{p}{1-p}\right) = 3.398 + (1.540)$* your faculty have a command on their subjects –(0.460)* Your faculties provide you with relevant information/explanations to explain the points of subjects matter –(0.524)* your faculties behave more nicely to some students –(0.656)* Your faculties lay emphases on completion of course work in time + (0.460)* Your faculties like some students more and favour them unduly + (0.489)* Your faculties become nervous when clarifications are sort-(0.488)* Your faculties use different teaching techniques.

$\log_e\left(\frac{p}{1-p}\right) = 1.657 + (0.883)$* Your faculty have a command on their subjects –(0.589)* Your faculties are friendly and approachable –(0.737)* Your faculties come to class well prepared for the lecture teaching –(0.57) Your faculties are efficient in their duties + (0.580)* Your faculties become nervous when clarifications are sort-(0.757)* Your faculties use their voice effectively during teaching + (0.883)* Your faculties relate the topic with the real-life situation through different examples.

So, this study identified seven important factors that influence the student–teacher relationship and their impact on academic achievements. From the regression analysis, it is revealed that faculty's subject command and faculty's ability to relate the topic with the real-life situation through different examples have a significant positive impact on academic growth. The developed logit models are good as Nagelkerke's R square value is 0.516. Thus, the findings of the above study help improve the student–teacher relationship that results in enhancing the academic growth of students.

References

1. H.G. Murray, Low-inference classroom teaching behaviours and student ratings of college teaching effectiveness. J. Educ. Psychol. **75**, 138–149 (1983)
2. J., Brophy, T. L. Good, Teacher behaviour and student achievement. in *Handbook of Research on Teaching*, ed. by M.C. Wittrock, 3rd edn. (MacMillan, New York, NY, 1986)
3. M.C. Wang, G.D. Haertel, H.J. Walberg, What helps students learn? Educ. Leadersh. (Alexandria, VA) **51**(4), 74–79 (1993)
4. A.W. Astin, *What matters in college? Four critical years revisited* (Jossey-Bass, San Francisco, CA, 1993)
5. D.G. Rawnsley, Associations between classroom learning environments, teacher interpersonal behaviour and student outcomes in secondary Mathematics classrooms. Unpublished Doctoral Dissertation. Curtin University, Science and Mathematics Education Centre, Perth, 1997
6. H.M. Evans, A study of students' cultural background and teacher-student interpersonal behaviour in secondary Science classrooms in Australia. Unpublished Doctoral Dissertation. Curtin University, Science and Mathematics Education Centre, Perth, 1998
7. M. Brekelmans, Th. Wubbels, P. den Brok, Teacher experience and the teacher-student relationship in the classroom environment. in Studies in Educational Learning Environments: An International Perspective, (New World Scientific, Singapore, 2002), (pp. 73–100), ed. by S.C. Goh and M.S. Khine
8. A.M. Klem, J.P. Connell, Relationships matter: Linking teacher support to student engagement and achievement. J. Sch. Health **74**(7), 262–273 (2004)
9. P. den Brok, M. Brekelmans, T. Wubbels, Interpersonal teacher behaviour and student outcomes. School Effectiveness School Improvement **15**(3/4), 407–442 (2004)
10. P. den Brok, J. Levy, M. Brekelmans, T. Wubbels, The effect of teacher interpersonal behaviour on students subject-specific motivation. J. Classroom Interact. 40(2) (2005)
11. E.A. Hanushek, S.G. Rivkin, Generalizations about using value-added measures of teacher quality. Am. Econ. Rev. **100**, 267–271 (2010)
12. K.G. Lewis, Pathways toward improving teaching and learning in higher education: international context and background. New Dir. Teach. Learn. **2010**(122), 13–23 (2010)
13. R. Arum, J. Roksa, *Academically adrift: Limited learning on college campuses* (The University of Chicago Press, Chicago, IL, 2011)
14. C.N. Loes, K. Saichaie, R.D. Padgett, E.T. Pascarella, The effect of teachers behaviour on students inclination to inquire and lifelong learning. Int. J. Scholarsh. Teach. Learn. **6**(2) (2012). Article 7
15. Y. Mehdipour, D. Balaramulu, The influence of teachers behaviour on the academic achievement. Int. J. Adv. Res. Technol. **2**, 2278–7763 (2013)
16. B.K. Hamre, R.C. Pianta, J.T. Downer, J. DeCoster, A.J. Mashburn, S.M. Jones, M.A. Brackett, Teaching through interactions: testing a developmental framework of teacher effectiveness in over 4000 classrooms. Elem. Sch. J. **113**, 461–487 (2013)

17. B. Hamre, B. Hatfield, R. Pianta, F. Jamil, Evidence for general and domain-specific elements of teacher-child interactions: associations with preschool children's development. Child Dev. **85**, 1257–1274 (2014)
18. R. Chetty, J.N. Friedman, J.E. Rockoff, Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. Am. Econ. Rev. **104**, 2593–2632 (2014)
19. C.A. Hafen, B.K. Hamre, J.P. Allen, C.A. Bell, D.H. Gitomer, R.C. Pianta, Teaching through interactions in secondary school classrooms: revisiting the factor structure and practical application of the classroom assessment scoring system-secondary. J. Early Adolesc. **35**, 651–680 (2015)
20. S. Gershenson, Linking teacher quality, student attendance, and student achievement. Educ. Finance Policy **11**, 125–149 (2016)
21. D. Blazar, M.A. Kraft, Teacher and teaching effects on student's attitudes and behaviours. Educ. Eval. Policy Anal. **39**(1), 146–170 (2017)
22. M.S. Bartlett, A note on the multiplying factors for various chi square approximations. J. Roy. Stat. Soc. B **16**, 296–298 (1954)
23. S. Chatterjee, A.S. Hadi, Regression Analysis by Examples, 5th edn. (Wiley, 2014)
24. R. Myers, *Classical and Modern Regression with Applications*, 2nd edn. (Duxbury Press, Boston, MA, 1990)
25. J.P. Stevens, *Applied Multivariate Statistics for the Social Sciences*, 4th edn. (Erlbaum, Mahwah, N.J., 2002)

Smart Water Metering Implementation

Urja Mankad and Harshal Arolkar

Abstract Throughout the past years, the Indian government is putting its best efforts in developing smart cities by applying smart solutions for infrastructures and services. Smart water management is one of the key service areas where now it is a high time to work for. Many industries and researchers are contributing to incorporate smart techniques for smart water system like sensor-based monitoring, real-time data transmission and controlling, leakage management, water distribution management and many more. Nevertheless, the design and proposal of such a smart water system are still not fairly standardized. The enormous applications of smart water management still do not have systematic framework to guide real-world design and deployment of a metering system. To address this challenge, a framework is designed and tested with prototype implementation. This paper mainly focuses on minimizing smart water metering application development efforts. The framework uses open-source technologies and has mechanisms like device installation, water usage statistics, bill payments and a distributed data transmission architecture.

Keywords IoT · Smart city · Smart water metering framework · Smart water network · LoRaWAN

1 Introduction

The exponential growth in networking has given rise to Internet of things (IoT) which today touches all aspects of human life. Smart water networks (SWN) have enormous possibility to enable more efficient water usage management [1]. There are heterogeneous IoT devices, sensors and technologies available from different vendors to form smart water networks [2]. The SWN consists of mechanism devised for deployment of smart devices, control the devices, collect the data, store and analyze

U. Mankad
L.J. University, Ahmedabad, India

H. Arolkar (✉)
GLS University, Ahmedabad, India
e-mail: harshal.arolkar@glsuniversity.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_68

the data as part of a smart water network [3]. In this context, this paper discusses different types of water meters that are available in India, how to communicate with these smart devices and collect the water usage data to generate the bill according to the individual customer's consumption of water. Different utility providers want to charge their customers for their water usage where a low-cost highly secure system design is a challenge [4]. To address this challenge, Sect. 2 discusses the related work carried out to design a smart water metering framework which gives the direction to propose a compact framework with low-cost and high-security features. Section 3 describes different components of the proposed smart water metering framework. Section 4 gives the prototype implementation and testing results. Section 5 gives future direction, and Sect. 6 concludes with the future development and scope of implementation.

2 Literature Review

Lloret et al. [5] have given an idea for an IoT-based architecture for smart metering where they have discussed different components and networking technologies which are needed to form a smart metering system for electricity, water and gas. Their review mainly focused on smart meters features, current communication protocols and big data analytics.

Pacheco et al. [1] have discussed about the security in the smart water system. Here, they have presented an IoT-based hierarchical architecture for smart water services, which mainly focuses on the connection type and communication mechanisms among smart meter devices and their end-user applications

Suresh et al. [6] have designed an application that will help to read the meter reading and generate bill for water distribution in large urban areas. Their smart meter system consists of a hardware component called EIM which communicates on radio frequency and have Ethernet stack and port for interface to TCP/IP network or Wi-Fi or GSM/3G/4G router connectivity. Proposed architecture is not for distributed network.

Shahra et al. [7] have identified the challenges of heterogeneous devices used to form a smart water network. Their main aim is to provide a SWN architecture and its components. According to them, the smart water network must have five layers which includes water meters at physical layers which have sensing, controlling layer above it which collects and communicates the water usage data which is then managed, analyzed and displayed at the top most layer.

Gonçalves et al. [8] provide an IoT framework for smart water management, having efficiency in measuring and identifying leaks. They have worked on water usage data analysis to identify the leakage problems in water supply system. They have used complex event processing (CEP) technology and declarative business processes to control the operational activities of water supply systems.

Li et al. [4] studied and classified many literatures considering smart water system framework, but none of them is complete and requires future research work. They

provided a conceptual framework of smart water system with smartness and cyber wellness as their metrics. For research recommendations, they are suggesting a systematic framework, cyber security of water meter network, making water metering more robust and end-user data analysis.

Mourtzios et al. [9, 10] have designed a custom-based IoT gateway that can support multiple protocols, which helps to connect different types of water meters and water valves. The gateway communication is bidirectional with the end devices that operate on wM-Bus and LoRa as a wireless communication protocol. They are developing system infrastructure for the management of water consumption in Thessaloniki based on fix wireless network where Internet connectivity is provided via GSM and NB/IoT.

Alvisi et al. [11] have given a solution for smart water metering middleware where interoperable wireless IoT middleware interfacing with smart meters functioning with different protocols. They have used Raspbian Jessie with Raspberry Pi 3 Model B connected with wireless radio modules. M-Bus and LoRa protocols are used as wireless protocol manager. Data commands are sent over MQTT [12].

Fuentes and Mauricio [13] have proposed an architecture design for water consumption system with leak detection algorithm. Here, the data collected from the smart meter are preprocessed by the local server and sent to the cloud for leak detection. For local server, they have designed a gateway with RaspBerian OS on Raspberry Pi and have used Node-RED for preprocessing and data analysis. Mankad, Urja, and Harshal A. Arolkar has provided a detailed study of the open source framework OSGP/GXF for implementing Smart Water Metering [14].

Literature review concluded that components needed for smart water metering network [1, 5–8] are as follow.

2.1 Smart Water Meter and Gateway Devices

Smart water meter is a digital electronic device that collects water consumption at real time and provides data to the water providers. Meters are of different types, but for automatic meter reading, electromagnetic or fully electronic water meters are preferable to collect real-time data [15]. Electronic meters are best suitable for long-distance communication with better data rate as compared to others, but it requires separate power management module. As meters work on battery power, it requires power backup system and energy optimization techniques [5, 15]. Meters must be having some communication protocol implementation which can be used to send data securely on IoT platform. Data can be temporarily stored in the meter itself depending on its storage capacity and battery power. Meter indication for fault tolerance and alarm mechanism should be there. Gateways can be custom made to achieve interoperability between heterogeneous protocols and vendor-specific devices [9].

2.2 Connection and Communication Mechanism

Meter must be connected to a gateway and gateway must be connected to an IoT platform where data storage, data processing and analysis reports can be generated. Connection can be done through a GPRS, GSM, NB-IoT, UMTS, LTE network or through Wi-Fi or Bluetooth connectivity [5]. Communication can be done using RESTful and HTTPs protocols. Request and response objects are sent and receive among the metering device and gateway in a JSON format. DSS and MQTT protocols are also used which is having publish subscribed method to communicate with TLS/SSL security [7]. wM-BUS and LoRa work as IoT radio wave protocol are two noteworthy technologies for communication in smart metering applications [16, 17].

2.3 Data Storage

Water usage data need to be stored and processed at centralized location. It can be cloud-based NoSQL database, or it can be any server-based RDBMS. It will be a huge data, so enough storage capacity is required [5]. Data must be collected real-time and periodically different reports should be generated for water consumption details and billing.

3 Proposed Framework for Smart Water Metering

3.1 Architecture

The architecture of the proposed framework is divided into four layers called device layer, gateway layer, IoT platform and user application layer. Figure 1 shows the layered architecture of the said prototype. The four layers are discussed below:

Smart Water Meter—Device Layer: At this layer, smart water meters will work with flow sensors. Smart meters will have battery power. They will have wireless transceiver which will provide real-time water flow information to the gateway layer. This communication will be carried out using LoRa WAN protocol [18].

Gateway Layer: Gateway will collect water flow information from the meters on real-time basis. These data are stored temporarily in local storage where they are consolidated and then sent to the platform for further processing. It is an edge computing layer which will communicate with the IoT platform using HTTPs protocol.

Fig. 1 Smart water metering layered architecture

IoT Platform: IoT Platform composed of further four sub-layers as follow:

Web API Layer: It will provide Web interface to internal logic. Gateways will post the usage data through Web API. User interfacing applications will access metering and other data through this layer.

Platform Logic Layer: This layer will contain logic to receive, aggregate and store usage data.

Functional Layer: It will provide all end-user functions like administration and maintenance, bill generation, reporting and customer portal.

Data Storage Layer: Database will store all the customer information along with their meter assignment details. Monthly water usage will also be recorded for each water meter and its customer. Also, the payment details will be stored against the generated and stored invoices.

User Application Layer: This layer represents different types of user applications from where end users will access the functionalities of platform. Using the mobile application, consumers can see their daily, monthly and year-wise water usage. Also, they can pay their bills at their convenience. Administration dashboard helps the admin of the utility provider to register the users and their meters. Admin can also view the water consumption and payment details of all the users.

3.2 Modules

To implement the proposed framework, we needed water meter with flow sensor, gateway, communication protocol, database and an IoT-based application. Figure 2 shows different components connected with each other to implement smart water meter system.

Components shown in Fig. 2 interact with each other in a step-by-step manner as follow:

Collection of Data: Water flow sensors will sense the amount of water flow per second, and according to that, the meter's magnetic shaft will rotate to display reading. Water meters can be of different types like displacement water meters, velocity water meters, electromagnetic meters and ultrasonic meters. Out of them all, electromagnetic meters are more suitable to display accurate flow reading as its turbidity is high compared to rest of conventional meters and can handle rapid change in flow. Magmeter's performance is not affected by temperature, pressure or viscosity.

Transmission of Data: Customer meter number and reading will be sent from the smart water meter to the gateway using LoRa WAN communication protocol. Per day, two times data are sent which is depending on network load and channel availability.

Fig. 2 Smart water metering architecture component diagram

Aggregation of Data: The data received from a meter will be sum aggregated on daily basis at the gateway. This is edge computing, which is useful for minimizing the network load and storage space on cloud server.

Processing of Data: Data will be processed at the IoT platform to generate monthly usage report, which is the bill to be paid by the customer. Processed data will be shown to the customers in the form of line chart and bar graphs.

Storing Data: Customer personal information with their location address and contact details will be stored. Meter information like meter ID, meter type and assigned to which customer will be stored. Daily water usage will be stored with customer's name and meter ID. All these data will be stored in real-time database in cloud.

3.3 Use cases

IoT-based implementation of smart water metering system will have different use cases as discussed below, and Fig. 3 represents the use cases.

Register Meter: When a meter is installed at customer site, he has to register himself in the system where personal details need to be feed. After registration, sign up process is carried out where meter need to be connected using Bluetooth connection where a random key is generated and assigned to meter as a key. That key and customer ID is stored in the database.

Fig. 3 Use case diagram

User Management: Admin can view each user and their history. Admin can assign meter to the registered user. If the user pays the bill or if there are pending payments, then admin can view details.

View Usage Report: Graphical representation can be used to facilitate the users to analyze their daily and monthly usage. Paid and unpaid bills can be shown in different colors to give alerts to the customers.

Payment: Online payment gateways can be used for the bill payments.

Figure 3 shows different use cases that can be implemented as a minimum requirement to test the proposed framework for smart water metering application. The system will contain variety of users who are connected to the platform for performing different types of operations using platform API. There may be variety of applications like Web portal, mobile applications and so on to facilitate the utility providers and customers.

Smart water meter will just measure and post the data. Meters will be allocated to the customers by the administrator (water provider) using Web portal or any mobile application. Using the application, admin can manage users and can view the customers' usage and payments. Customers should get notifications for their water usage and payment reminders. Different types of reports can be viewed by them for their convenience.

4 Prototype Implementation

To implement smart water metering prototype, the proposed framework is implemented with different components as discussed below.

4.1 Water Meter

Three water meters are taken to implement the prototype. The flow sensor is fitted in line with the water line and measures the flow of liquid that moves through it. Sensor detects each revolution and generates an electrical pulse for every revolution. Total flow of the water can be calculated by counting the number of pulses generated by the sensor.

4.2 Gateway

One gateway module is implemented which is connected with the three water meters and collects its meter reading and stores the data on cloud-based server. Gateway has microcontroller unit having Wi-Fi module and a LoRa module inbuilt.

4.3 Communication Protocol

The meter to gateway communication will occur using LoRaWAN protocol, and the gateway to cloud communication is using Internet.

4.4 Application

To interact with all these components of the system, mobile-based application is developed, where the customers and *n* domain of utility provider will have their own login to perform above mentioned use cases.

Figure 4 shows the implementation diagram for smart water metering system.

In the implemented system, the meters send the water consumption reading to the gateway periodically. The gateway will collect all three meter's data frequently and aggregate day-wise usage and send it to the database server. Implemented prototype is tested for communication delay and packet lost scenario. It is found that as the readings are collected after every 6 h and aggregated for daily consumption, if any

Fig. 4 Implementation diagram for smart water metering system

data packet is lost and meter readings are not taken once a while; then in the next cycle, it will get fresh reading which will not affect the overall water consumption and billing system. LoRa communication delay is found to be just 0.4 s, which is negligible.

5 Future Work

Currently, the framework is tested with a prototype model having just three water meters for metering system, but in future, the framework can be tested for meter management, leakage management, water line fault detection and many more. This prototype can be simulated to test the network load and feasibility to implement the smart water network in the cities.

6 Conclusion

An IoT-based smart water metering framework has been designed and prototype implemented to help the utility providers and water consumers to smoothen out the process of consumption tracking and billing. A mobile app is developed to easily access the data and information from both the ends. The researchers have used an open-source technology, so the prototype implementation cost is not too high and the data are stored and retrieve using a secure cloud platform. Implementing this type of system in highly populated cities of any developing countries like India, where there is much wastage and scarcity of water, we can control the usage and can save our natural resource.

References

1. J. Pacheco et al., IoT security framework for smart water system. in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)* (IEEE, 2017)
2. U. Mankad, H.A. Arolkar, Survey of IoT frameworks for smart water metering. Management **24**, 26 (2018)
3. S. Sukode, S. Gite, H. Agrawal, Context aware framework in IoT: a survey. Int. J. **4**(1), (2015)
4. J. Li, X. Yang, R. Sitzenfrei, Rethinking the framework of smart water system: a review. Water **12**(2), 412 (2020)
5. J. Lloret et al., An integrated IoT architecture for smart metering. IEEE Commun. Mag. **54**(12), 50–57 (2016)
6. M. Suresh, U. Muthukumar, J. Chandapillai, A novel smart water-meter based on IoT and smartphone app for city distribution management. in *2017 IEEE Region 10 Symposium (TENSYMP)* (IEEE, 2017)

7. E.Q. Shahra, W. Wu, M. Romano, Considerations on the deployment of heterogeneous IoT devices for smart water networks. in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (IEEE, 2019)
8. R. Gonçalves, J.J.M. Soares, R.M.F. Lima, An IoT-based framework for smart water supply systems management. Future Internet **12**(7), 114 (2020)
9. C. Mourtzios, K. Galanis, N. Papadimitriou, An experimental study of a multi-protocol IoT gateway. in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)* (IEEE, 2020)
10. C. Mourtzios et al., Work-In-progress: SMART-WATER, a νovel τelemetry and remote-control system infrastructure for the management of water consumption in Thessaloniki. in *Interactive Mobile Communication, Technologies and Learning* (Springer, Cham, 2019)
11. S. Alvisi et al., Wireless middleware solutions for smart water metering. Sensors **19**(8), 1853 (2019)
12. OASIS, OASIS MQ Telemetry Transport (MQTT) Technical Committee. https://www.oasis-open.org/committees/mqtt/charter.php. Accessed 10 Nov 2020
13. H. Fuentes, D. Mauricio, Smart water consumption measurement system for houses using IoT and cloud computing. Environ. Monit. Assess. **192**(9), 1–16 (2020)
14. U. Mankad, H.A. Arolkar, A study of the opensource framework OSGP/GXF for implementing Smart Water Metering. in 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) (IEEE, 2020)
15. X.J. Li, P.H.J. Chong, Design and implementation of a self-powered smart water meter. Sensors **19**(19), 4177 (2019)
16. A. Ikpehai et al., Low-power wide area network technologies for Internet-of-things: a comparative review. IEEE Internet Things J. 1 (2018)
17. V. Slaný et al., An integrated IoT architecture for smart metering using next generation sensor for water management based on LoRaWAN technology: a pilot study. Sensors **20**(17), 4712 (2020)
18. LoRa Alliance Organisation, LoRaWAN® Specification v1.1. https://lora-alliance.org/resource-hub/lorawanr-specification-v11, Accessed 12 Nov 2020

Artificial Bee Colony Optimization-Based Load Balancing in Distributed Computing Systems—A Survey

Vidya S. Handur and Santosh L. Deshpande

Abstract Distributed computing allows the interoperability of components in distributed system in which software or hardware components located at networked computers coordinate and communicate their actions by message passing. The tasks in such a system are carried out independently. In distributed computing systems, load balancing is one of the issues, which is a means to distribute the tasks such that the computational nodes are neither overloaded nor underloaded, and the performance of the system is improved. Among the various solutions proposed for load balancing, metaheuristic-based algorithms are one of them. This paper discusses the variants and recent developments of artificial bee colony optimization algorithm for solving load balancing in distributed systems. As a result of load balancing, various performance metrics measured are throughput, response time, makespan, CPU utilization, memory utilization, and network utilization. The performance of the optimization algorithms is also measured with the time taken to converge in finding the optimal solution for load balancing.

Keywords Distributed computing systems · Load balancing · Artificial bee colony optimization · Convergence rate

1 Introduction

With the development in technology and increase in the data, getting optimal solutions to data intensive applications are a challenge. To overcome this, there is need for techniques that handle the data efficiently which in turn makes it necessary that all the resources being utilized appropriately. Resource management deals with managing the resources in such a way that all the resources are utilized efficiently to improve the performance of the system. Resource management is achieved by proper resource

V. S. Handur (✉)
KLE Technological University, Hubballi, India
e-mail: vidya_handur@kletech.ac.in

S. L. Deshpande
Visveswaraya Technological University, Belagavi, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_69

Fig. 1 Distributed system

allocation to the applications so to maximize the efficiency of the system. As the resource allocation varies with requirements of the users, there is a need to check for the availability of the resources and allocate. The resource allocation should be performed such that no resource is either underutilized or over utilized. Load balancing is a technique to discover the inefficiently used or over used resources and to balance the load among the resources proportional to their capacity [1, 2]. Load balancing is a major concern in a network that is distributed in nature. Figure 1 shows various computational nodes that communicate and coordinate in distributed manner to accomplish the task. The major concerns with distributed system are that all the computational nodes are autonomous, and there is no global clock. Other characteristics of distributed system include sharing of resources, unreliability of systems, openness, and heterogeneity of computational nodes. Due to the distributed characteristics, such systems have the advantage of better performance, availability, and scalability [3–5]. There are wide range of applications of distributed systems such as fog computing, edge computing, grid computing, cloud computing, World Wide Web, and automated banking systems.

With the increase in the technology and the data rates, there is need for techniques to handle the huge traffic to enhance the performance of the system. Also due to the differences in the computational capacities and heterogeneity of computing and network resources, the pattern of job arrival and the workloads on each node may vary significantly because of which the performance of the distributed system may deteriorate. Therefore, there is a need to distribute the workloads proportional to the capacity of the nodes. This imbalance of load can be addressed with load balancing.

The paper is systematized as follows: Sect. 2 briefs about load balancing, Sect. 3 introduces the artificial bee colony (ABC) optimization algorithm, Sect. 4 briefs the recent developments of the ABC algorithm for load balancing and followed by conclusion.

2 Load Balancing

Load balancing allows distributing the load on the system such that the computational nodes receive the load proportional to their capacity. The load balancing is either static or dynamic [6–8]. Load balancing is static when the load is assigned to the computational nodes without considering the current load of the nodes and the resources needed to carry out the tasks. Such an assignment is preferable when the load on the system is constant, and it provides good result for homogeneous environments. Due to growth in the technology usage, the traffic is not constant and is increasing at galloping speed, and static method of allocating load to the nodes is inefficient [8]. There is a need for appropriate technique to distribute the traffic among the nodes in the system. Therefore, to meet the current needs of the system, dynamic load balancing is suitable. In dynamic load balancing, the present conditions of the workload on the computational nodes are considered before executing the load. Some of the load balancing methods are round robin, min–max, game theory-based [4], throttled algorithm, etc. Nature-inspired metaheuristic techniques have also been proposed [9]. These techniques have the characteristics that they are decentralized, self-organizing, agent-based, flexible, robust, scalable, and adaptive to changes in the network. Due to ease of implementation, swarm intelligence-based techniques have been extensively employed [10].

With dynamic load balancing algorithms, there are several research questions that need to be addressed such as how frequently the load balancing should be invoked, which host initiates the decision of load balancing, how to collect the load information of the nodes, and migration of load between the hosts. Various researchers have proposed many different solutions for load imbalance. With advancement in the technology, load balancing has become more significant in fog and edge computing. The goal of load balancing is to improve the overall performance of the system by reducing the load imbalance. The various performance metrics measured are the response time, makespan, resource utilization, fault tolerance, throughput, and migration time. This paper focuses on how the artificial bee optimization algorithm and the metaheuristic-based solution can be applied for balancing the load in distributed systems.

3 Artificial Bee Colony Optimization

Artificial bee colony (ABC) optimization algorithm was originally published by Karaboga in 2005 for numerical optimization problems. ABC optimization algorithm is a nature-inspired, swarm intelligence, metaheuristic algorithm that is based on the foraging behavior of bees [11–13]. According to Ullah et al. [14], the ABC algorithm is adaptive to the heterogeneous environment and also to the varying nature of load [15]. The algorithm has few advantages over other swarm intelligence-based algorithms. The advantages are that the algorithm uses very few control parameters,

and it is simple and robust, can be easily hybridized with other algorithms of optimizations, and has fast convergence rate [16]. The algorithm also has better exploration capability as compared to other swarm intelligence algorithms [17]. Some of the application areas of ABC algorithm are cluster analysis, software testing, cluster problem optimization, structural optimization, multilevel thresholding, MR brain image classification, advisory system, numerical assignment problem, bioinformatics, face pose estimation, parameter estimation in software reliability models, wireless sensors [18], big data analytics, edge and fog computing [12, 19].

3.1 Phases of Artificial Bee Colony Optimization Algorithm

Artificial bee colony optimization algorithm has three types of bees: employed bees, onlooker bees, and scout bees.

- Employed bees: The task of the employed bees is to arbitrarily search for food source which represents the potential solutions for the problem. These bees share the information of food source by dancing in the bee hive's dance area. This shared information specifies the quality of solution. The number of employed bees and the number of food sources for the bee hive are equal.
- Onlooker bees: These are the bees waiting in the dance area to assess several dances before choosing a position of food source. The selection is based on the probability proportional to the quality of that food source.
- Scout bees: These bees randomly look for possible new food source [13, 14, 20]

Figure 2 represents the flow diagram of basic ABC optimization algorithm.

4 Recent Developments of ABC Optimization for Load Balancing

With the increase in data and variations in the data rates, there is a need to manage the varying traffic across the system. Load balancing is a technique to balance the load so that all the resources are utilized efficiently. There are several solutions proposed for load balancing. This section introduces few of the recent solutions proposed using ABC optimization algorithm.

The authors in Kruekaew et al. [6] have simulated ABC optimization algorithm for cloud computing by applying a heuristic method which uses certain rules or random methods to find an optimal solution. In the proposed approach, the heuristics is based on the priority for the process selection in the system. The priority criteria used are first come first serve, shortest job first, and largest job first. The system is tested under homogeneous and heterogeneous environments and has compared with other optimization algorithms like ant colony optimization and particle swarm

Fig. 2 Flow diagram of ABC optimization algorithm

optimization. In the proposed solution, the performance metric measured is makespan and algorithm with largest job first outperforms the other priority criteria.

According to the solution proposed in Hashem et al. [21], the task allocation is carried out by determining the processing time variation of each virtual machine (VM) with respect to the average processing time of all the virtual machines. The proposed solution determines the utilization of a virtual machine by setting a predefined threshold. The system is modeled to be non-pre-emptive. To solve load imbalance, some of the parameters determined are

- The processing time of host and each VM
- Average processing time of all hosts
- Average processing time of all VMs
- Load standard deviation.

The proposed algorithm is compared with round robin and throttled algorithms. The performance metric measured is the response time and the task execution time, and honey bee-based solution outperforms the other algorithms.

The solution proposed in Bhavya et al. [22] is that the authors assign the tasks dynamically according to the change in the user's demands. The algorithm works by grouping the virtual servers, and each virtual server maintains a queue of processes. After processing the request, the profit is determined. With high value of profit, the server stays and low value of profit causes a return to foraging. In this method, there is an overhead in computing the profit which affects the throughput. The load balancing is initiated by assigning the load to a virtual machine (VM) which is underloaded and with maximum throughput. If the load remains below the preset threshold, the load is assigned otherwise a VM with next highest value of throughput is selected. The performance metric measured is throughput.

The authors Korat and Gohe [23] proposed a solution based on multi-objective optimization Pareto dominance, and the weighted sum method is used for selecting the optimal virtual machine (VM) for balancing the load and for priority assignment to the task. All the virtual machines are grouped according to their load. The tasks considered are the pre-emptive tasks, and priority is used to migrate the tasks from one VM to another. To calculate the priority of the task, the parameters used are latency time of task, user type of task, expected priority of the task, and the length of the task. The performance metrics measured are the makespan and the throughput.

In the proposed study, Mallikarjuna et al. [24], the authors initially set various parameters for the virtual machines. The parameters set is million instructions per second (MIPS) which is a measure of the speed of CPU, storage space for each VM, RAM space, and the bandwidth. Each task is identified by the parameters task id and length of the task. The capacity of any virtual machine is calculated as in Eq. 1.

$$\text{Capacity} = \text{MIPS} \times \text{Number of CPUs} + \text{Bandwidth} \tag{1}$$

To balance the load, an iteration method is adopted to select the most appropriate virtual machine, and a newton gravity function is used. The performance metric measured is the resource utilization.

Table 1 Performance metrics measured

Publication	Performance metrics				
	Response time	Makespan	Resource utilization	Throughput	Task execution time
[22]				√	
[21]	√				√
[6]		√			
[23]		√		√	
[24]			√		

Table 1 summarizes the performance metrics measured by applying ABC optimization algorithm to balance load.

5 Conclusion

With the advancement in technology such as fog, cloud, and edge computing, load balancing or the offloading plays a significant role. There are a number of solutions to reduce the load imbalance in the network, however, there is still a requirement for the load balancing. Many researchers have proposed swarm intelligence-based solutions for load balancing. This paper is an attempt to show the recent developments of artificial bee colony optimization algorithm to solve load imbalance in distributed system. Among various performance metrics measured, convergence rate of the algorithm is one of them which needs to be addressed.

References

1. S. Handur Vidya, R.M. Prakash, Response time analysis of dynamic load balancing algorithms in cloud computing. in Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), vol. 2020 (2020), pp. 371–375
2. S. Handur Vidya, S. Deshpande, P.R. Marakumbi, Particle swarm optimization for load balancing in distributed computing systems–a survey. Turk. J. Comput. Math. Educ. (TURCOMAT) **12**(1S), 257–265 (2021)
3. V.S. Handur, S. Belkar, S. Deshpande, P.R. Marakumbi, Study of load balancing algorithms for cloud computing. in *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)* (2018), pp. 173–176
4. S. Penmatsa, A.T. Chronopoulos, Game-theoretic static load balancing for distributed systems. J. Parallel Distrib. Comput. **71**(4), 537–555 (2011)
5. S.B. Kshama, K.R. Shobha, A novel load balancing algorithm based on the capacity of the virtual machines. ICACDS CCIS **905**, 185–195 (2018)
6. B. Kruekaew, W. Kimpan, Enhancing of artificial bee colony algorithm for virtual machine scheduling and load balancing problem in cloud computing. Int. J. Comput. Intell. Syst. **13**(1), 496–510 (2020)

7. A. Hanamakkanavar, V.S. Handur, Load balancing in distributed systems—a survey. Int. J. Emerg. Technol. Comput. Sci. Electron. (IJETCSE) **14**(2) (2015). ISSN: 0976-1353
8. S.P. Belkar, V. Handur, Comparative study of static load balancing algorithms in distributed system using cloudsim. Int. J. Adv. Res. Basic Eng. Sci. Technol. (IJARBEST) **5**, 26–30, (2017)
9. S.T. Milan, L. Rajabion, H. Ranjbar, N.J. Navimipou, Nature inspired meta-heuristic algorithms for solving the load-balancing problem in cloud environments. Comput. Oper. Res. 159–187 (2019)
10. C. Bansal, A. Gopal, A.K. Nagar, Analysing convergence, consistency, and trajectory of artificial bee colony algorithm. IEEE Access **6**, 73593–73602(2018). https://doi.org/10.1109/ACCESS
11. Sudha Senthil Kumar, K. Brinda, R. Rathi, Angulakshmi, Jothi, Yash Vardhan Thirani: Honeybee foraging algorithm for load balancing in cloud computing optimization. Int. J. Eng. Sci. Comput. 7(12) (2017)
12. S.S. Ilango, S. Vimal, M. Kaliappan, Optimization using artificial bee colony based clustering approach for big data. Clust. Comput. **22**, 12169–12177 (2019)
13. D. Karaboga, An idea based on honey bee swarm for numerical optimization. *Technical Report TR06* (Erciyes University, Engineering Faculty, Computer Engineering Department, 2005)
14. A. Ullah, N.M. Nawi, J. Uddin, S. Baseer, A.H. Rashed, Artificial bee colony algorithm used for load balancing in cloud computing: review. IAES Int. J. Artif. Intell. (IJ-AI) **8**(2), 156–167 (2019)
15. A.L. Bolaji, A.T. Khader, M. Azmi Al-Betar, M.A. Awadallah, Artificial bee colony its variants and applications: a survey. J. Theor. Appl. Inf. Technol. **47**(2) (2013)
16. S. Anam, Multimodal optimization by using hybrid of artificial bee colony algorithm and BFGS algorithm. IOP Conf. Ser.: J. Phys.: Conf. Ser. **893** (2017)
17. D. Yazdani, M.R. Meybodi, A novel artificial bee colony algorithm for global optimization. in *International Conference on Computer and Knowledge Engineering* (2014)
18. S. Kaswan, S. Choudhary, K. Sharma, Applications of artificial bee colony optimization technique: survey. in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)* (2015), pp. 1660–1664
19. S.A. Zakaryia, S.A. Ahmed, M.K. Hussein, Evolutionary offloading in an edge environment. Egypt. Inf. J. (in press)
20. L. Shen, J. Li, Y. Wu, Z. Tang, Y. Wang, Optimization of artificial bee colony algorithm based load balancing in smart grid cloud. in IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia), vol. 2019, (2019), pp. 1131–1134
21. W. Hashem, H. Nashaat, R. Rizk: Honey bee based load balancing in cloud computing. KSII Trans. Internet Inf. Syst. **11**(12) (2017)
22. V.V. Bhavya, K.P. Rejina, A.S. Mahesh: An intensification of honey bee foraging load balancing in cloud computing. Int. J. Pure Appl. Math. **114**(11), 127–136 (2017)
23. C. Korat, P. Gohe, A novel honey bee inspired algorithm for dynamic load balancing in cloud environment. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. 4(8), (2015)
24. B. Mallikarjuna, P. Venkata Krishna, A nature inspired bee colony optimization model for improving load balancing in cloud computing. Int. J. Innov. Technol. Exploring Eng. (IJITEE) 8(2S2) (2018). ISSN: 2278-3075

Distributed and Scalable Healthcare Data Storage Using Blockchain and KNN Classification

Manjula K. Pawar, Prakashgoud Patil, and Amit Singh Patel

Abstract Medical data from patients is saved electronically in the healthcare industry. The paper suggests storing medical data on a distributed on-chain utilizing InterPlanetary File Storage (IPFS) and blockchain. The use of InterPlanetary File Storage can help an organization manage the storage space efficiently. The proposed method gives details about the intersection of artificial intelligence and blockchain. This proposed approach eliminates the requirement for a trusted centralized authority, intermediates, and transaction records, resulting in increased efficiency and safety while retaining high integrity, scalability, reliability, and security. It also emphasizes smart contracts to manage and control all interactions and transactions between all healthcare participants. Using artificial intelligence, a revolutionary new system can be built in which a huge amount of medical data can be processed efficiently and can be used to predict diseases. The paper utilizes the assistance of binary classification method by K-nearest neighboring algorithm for machine learning to foresee the patient's heart health. Using InterPlanetary File Storage (IPFS) improves the storage system framework by making data distributed and, subsequently, more reliable, scalable, and accessible to the clients.

Keywords Artificial intelligence · Machine learning · Blockchain · InterPlanetary File Storage · KNN classification · Scalability

1 Introduction

A lack of patient agency, slow access to the medical records, a lack of optimized system interoperability, and the quality and amount of data for medical research are all issues that blockchain and machine learning technology can help. Blockchain technology facilitates and secures data storage easier and more secure. Doctors can

M. K. Pawar (✉) · P. Patil · A. S. Patel
KLE Technological University, Hubballi, Karnataka, India
e-mail: manjulap@kletech.ac.in

P. Patil
e-mail: prakashpatil@kletech.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_70

see a patient's whole medical history. Still, researchers can only access statistical data rather than any personal information. Machine learning may use this data to discover trends and make accurate predictions, providing more assistance to patients and research domains where accurate data is necessary to predict trustworthy outcomes. Blockchain technology offers a decentralized storage method that allows a medical record of any patient registering themselves to the network to be easily shared between peers in a healthcare system (mainly targeting hospitals having more than one center to share the data among them). This technology ensures data consistency, chain integrity, block immutability, and data privacy. Those are all crucial necessities in today's medical record storing system.

Though blockchain has better features, it faces certain challenges such as scalability [1]. There are several solutions for improving scalability for blockchain systems [1–3]. IPFS is a Merkle tree-based peer-to-peer (P2P) distributed storage format for storing huge data in blocks. IPFS preserves data with its content-addressed hash in a hash table which is distributed and using the history of version control. It removes duplicates. IPFS is a block storing system with high throughput, transaction security via hash mapping, and network peers' concurrent access to transactions. To access the data next time with guaranteed quick access, it stores local hash values of requests with the high frequency.

The convergence of AI and blockchain technology could lead to substantial breakthroughs in healthcare, commercial development, and research. The proposed method uses AI-ML and blockchain technology in the dataset. The decentralized AI is chosen because it is being utilized in many healthcare areas to protect data misuse stored in hospital records, medical research labs [4–6], etc.

Instead of using "Location-Based Addressing" (LBA) IPFS [7] uses "Content-Based Addressing" (CBA). In LBA, whenever we want to search the dataset, we tell the exact location to the computer from where to fetch the data; this is called As LBA. But if the dataset is unavailable or say that the server is down, the data won't be fetched. The dataset may be available, or some other computer must have the dataset, but it cannot be accessed if the exact location is not provided. So to resolve this issue, IPFS has moved from LBA to CBA. In "content-based addressing," it uses hash values to search the required dataset of the patients.

IPFS [8, 9] uses hash values instead of storing the entire medical data records. Files are transferred or received using the hash values. It makes the work easier than sharing the whole dataset. But there can be tampering of files. As hash is used to request files, it can be verified. The file is asked with a specific hash, and when it is found, it will be checked to see if it matches. Deduplication is another mechanism used by hashing to address stuff. When numerous people publish the same file on IPFS, it will be produced only once, resulting in a very efficient network.

In the proposed methodology, AI for blockchain is used to implement the healthcare application by applying the KNN algorithm.

2 Related Work

In Ref. [10], the work in this study is motivated by the following observation: in blockchain networks, storage is a problem due to the append-only characteristic. In this paper, they offer an IPFS-based blockchain storage system that overcomes the limitations of existing blockchain storage. The author also campaigned for transaction access that is content addressed rather than simply location-based.

In Ref. [11], the author introduces explainable artificial intelligence (XAI) that generates the explanations for auditing the responsibility by regulators in case of catastrophic failure. But since AI works in a centralized manner which has security limitations. And there may be a chance of tampering with the data for their benefit to avoid penalty. To overcome these issues author adapts the framework using blockchain as proof-of-authenticity about XAI. It stores the data in IPFS to overcome Ethereum storage limitations. Smart contracts are designed and deployed to retrieve the explanations and supervise the storage. To induce cryptographic security, it uses the hash value and is stored in blockchain. The author also performs cost and security analysis of implemented framework.

In Ref. [12], author uses a modified IPFS and is used to deploy Ethereum smart contracts. These smart contracts provide the ability to share files according to access control. To maintain the list of access controls, smart contracts are used. To enforce it, they modified the IPFS software. Its smart contracts interact with it whenever a file upload is done, a download, or a file transfer.

In Ref. [13], the author overcomes the drawbacks of POW and POS. The author implements the proof of AI with an AI algorithm for selecting the miners by training the neural network model without compromising the basic properties of blockchain technology. Furthermore, the author implements the healthcare domain dataset on top of the blockchain network to train the NN model. This implementation provides an advantage to the state-of-the-art consensus models by increasing efficiency.

In Ref. [14], the combination of blockchain, and AI is called decentralized AI. As AI is centralized and there are chances of data being tamper, there is no guarantee that source data is safe and not misused. Contrarily, blockchain is decentralized in nature and provides more authenticity to the data source, and when merged with AI, it provides unchangeable and reliable characteristics.

In Ref. [15], the author discusses that AI is decision-making on systems like human thinking, whereas blockchain is a decentralized platform in a secure way. The author discusses that integrating AI and blockchain builds decentralized AI, which is nothing but decision-making on a secured platform for sharing the ledgers without involving a third party. The author discusses issues and problems involved in the integration of AI and blockchain. It helps in identifying the malware blocks in the system.

3 Proposed Model

As in Fig. 1, the proposed model presents the use of content addressed (rather than location-based addressing) hash ID of the data as the hash of block in the blockchain. The user enters the medical data for each suggested parameter, and this data is stored in the backend. This data follows a particularly strict format of comma-separated values such that query data dimension matches training data dimension for the used machine learning model. It undergoes two stages such as data classification and data storing.

Fig. 1 Proposed framework

3.1 Data Classification

The algorithm, Algorithm 1, takes account for preparing the trained ML model. First, the raw medical data is preprocessed. Then the preprocessed medical data is divided randomly into 80% to train the model, and the remaining 20% to test the model. Using this training data, the model is KNN model is trained with k value as 3. The trained model is then saved for future use, and preprocessed test data is also saved for testing the model in the future.

Algorithm 1 for ML Model Generation
Input: Raw data
Output: Hash of the data

1. Begin
2. Split the full data into train data(80%) and test data (20%)
3. Select the preprocessed training data
4. Declare the k value equal to 3
5. Train the model with training data
6. Now save the trained model using pickle.dumps("model_ name")
7. Save the data
8. End.

3.2 Data Storing

The algorithm for data storing described below in Algorithm 2 is used to classify the medical data and show its result into the front end. First, the application's backend is connected to infura sever and used to connect to the IPFS network. The user enters the IPFS hash address of the data and stores the value to preprocess and get evaluated. Here, the saved ML model is loaded and used to classify the medical data for the patients' heart condition.

Algorithm 2 Algorithm for fetching data from IPFS
Input: IPFS hash
Output: Computed result for the patient

1. Begin
2. api=ipfsapi.Client(host='https://ipfs.infura.io/', port=5001) // connect to infura serverwith port number 5001
3. hashstring = request.POST['hashfromweb'] //Get the hash value from user from frontend
4. api.get(hashstring) //Get the data at this hash from ipfs api and download the file name hashstring
5. str = open(hashstring,'r').read() //Open and read the content of the file hashstring into str
6. preprocess(str)7 loadedmodel = pickle.load(open('modelFileName.sav','rb')) //Load saved trained model tothis django application

```
7.  prediction = loadedmodel.predict(str)[0] //predict the output of the model after
    fetchingtest data in it and store its result in prediction
8.  if(prediction==0): result="The Patient Is Not Having Heart Disease"
    if(prediction==1): result="The Patient Is Having Heart Disease"
9.  else
10. result = "Invalid Result"
11. End.
```

4 Implementation

The implementation of this model has been divided into two different modules: data storing and data classification. The modules are designed carefully and meant to work independently. There are two different environments for implementations, one for React-based Web applications and the other for Django-based Web applications. The modules are carefully planned to be separated because both models have different functionalities. The first user has to store the data on to IPFS and attest its hash to the smart contract. While another module contains the ML model required to run on Python, thus Django makes a suitable choice for this module's implementation. To implement this model a Linux-based (Ubuntu) OS was used because it is more secure, less vulnerable, less graphic extensive, and does not collect a lot of data, unlike Microsoft Windows.

The specification for hardware is Intel(R) i5-8300H CPU @4 GHz, OS: Ubuntu × 64, 8 GB of primary storage, 128 GB SSD, and 1 TB HDD of secondary storage. The details of each implemented module are given as follows.

4.1 Data Storing

This module is built on a React-based Web application [9, 16, 17], and its backend includes the use of IPFS API, the user uploads the patient's medical data to the IPFS network. Stores its content-addressed hash on smart on file by the name "attestor. Sol" smart contract (written on solidity language) to be attested on blockchain for calling and implementing various custom functionalities stated in the smart contract to be implemented using the blockchain.

User needs to send ETH value of 0.01 ETH from faucet address to burner address. For completing any transaction on the Ethereum chain, some amount of ETH is required as a gas price. A burner address of zero balance cannot initiate the transaction for creating a new block of Hash Id equals the hash value of the data uploaded to IPFS.

4.2 Data Classification

This module is essential for classifying the medical data. This module is a Django Web application that takes the IPFS hash of the data from the frontend and uses this Hash ID to retrieve the data from the IPFS network. For testing, this data retrieved is test data and was not used to train the ML model. The trained ML model is loaded to the application's backend and used for predicting the output by fetching the test data. The output is sent to the frontend of the application and shows the result of classification done by ML.

5 Results and Discussion

The proposed model testing gets an accuracy of 88.52% when the value k value is set with 3. While uploading to the IPFS network, the size of data does not affect the upload time. Still, while uploading the same data again, the data gets uploaded in much less time because the data is already present in the network and hosted by the setup computer. Hence, uploading of the data is not needed again. Just the same hash is returned to the user.

In Fig. 2, the graph defines the comparison between data size vs. load time. Figure 2 shows that the time does not vary much as the size increases while uploading the data. There are very slight increases when re-uploading the data because it uses already uploaded hash value data, so time does not vary much on re-uploading.

Fig. 2 Data size vs. load time

Fig. 3 KNN algorithm performance

Figure 3 shows that it will reach the maximum score when it is set with the k value equal to 3, 7, and 8. The figure shows that the maximum score is 88.52%. This plot contains the k values as the x-axis and score or accuracy as the y-axis. While training the KNN model for different values of k from 0 to 20, the model gives different accuracy for different values of k. Among that values, $k = 3$, 7, and 8 gave the highest accuracy of 88.52%. So it is feasible to train the model among any one of the three values.

The standard model with centralized data storing models, such as large data centers, is purely dependent and governed by the storage providers. Any malfunctions, hacks, and data breaches will result in losing or leaking of sensitive medical data. But the proposed decentralized model is secured from problems like server malfunction because if a node is offline, another node can process the upcoming requests.

6 Conclusion

This paper presents designing and implementing a decentralized medical data storing platform using IPFS-based blockchain-based on smart contracts and this decentralized data to classify a patient's report for their heart condition using AI. The proposed model is more efficient and scalable than the conventional data soring model because it stores only the hash of the data instead of storing data in each blockchain block. This model demonstrates the use of machine learning on blockchain data. The model demonstrates the requirement of an account ID to add the block on the chain, which showcases that the security of the data is not compromised even the data is decentralized.

6.1 Limitations of the Study

The proposed methodology uses IPFS, which is a public protocol. Using a public protocol for data storing is not secure enough as medical data is sensitive and needs data confidentiality. However, the proposed method is implemented using a P2P (peer-to-peer) network. It ensures data confidentiality, integrity, and availability for retrieving data so that accessibility is maintained to the patient's data and does not allow any other person to view medical records. Also, IPFS improves the scalability by saving the space in blocks as it stores only hash values in blocks.

The future scope of the paper is to implement a decentralized storage system of healthcare data that can be shared among a group of hospitals with the same organization or hospitals within the same city [18].

References

1. M.K. Pawar, P. Patil, P.S. Hiremath, A study on blockchain scalability. in: *ICT Systems and Sustainability*, ed. by M. Tuba, S. Akashe, A. Joshi. Advances in Intelligent Systems and Computing, vol. 1270 (Springer, Singapore, 2021)
2. J. Liu, H. Tang, R. Sun, X. Du, M. Guizani, Lightweight and privacy-preserving medical services access for healthcare cloud. IEEE Access (2019)
3. M.K. Pawar, P. Patil, P.S. Hiremath, V.S. Hegde, S. Agarwal, P.B. Naveenkumar, Scalable blockchain framework for a food supply chain. in *Advances in Computing and Network Communications*, ed. by S.M. Thampi, E. Gelenbe, M. Atiquzzaman, V. Chaudhary, K.C. Li. Lecture Notes in Electrical Engineering, vol. 735 (Springer, Singapore, 2021). https://doi.org/10.1007/978-981-33-6977-1_35
4. S. Yu, K. Lv, Z. Shao, Y. Guo, J. Zou, and B. Zhang, A high performance blockchain platform for intelligent devices. in *Proceedings of 1st IEEE International Conference on Hot Information-Centric Networking* (Shenzhen, China, 2018), pp. 260–261
5. A. Ekblaw, A. Azaria, J.D. Halamka, A. Lippman, A case study for blockchain in healthcare MedRec prototype for electronic health records and Medical research data. in *Proceedings of IEEE Open Big Data Conference* (2016), p. 13
6. T.N. Dinh, M.T. Thai, AI and blockchain: a disruptive integration. Computer **51**(9), 48–53 (2018). https://doi.org/10.1109/MC.2018.3620971
7. Q. Zheng, Y. Li, P. Chen, X. Dong, An innovative IPFS-based storage model for blockchain. in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (2018), pp. 704–708. https://doi.org/10.1109/WI.2018.000-8
8. R. Kumar, N. Marchang, R. Tripathi, Distributed off-chain storage of patient diagnostic reports in healthcare system using IPFS and blockchain. in *2020 International Conference on Communication Systems Networks (COMSNETS)* (2020), pp. 1–5. https://doi.org/10.1109/COMSNETS48256.2020.9027313
9. T.T. Huynh, T.D. Nguyen, H. Tan, A decentralized solution for web hosting. in *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)* (2019), pp. 82–87. https://doi.org/10.1109/NICS48868.2019.9023837
10. R. Kumar, R. Tripathi, Implementation of distributed file storage and access framework using IPFS and Blockchain. in 2019 Fifth International Conference on Image Information Processing (ICIIP) (2019), pp. 246–251. https://doi.org/10.1109/ICIIP47207.2019.8985677

11. D. Malhotra, S. Srivastava, P. Saini, A.K. Singh,Blockchain based audit trailing of XAI decisions: storing on IPFS and Ethereum Blockchain. in *2021 International Conference on Communication Systems Networks (COMSNETS)* (2021), pp. 1–5. https://doi.org/10.1109/COMSNETS51098.2021.9352908
12. M. Steichen, B. Fiz, R. Norvill, W. Shbair, R. State, Blockchain-based, decentralized access control for IPFS. in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (2018), pp. 1499–1506. https://doi.org/10.1109/Cybermatics2018.2018.00253
13. N. Kumar, C. Parangjothi, S. Guru, M. Kiran, *Peer Consonance in Blockchain-based Healthcare Application using AI-based Consensus Mechanism* (2020), pp. 1–7. https://doi.org/10.1109/ICCCNT49239.2020.9225550
14. N. Nasurudeen Ahamed, P. Karthikeyan, A reinforcement learning integrated in heuristic search method for self-driving vehicle using blockchain in supply chain management. Int. J. Intell. Netw. **1**, 92–101 (2020). ISSN: 2666-6030. https://doi.org/10.1016/j.ijin.2020.09.001
15. M.K. Pawar, P. Patil, M. Sharma, M. Chalageri, Secure and scalable decentralized supply chain management using Ethereum and IPFS platform. in *2021 International Conference on Intelligent Technologies (CONIT)* (2021), pp. 1–5. https://doi.org/10.1109/CONIT51480.2021.9498537
16. M.A. Uddin, A. Stranieri, I. Gondal, V. Balasubramanian, Continuous patient monitoring with a patient centric agent: a block architecture. IEEE Access **6**, 32700–32726 (2018)
17. R. Kumar, R. Tripathi, Traceability of counterfeit medicine supply chain through blockchain. in *2019 11th International Conference on Communication Systems Networks (COMSNETS)* (IEEE, 2019), pp.568–570
18. S.D. Smriti, M. Hooda, Possibilities at the intersection of AI and Blockchain technology. Int. J. Innov. Technol. Exploring Eng. (IJITEE) **9**(1S) (2019). ISSN: 2278-3075

A Sustainable Green Approach to the Virtualized Environment in Cloud Computing

Anjani Gupta, Prashant Singh, Dhyanendra Jain, Anupam Kumar Sharma, Prashant Vats, and Ved Prakash Sharma

Abstract Cloud technology is a dynamic industry of communication and information technology (especially the Internet) that has revolutionized current computing while also introducing new environmental concerns. The cloud is a huge improvement over the conventional computing as it has redesigned the way businesses operate. With the cloud, the information is virtualized and reduces the need of infrastructural based model. Businesses can operate through this tangible area, significantly reducing energy consumption and the need for excessive material resources. In recent years, computational technologies and ideas have transitioned to distant data centers and pay-per-use hardware and software solutions. However, as the number of information centers supporting cloud-based applications grows, vast quantities of energy are consumed, resulting to high prices and greenhouse emissions in the atmosphere. Going green emerged as a viable solution to this challenge. Green computing is a study field that intends to overcome energy and climate issues. Green sustainable cloud technology aims to accomplish not only effective processing and computing environment usage, but also to improve energy efficiency. This paper will address the possibilities of sustainable cloud technology, emerging developments, and strategies to increase data center efficiency, which also will minimize carbon emissions. This creates research hurdles when such power devices are necessary to reduce the negative environmental effect of cloud technology.

Keywords Cloud computing · Energy efficiency · Efficient data centers · CO_2 emissions

1 Introduction

The emergence of cloud-based solutions, which allows including on deployment of flexible resources via a pay-as-you-go approach, has changed the communication and information technology (especially information assurance) business. Large

A. Gupta (✉) · P. Singh · D. Jain · A. K. Sharma · P. Vats · V. P. Sharma
Dr. Akhilesh Das Gupta Institute of Technology & Management, GGSIP University, New Delhi, India
e-mail: anjaniaggarwal.06@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_71

companies and public agencies have transferred their information and operation infrastructure to the cloud in recent years. As we approach the latest iteration of wireless communication networks (5G), telecom operators (access network) must satisfy the growing increase in economic capacity and low-latency applications [1]. As a result, they take use of Internet computational power and operate their networking pieces as dispersed public cloud. Many businesses' use of cloud technology has led in the construction of massive data centers throughout the world, each comprising dozens of computers and communication equipment. Because of the high demand for processing resources, the number of cloud support data centers has grown dramatically in recent years. Because more and more data centers have come into existence, the energy consumption of these data centers has grown significantly [2]. The principles of green computing are similar to raw chemistry, reduce the use of hazardous substances, increase the working capacity during the life of the product, and encourage recycling or decay of non-existent products and industrial waste. Research is underway in key areas such as making the use of computers as energy efficient as possible, as well as designing algorithms and computer technology systems related to efficiency. With conventional data centers, organizations are aware of how much energy they are using in their energy bill. In the clouds, such information is not readily available. Although data centers are much more efficient than local ones, their much-needed model for rapid upgrades and their easily accessible computer resources may mean that it is easier for organizations to use additional resources indirectly. In addition to excessive energy consumption, there is an additional impact on the environment in the guise of carbon emissions (Fig. 1).

This research provides the sustainable solution to cloud computing architecture by utilizing the resources efficiently, virtualization, and enabling energy. Green cloud computing aims to incorporate energy-saving technologies, environmentally friendly production processes as well as advanced waste disposal and recycling processes.

(A) Green Design: Designing energy-efficient servers, infrastructure, computers, software applications, and other digital devices will reduce the amount of energy used in future.
(B) Green Production: Reducing energy and waste by recycling and recycling during digital device production can have a far-reaching effect on a more sustainable environment.
(C) Green Consumption: Reducing the amount of energy a product uses while using

it can reduce up to 27% of energy.

Sustainable public cloud gained traction after the National Institute of Standards and Technology (NIST) identified a disadvantage of cloud applications. As according NIST, "the main objective of this application would be to provide common resources, but the drawback has been its high installation cost, excessive power use, and carbon emissions, as seen in Fig. 2." As public cloud grows more popular and World Wide Web-related networking volume increases, large corporations providing Internet services use greater and greater electricity for operating data centers.

Fig. 1 Cloud computing

Fig. 2 Cloud Computing and its effect on climate change

To overcome this issue, data center resources must be handled in an energy-efficient way in order to support green cloud technology. Cloud storage, in instance, must be allotted not just to meet the service quality needs provided by users, but also to decrease energy consumption.

2 Key Necessity of Sustainable Cloud-Based Solutions

Computer technology must be "green," which means we must be aware of how our lives and work affect the environment. Going green, often referred to as green computing, refers to the ecologically sustainable and environmentally sensitive use of machines and their supplies. This includes the use of energy-efficient central processing unit (CPU), workstations, and accessories and also decreased resource utilization and efficient and effective disposal of wastes (e-waste). This also includes the creation of environmentally friendly materials, energy-efficient computing, and the recycle of systems and devices. Raised power usage has raised IT operating costs [3]. Modern data centers, operating under the cloud computing model, host a variety of applications ranging from those that run for a few seconds (e.g., providing Web applications such as e-commerce and temporary social networking sites) to long-term ones (e.g., simulation or large data set processing) in shared hardware forums.

The need to manage multiple applications in a data center poses a challenge to providing much-needed service and distribution in response to dynamic workloads. In general, data center resources are assigned to applications mathematically, based on heavy load factors, in order to maintain segmentation and provide performance guarantees. Green sustainable cloud computing has emerged as a result of the growing threats to the environment and the need to minimize the energy use. Sustainable computing's primary objectives are to limit use of such poisonous and hazardous substances, promote sustainable development, and recycle e-waste. This method entails the effective installation of Webserver and other components as well as the reduction of energy usage [4]. Investigation in important areas including such developing computer being used as efficient as feasible, as well as inventing computational methods for efficiency-related information systems, continued.

3 Green Cloud Services Methods

Using "virtualization" technique is vital for energy conservation in clouds since it allows for large improvements in energy efficiency while utilizing the very same architecture. Organizations that use VMware may save a lot of money in terms of space, administration, and electricity.

(A) Dynamic Cloud Provisioning

Traditionally, computer systems and commercial infrastructures were managed to meet deadliest demands. As a result, most IT firms build more capacity than is necessary. The following are the explanations behind this situation [4]:

•It is quite hard to anticipate demand in advance.

•To ensure that service continuity and to sustain a specific degree of customer satisfaction for end customers.

(B) SaaS-Based Multi-tenancy

The use of virtualized environment decreases total energy consumption and greenhouse gas emissions. Broader organization is served by SaaS application services using the same technology and equipment. This strategy is much more environmentally friendly than installing several backup copies on various infrastructure, which could also reduce the requirement for infrastructure upgrades. The lower the volatility in consumption, the greater the forecast, and the bigger the energy management system.

(C) Server Utilization

Multiple programs can indeed be maintained and operated in solitude on the very same domain controller using VMware vSphere, resulting in utilization concentrations of up to 70 percent of the overall [5]. While also increasing server utilization leads in increased power requirements, a feedforward at high utilization can perform greater workloads with the same amount of energy.

(D) Data Center Efficiency

Cloud service providers may considerably enhance the power demand performance (key performance indicator) of existing computer servers by utilizing far more energy-efficient technology. Cloud computing provides applications to be transferred across several data centers that have lower spectral efficiency levels.

4 Energy Consumed into Large Data Centers

An information center's design is complicated since it includes not just equipment but then also application that operates in the IT infrastructures. As a result, we may divide its components into 2 layers: physical and logical, as seen in Fig. 3.

Energy is usually used to operate data centers. Contemporary data centers, on the other hand, utilize green technologies such as wind and solar power, waves, and geothermal PV as part of a drive to reduce emissions and conform with proactive environmental standards. Electricity is sent from exterior power stations to inner basic infrastructures, information and technology (telecom) gear, and many other support networks. Electricity is sent to interior IT operations using uninterruptible power supply supplies (appliances and equipment) to provide uninterruptible power delivery though in the event of a power outage [6].

Fig. 3 Categorization of Data Centers: **Hardware and Software**

5 Initiatives Development using Sustainable Cloud Technology

There really are equipment/software technical solutions which drastically reduce power requirements and environmental effect. The following are the primary approaches for reducing energy use, CO2 emissions, and development approaching greener data centers [7]:

(A) Application Oriented

The software-as-a-service (SaaS) concept has altered how applications and software are delivered and consumed. To reduce IT costs, an increasing number of businesses are migrating to SaaS clouds. As a result, this has become necessary to resolve energy consumption somewhere at application server.

(B) Virtualization

It is just an energy-efficient way for sharing material assets across virtual server machine platforms (VMs). Variable resource distribution is a technology that allows us to use as well as allocate periodically free computational power across VMs while conserving a substantial portion of power [8]. VMware is an effective method for providing server centralization and powering off idle servers. In so many circumstances, a dormant server uses 70 percent of the overall of the electricity that a server operating at full CPU utilization requires [9].

(C) Energy-efficient Hardware

Computing system is comprised of equipment, including as processors, inbuilt displays, discs, fans, and so on. This gear must utilize as little electricity as possible. Several well-known approaches can be used to preserve computer power [10]. To begin, the program should indeed be turned off if it detects a disconnect. Secondly, we may decelerate the processor clock frequency, which would be called as clock PWM gating, and then, we can shut down the chips, which would be called as microchip inhibition. Tempo shift, energy still, cool and quiet [11], and regulation switches are some other strategies for shutting down the main CPU.

(D) Monitoring/Metering

It is claimed that what you should cannot evaluate can indeed be improved. It is critical to build electricity capabilities that help the computer to understand how much power is spent by a certain component and how it could be decreased. The sustainable grid has suggested two particular measures known as power usage effectiveness (peak power) and datacenter infrastructural effectiveness (application procedures) to assess the combined efficiency of a datacenter and increase the reaction rate per watt [12].

Power Usage Effectiveness

$$= \text{Facilities Total Available Power/Provided IT based Equipmental Power} \quad (1)$$

Datacenter Infrastructure EffeTctiveness Efficiency

$$= 1/\text{Power Usage Effectiveness}$$

$$= \left[\text{Provided IT based Equip mental Power/Facilities Total Available Power}\right] \times 100\% \quad (2)$$

(E) Network Infrastructure

Energy-saving concerns in communication are sometimes alluded to as "environmental networking," which refers to the incorporation of electricity into network infrastructure, equipment, and procedures.

6 Conclusion and Future Scope

This article provides a quick overview of sustainable green cloud computing in a public cloud scenario. It aims to accomplish not only effective processing and computing infrastructure usage, but to improve energy efficiency. Power demand control in storage systems has resulted in numerous of significant advancements in energy conservation. To summarize, no technique or approach is greener, but it is the person who employs it make it efficient by employing it properly and in accordance with its specifications. Cloud service providers must reduce their energy use and take massive moves toward embracing renewable resources instead of focusing solely on reducing costs. Sustainable computing technology, on the other hand, has to be investigated.

References

1. K. Bilal, S. Malik, S. Khan, A. Zomaya, Trends and challenges in cloud datacenters. IEEE Cloud Comput. **1**(1), 10–20 (2014)
2. P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, Pratt and I. Warfield The art of virtualization. ACM IGOPS Operating Syst. Rev. **37**(5), 164–177
3. B. Saha, Green computing. Int. J. Comput. Trends Technol. (IJCTT) **14**(2) (2014)
4. A. Tahiliani, A. Purohit, A review on energy saving using green computing system. Int. J. Eng. Dev. Res. ISSN 2321-9939
5. A. Atrey, N. Jain, N. Iyengar, A study on green cloud computing. Int. J. Grid Distrib. Comput. **6**(6), 93–102
6. M. Aljaberi, S. Khan, S. Muammar, Green computing implementation factors: UAE case study. 1–4 (2016). https://doi.org/10.1109/ICEDSA.2016.7818528
7. B. Whitehead, D. Andrews, A. Shah, G. Maidment, Assessing the environmental impact of data centres—Part 1: background, energy use and metrics. Building Environ. **82**, 151–159 (2014)
8. H. Liu, C.-Z. Xu, H. Jin, J. Gong, X. Liao, Performance and energy modeling for live migration of virtual machines, in *Proceedings of 20th International Symposium HPDC* (2011), pp. 171–182
9. A. Beloglazov, J. Abawajy, R. Buyya, Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Future Gener. Comput. Syst. (2012, May)
10. A. Beloglazov, R. Buyya, Energy efficient resource management in virtualized cloud data centers, in *Proceedings of 10th IEEE/ACM International CCGrid* (2010, May), pp. 826–831
11. E. Feller, C. Rohr, D. Margery, C. Morin, Energy management in IAAS clouds: a holistic approach, in *Proceedings of IEEE 5th International Conference*s.
12. P. Corcoran, A. Andrae, Emerging trends in electricity consumption for consumer ICT (National University of Ireland Galway, Ireland, Tech. Rep., 2013)
13. C. Thanmai, K.N. Narsimaha Murthy, S. Ambareesh, A survey on evolution of green computing in the cloud environment. Int. J. Innov. Res. Comput. Commun. Eng. **3(**Special Issue 5) (2015)

14. L. Krug, M. Shackleton, F. Saffre, Understanding the environmental costs of fixed line networking, in *Proceedings of 5th International Conference on Future Energy Systems* (2014), pp. 87–95
15. S. Mingay, Green IT: The New Industry Shock Wave (2007). Available online http://www.ictliteracy.info/rf.pdf/Gartner_on_Green_IT.pdf. Accessed on 18 July 2017
16. C.T. Yang, K.C. Wang, H.Y. Cheng, C.T. Kuo, C.H. Hsu, Implementation of a green power management algorithm for virtual machines on cloud computing, in *Proceedings of the 8th International Conference on Ubiquitous Intelligence and Computing*, Banff, AB, Canada, 2–4 September 2011 (Springer, Berlin, Germany, 2011), pp. 280–294
17. X.L. Xu, G. Yang, L.J. Li, R.C. Wang, Dynamic data aggregation algorithm for data centers of green cloud computing. J. Syst. Eng. Electron. **34**, 1923–1929 (2012)
18. H. Lee, Y.S. Jeong, H.J. Jang, Performance analysis based resource allocation for green cloud computing. J. Supercomput. **69**, 1013–1026 (2014)
19. M. Azaiez, W. Chainbi, H. Chihi, A green model of cloud resources provisioning, in *Proceedings of the 4th International Conference on Cloud Computing and Services Science*, Barcelona, Spain, 3–5 April 2014 (SCITEPRESS—Science and Technology Publications, Lda., Setubal, Portugal, 2014), pp. 135–142
20. J. Kolodziej, S.U. Khan, L.Z. Wang, M. Kisiel-Dorohinicki, S.A. Madani, E. Niewiadomska-Szynkiewicz, C.Z. Xu, Security, energy, and performance-aware resource allocation mechanisms for computational grids. Future Gener. Comput. Syst. **31**, 77–92 (2014)
21. L. Xu, K. Wang, Z. Ouyang, X. Qi, An improved binary PSO-based task scheduling algorithm in green cloud computing, in *Proceedings of the 9th International Conference on Communications and Networking in China (CHINACOM)*, Maoming, China, 14–16 August 2014 (IEEE, New York, NY, USA, 2014), pp. 126–131
22. G. Kaur, S.A. Midha, Preemptive priority based job scheduling algorithm in green cloud computing, in *Proceedings of the 6th International Conference Cloud System and Big Data Engineering*, Noida, India, 14–15 January 2016 (IEEE, New York, NY, USA, 2016), pp. 152–156
23. Y. Liu, W. Shu, C. Zhang, A parallel task scheduling optimization algorithm based on clonal operator in green cloud computing. J. Commun. **11**, 185–191 (2016)
24. D. Zhang, Z. Chen, L.X. Cai, H. Zhou, J. Ren, X. Shen, Resource allocation for green cloud radio access networks powered by renewable energy, in *Proceedings of the Global Communications Conference*, Washington, DC, USA, 4–8 December 2016 (IEEE, New York, NY, USA, 2016), pp. 1–6
25. A. Khosravi, S.K. Garg, R. Buyya, Energy and carbon-efficient placement of virtual machines in distributed cloud data centers, in *Lecture Notes in Computer Science, Proceedings of the Euro-Par 2013 Parallel Processing*, Aachen, Germany, 26–30 August 2013, ed by F. Wolf, B. Mohr, D. Mey (Springer: Berlin, Germany, 2013), pp. 317–328.
26. F. Cao, M.M. Zhu, C.Q. Wu, Green cloud computing with efficient resource allocation approach, in *Green Services Engineering, Optimization, and Modeling in the Technological Age*, 1st edn. (IGI Global, Hershey, PA, USA, 2015), pp. 116–148. ISBN 9781466684478
27. Z. Deng, G. Zeng, Q. He, Y. Zhong, W. Wang, Using priced timed automaton to analyse the energy consumption in cloud computing environment. Clust. Comput. **17**, 1295–1307 (2014)
28. J. Huang, K. Wu, M. Moh, Dynamic virtual machine migration algorithms using enhanced energy consumption model for green cloud data centers, in *Proceedings of the International Conference on High Performance Computing & Simulation*, Bologna, Italy, 21–25 July 2014 (IEEE, New York, NY, USA, 2014), pp. 902–910
29. C.M. Wu, R.S. Chang, H.Y. Chan, A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters. Future Gener. Comput. Syst. **37**, 141–147 (2014). (Symmetry 2017, 9, 295 16 of 20)
30. J.A. Aroca, A.F. Anta, Empirical comparison of power-efficient virtual machine assignment algorithms. Comput. Commun. **96**, 86–98 (2016)
31. F. Farahnakian, T. Pahikkala, P. Liljeberg, J. Plosila, H. Tenhunen, Utilization prediction aware VM consolidation approach for green cloud computing, in *Proceedings of the 8th International*

Conference on Cloud Computing (CLOUD), New York, NY, USA, 27 June–2 July 2015 (IEEE, New York, NY, USA, 2015), pp. 381–388
32. B. Kaur, A. Kaur, An efficient approach for green cloud computing using genetic algorithm, in Proceedings of the 1st International Conference on Next Generation Computing Technologies, Dehradun, India, 4–5 September 2015 (IEEE: New York, NY, USA, 2015), pp. 10–15
33. G. Koutsandria, E. Skevakis, A.A. Sayegh, P. Koutsakis, Can everybody be happy in the cloud? Delay, profit and energy-efficient scheduling for cloud services. J. Parallel Distrib. Comput. **96**, 202–217 (2016)
34. Z. Long, W. Ji, Power-efficient immune clonal optimization and dynamic load balancing for low energy consumption and high efficiency in green cloud computing. J. Commun. **11**, 558–563 (2016)

Threats and Challenges of Artificial Intelligence in the Healthcare Industry

Priti Ranjan Sahoo, Smrutirekha, and Mou Chatterjee

Abstract The authors provide a crisp yet in-depth summary of the relevance of artificial intelligence (AI) in providing healthcare solutions. Artificial intelligence is a relatively new concept in the field of health care. AI aids in the prediction of disease patients for medical procedures. Patients, pharma companies, health services, insurance companies, and medical institutions benefit from AI's application in health care. Artificial intelligence supports different concepts, counting computing, computer program improvement, and information exchange. Machine learning, profound learning, normal dialect generation, discourse acknowledgment, robots, and biometric distinguishing proof are illustrations of artificial intelligence's innovation. Artificial intelligence is used in a variety of areas, including health care, manufacturing, and business. It is also used in the automotive industry. The authors have discussed the current scenario of artificial intelligence and the threats and challenges posed for AI in the healthcare industry.

Keywords Artificial intelligence · Health care · Technology · Machine learning

1 Introduction

The ability to gather, partake in, and provide data is becoming increasingly crucial as digitization eliminates all interruptions, including health care. Machine learning, big data, and artificial intelligence (AI) can all help with the issues that enormous datasets issues. Machine intellect can also assist healthcare chambers in meeting rising medical demands, improving operations, and lowering costs.

Alan Turing, a British mathematician, and philosopher questioned whether machines could think in 1950. AI was not fully conceptualized and named until the Dartmouth Summer Research Project in 1956. AI made its way into health care in the early 1970s and began to be applied to biomedical problems. Artificial intelligence in medicine, an international AI journal, was launched, and the American

P. R. Sahoo (✉) · Smrutirekha · M. Chatterjee
KIIT School of Management, KIIT University (Institution of Eminence), Bhubaneswar, India
e-mail: prsahoo@ksom.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_72

Association for artificial intelligence was founded in 1980, focusing on medical applications. In the decades that followed, AI was introduced into clinical settings. Among the new applications of AI in healthcare are fuzzy expert systems, Bayesian networks, artificial neural networks, and crossbred intelligent systems.

Geoffrey Hinton and colleagues released AlexNet in 2012. This began the current surge in interest in significant learning. Significant neural frameworks have as of late outlined multitudinous errand execution comparable to characteristic pros. Over the decade, there has been an immediate increase in research regarding AI knowledge and it's significant capability in healthcare applications. By 2017, AI inside the treatment had finished up the preeminent common application of AI in adding up to esteem help.

At last, two sorts of fake bits of knowledge utilized in health care have risen: physical and virtual. Physical AI alludes to utilize of proclivity and robots to assist cases and suppliers in giving care. Machine, or significant capability, might incorporate virtual AI in which calculations are made through replication and inclusion [8].

AI-based innovations try to emulate human cognitive processes, and robots' ability to analyze and manage massive datasets using AI algorithms has evolved rapidly. In the era of big data, one advantage of using AI in health care is that complicated patterns, and correlations can be identified algorithmically without recruiting additional healthcare specialists for data analysis.

Artificial intelligence AI refers to machines that can perform tasks that would require intelligence if performed by humans. Machine learning refers to computer learning that is not explicitly programmed and includes a variety of techniques for achieving AI. Artificial neural networks, which simulate the human brain, are an example of machine learning technology architecture. The human brain comprises billions of nerve cells or neurons that are organized in complex interconnected networks that allow us to generate complex thought patterns and actions.

2 Role of Artificial Intelligence in Healthcare Solutions

Artificial intelligence provides numerous technology solutions for doctors, patients, and management consistently for the ease of each of the entities involved in the healthcare industry.

For example, doctors use the diagnostic imaging interpretation application for making a decision. The healthcare industry is much obliged to profound learning programs, innovation categorization, and AI-based imaging frameworks are prepared with calculations to filter pictures. MRIs and CT scans are examples of this. One of the most widely used AI-based predicated judgment checkers, buoy health, employs an algorithm that aids in treating sickness. The chatbot is aware of the case's symptoms as well as the health complications. This aids in decision-making and gives a direction for further treatment.

As the well-known saying goes, "Prevention is better than Cure." The doctors' new approach focuses on preventative care based on data collection. This includes

everything from gathering in-born information to wearable devices that advance the electronic healthcare system. Fitbit, Apple, Garmin, and other companies monitor heart rate and physical activities effectively. They can send warnings to the case about exercise and medication dosage. The fashionable AI solution gathers data from mutation genetics and DNA. This facilitates the system's inspection and detection of cancer and vascular disorders in advance.

A virtual health assistant is a one-of-a-kind method for patients that can reduce frequent clinic visits and relieve the pressure on the medical staff. They also serve as a conduit between service providers and cases. Chatbots also play a significant role in serving the patients. Interactive chatbots have been created using natural language processing (NLP) and sentiment analysis [11]. They enable cases to inquire about bill payment, furniture, or pharmaceutical kitties. Cases enduring persistent ailments get it how to care for their condition. In any case, communicating with them through bots makes a difference to ease the burden on therapeutic suppliers.

Babylon, a UK-based software, uses emotional artificial intelligence to offer the best medical arguments based on a person's history of medical diagnosis. Either way, the job suggests a course of action. Another AI software, sensibly, aids healthcare workers in covering cases' circumstances and providing remedies as a result.

For the manufacturers and service providers, AI aids in maintaining cybersecurity, in the manufacturing of drugs, and detection and checking of any fraudulence in the system.

To consider the practical implications of virtual nursing assistants, in the United States, 89 percent of patient googles their symptoms before consulting a doctor, and the consequences of self-diagnosis are often frightening. Molly is an excellent example of a multi-functional virtual nurse. It tracks a patient's weight, blood pressure, and other indicators generated by monitoring equipment and provides remote support for common medical issues. Patients can also use the app's chatbot to confidentially discuss their health concerns and make an appointment with their doctor. Molly and other virtual assistants have already gained much traction, with 64 percent of patients preferring to receive instructions from them [19].

Machine learning applications can make strides in the precision of treatment conventions and well-being results through algorithmic strategies. Profound learning, for illustration, could be a sort of complex machine learning that re-enacts how the human brain capacitate [17]. Profound learning programs can identify, recognize, and assess dangerous tumors from photos by utilizing neural systems that learn from information without supervision [13].

Machine learning systems can spot irregularities in photographs that the human eye cannot because of quicker processing speeds and cloud infrastructures, assisting in disease detection and treatment.

Breakthroughs in machine learning in the medicine-related industry will continue to shape business strategies in the future. For instance, as studied by [5], machine learning applications in advancement incorporate a diabetic retinopathy symptomatic apparatus and prescient analytics to distinguish breast cancer repeat based on therapeutic information and pictures.

Artificial intelligence helps with healthcare technology record keeping, particularly electronic health records (EHRs). AI-assisted EHR management can improve the quality of health care while reducing health care and administrative costs and streamlining operations.

Normal dialect handling, for illustration, permits experts to require and record clinical notes without having to depend on manual strategies. By advertising clinical choice help, computerizing imaging procedures, and binding together innovations, machine learning methods can simplify EHR organization frameworks for doctors. Due to insufficiencies in healthcare information, machine learning calculations might create false expectations, which can critically affect clinical decision-making [16].

Because healthcare data were created for electronic health records (EHRs), it must be prepared well in advance for the machine learning algorithms to use it effectively. Health informatics experts are in charge of data integrity. Health informatics specialists are in charge of gathering, analyzing, classifying, and cleansing data. Machine learning, medical technology, and predictive analytics are all tools that can be used to improve healthcare procedures, change clinical decision support systems, and improve health satisfaction [10]. The capacity of machine learning in health care to utilize well-being informatics to anticipate well-being results through prescient analytics comes about in more exact determination and treatment and moved forward clinician bits of knowledge for custom-made and cohort medicines.

Machine learning can supplement prescient analytics by deciphering information for decision makers to distinguish handle imperfections and upgrade generally healthcare operations [1].

Machine learning has the potential to have a positive impact on patient care delivery strategies. It can, for example, assist professionals in identifying, diagnosing, and treating disease. Machine learning applications in health care can also help to streamline tasks and improve operation planning, preparation, and execution [3, 8]. By examining enormous amounts of healthcare records and other patient data, machine learning algorithms can find patterns related to diseases and health conditions in one go. This reduces the operational cost and processing time.

Machine learning technologies have the potential to increase access to health care in developing countries while also improving cancer detection and therapy [2].

Machine learning has been shown to help clinical practitioners improve their productivity and accurateness. Machine learning in medical imaging is commonly used to detect cardiovascular irregularities, diagnose musculoskeletal injuries, and screen for cancer [6].

By combining real-time information, data from earlier fruitful surgeries and therapeutic records, machine learning can move forward the accuracy of automated surgical gear. Decreased human mistakes, back amid more complex medicines, and negligibly intrusive surgeries are fair a number of the benefits. Robots, for instance, could help in spine surgery by performing precise surgeries to unblock blood vessels. Surgical robots can provide surgeons with more than just automated assistance by helping them organize surgical operations and procedures [14].

Patients' abilities can be supplemented directly by robots. Assisting paraplegic patients to regain their ability to walk and execute tasks such as blood pressure

monitoring and medication reminders are just a few examples. Robot friendship can benefit even the sick and aged [9].

The quality of data input determines the output of machine learning algorithms. Erroneous data might jeopardize system reliability, raising questions about whether judgments based on the data are correct or incorrect. Flawed data can also contribute to a lack of cultural competency. In the end, issues regarding the reliability of the system and the complications due to misinformation used by machine learning algorithms could result in erroneous outputs, misinformed medical decision-making, and eventually harm people's health.

3 Threats and Challenges Faced by Artificial Intelligence in the Healthcare Industry

1. To train and optimize the performance of the algorithms, a substantial quantity of patient information is required for an AI solution to be successful. Obtaining access to these datasets in health care raises several issues. Case loneliness and data control ethics—entering idiomatic medical records is strictly prohibited. Data sharing between hospitals and AI companies have spawned competitors in recent generations, highlighting several ethical concerns.
2. Developing standards for a technology that is shadow based and constantly changing poses significant difficulties. What methods are available for securing cases? How can one ensure that a constantly learning and evolving result—rather than a unique, reading-controlled medical equipment—has adequate despotic oversight? Whether AI is a guru of treatment rather than just a device arises when AI outcomes entail direct patient relations without doctor monitoring. Will this sample expand to a desire for some medical license to operate—and would a public medical board agree to provide such a license? This raises the question of who is responsible if something goes wrong. Is the AI firm, however, responsible for the case's outcome? Will insurance firms ever invest in an AI tool, as an example?
3. Suppose this technology has control over opinion or treatment. Another barrier to utilization is the departure of a doper. With these tools, the human touch of interacting with a doctor can be lost. Are cases willing to rely on a decision made by a software algorithm rather than a human? Furthermore, in the meantime, are physicians prepared to accept these research discoveries? It may be ludicrous to expect relatively cold departing rates beyond the testimonial of sweeping generalization studies in conscientiousness that still universally uses the device.
4. Despite the revolution in incorporating AI in providing healthcare solutions, few threats as given in Fig. 1 have been found that need attention.

A risk highlighted by AI systems is that they demand enormous amounts of data from various sources, including pharmaceutical, electronic health, and insurance claims [7]. Because data are broken and cases changing insurance companies or

Fig. 1 Threats to AI in health care

Fig. 2 Forecast for the future of the industrial implications

providers daily, the data become more convoluted and less coherent, increasing the risk of inaccuracy and the expense of data gathering—data availability.

- Within the long run, the work of AI frameworks could lead to shifts within the therapeutic calling especially in ranges like radiology, where the farthest of the work gets computerized. This raises the concern that a tall degree of business of AI might lead to a drop in standard information and capacity over the dates, making suppliers fail in identifying AI wrongdoings and within the other advancement of vital information—switching of profession.

- Collecting massive datasets and exchanging information between healthcare systems and AI engineers to strengthen AI systems lead to combat cases acknowledging that this may abuse their confines, resulting in a record of claims. Another area where AI frameworks' work poses this concern is that AI can predict private data surrounding situations even when persistent has never provided this info—privacy issues.
- AI frameworks can assimilate the inclinations of the available knowledge because they assimilate and learn via the prepared material, for example, suppose the data gathered in AI is collected chiefly in scholarly therapeutic centers. In that case, the developing AI frameworks will have less information about, and as a result, will treat cases from populations that do not routinely visit scholarly therapeutic centers less successfully—biases and inequality.

For at least two reasons, AI errors are potentially distinct. While human healthcare professionals can commit mistakes at times, what makes this important is that an underlying fault, such as one in an AI system, might result in thousands of patients being injured in one single process cycle—errors and hazards [4].

4 Scope and Future Implications of Artificial Intelligence in Healthcare Solutions

Massive healthcare delivery organizations have tested and are testing AI on a broad scale, and the beauty of running an independent medical practice is that one gets to observe these giant providers test out technology before it becomes universal. Smart practices, on the other hand, are better pursuit to wait for the larger, more expensive AI applications. Keep a watch on large-scale suppliers as they increasingly include AI-powered technologies and learn from their experiences to determine which applications will be most helpful to one in the future. Figure 2 gives an overview of the forecast for the future of the industrial implications.

Digital healthcare professionals are on the cutting edge of this opportunity, assisting in integrating learning algorithms into medical and healthcare systems. In a profession that is becoming increasingly data driven, their in-depth understanding of technology and how it can improve the treatment protocol and the results are invaluable [15]. Machine learning's deep learning algorithms can reduce the time it takes to analyze a patient's medical history and the recurring disease details resulting in faster diagnosis, treatment, and recovery. The most important work of well-being informatics experts is ensuring the precision of healthcare data that becomes progressively crucial as healthcare organizations endeavor to coordinate machine learning into healthcare and restorative operations. A run of technology-driven healthcare activities will appear as a guarantee within the future in terms of making strides in care conveyance. Health informatics experts' jobs will be altered as a result of these improvements. Some of the artificial intelligence technologies that would affect and create opportunities in the healthcare industry are discussed below.

- Virtual Reality: Virtual reality (VR) alters health care by making it easier to train doctors and transform patients' lives. Surgeons wearing specific VR headsets, for example, can Webcast surgeries and provide medical trainees a unique perspective on a surgical process. In another case, virtual reality is being employed to aid in physical therapy recuperation. Physical therapy patients are frequently subjected to sensitive physical activities that might be taxing. Recovery programs can be tailored, and physical treatment activities can be made more pleasurable and engaging through VR training exercises with machine learning.
- Genome Sequencing: The physician can start creating personalized treatment plans for their patients using genomic data. Machine learning in information systems allows genetic mutations to be analyzed more quickly and aids in diagnosing conditions that can lead to disease. Deep learning applications that enable genome sequencing can improve cancer care and prognosis while minimizing the impact of infectious diseases. As genetic sequencing becomes more reasonably priced and machine learning becomes way more intelligent, healthcare technology professionals can help advance genomic medicine to treat the world's deadliest diseases.
- Augmented Reality: Augmented reality (AR) is one of the top three technologies revolutionizing health care. AR technologies in health care, like VR, can help medical students prepare better. Students can learn firsthand from surgeons doing real-life surgery using augmented reality technologies. Medical students, for example, may acquire precise, accurate renderings of human anatomy without having to study actual human beings thanks to augmented reality.
- Nanotechnology: According to the National Nanotechnology Initiative, nanotechnology is described as "the understanding and control of matter at the nanoscale, at dimensions between about 1 and 100 nm." Nano-medicine refers to the use of nanotechnology in health care. Nanotechnology can aid in executing activities involving chemicals, biological structures, and DNA, such as medicine delivery. Medications can be sent to specific sections of the human body that are not impacted by diseases.
- 3D Printing: This is an innovation that offers an array of opportunities in the healthcare sector. These techniques can benefit drug formulations, implants, prostheses, biosensor devices, and human tissues and organs. It enables the personalization of medical treatments, improves healthcare quality, lowers costs, and lowers production risks.
- Wearable Watch: Various consumer wearable devices help people become more fit, from counting steps to monitoring heart rhythms. Other wearable devices, such as heart rate, blood pressure, temperature, and heart rate, can give clinicians information regarding a patient's health. According to Pew Research Center, wearable devices, such as activity trackers and smartwatches, are used by roughly 21% of Americans. As more people adopt wearable technology, health informatics specialists can improve data communication and accuracy between these devices and the health information systems used by clinicians.

5 Conclusion

The construction of an ethical global governance structure and system and exceptional standards for frontier AI uses in medicine are needed to assure "trustworthy" AI applications in health care and medicine.

AI tools are making their way into the medical field. It is now a fact that we must face in order of their arrival. While there appears to be some hyperbole in the discussion of AI in health care, we also discovered that diverse considerations and knowledge have emerged among each stakeholder category. On the one hand, the health sector and researchers put a premium on high-quality health data; on the other side, doctors are still waiting for proof of these tools' utility, and they are concerned about being found liable in the occurrence of an injury caused by such an artificial intelligence system they do not fully comprehend [12].

Crossbred models, in which practitioners are aided in determination, treatment planning, and correlating peril variables, but retain final accountability for the case's care, are the best chances for AI in health care over the following countless areas. This will result in healthcare providers leaving faster due to the reduced perceived risk and bringing measurable advances in patient concerns and efficiency at scale.

Massive challenges continue as a result of the magnitude and complexity of health data. However, by creating new example recognition systems, adaptive computations, and unique methodology that leverage vast volumes of health data to answer broad questions, the community dwelling in the lap of artificial intelligence is well on its way to surpassing these obstacles [18].

Counterfeit insights are a science with its implementations in various fields, including the medical and healthcare system. The development of untrue mindfulness has decreased human intercessions and, as a result, has driven to fundamental and fast, the common-sense discovery of different complicated afflictions. Artificial intelligence is valuable and worth incorporating for planning one's daily life. The fundamental component of artificial intelligence in long-term care is calm finding and image examination; however, the long run holds incredible potential for applying AI to improve various aspects of long-term care management.

References

1. G.V.K.S. Abhinav, S. Naga Subrahmanyam, Artificial intelligence in healthcare. J. Drug Del. Ther. **9**(5-s), 164–66 (2019)
2. M.A. Ahmad, A. Patel, C. Eckert, V. Kumar, A. Teredesai, Fairness in machine learning for healthcare, in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2020), pp. 3529–3530
3. R. Alugubelli, Exploratory study of artificial intelligence in healthcare. Int. J. Innov. Eng. Res. Technol. **3**(1), 1–10 (2016)
4. J. Archana, E.A. Mary Anita, A survey of big data analytics in healthcare and government. Procedia Comput. Sci. **50**, 408–413 (2015)

5. J. Bajwa, U. Munir, A. Nori, B. Williams, Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthcare J **8**(2), e188 (2021)
6. A. Callahan, N.H. Shah, Machine learning in healthcare, in *Key Advances in Clinical Informatics* (Elsevier , 2017), pp. 279–91
7. P. Chowriappa, S. Dua, Y. Todorov, Introduction to machine learning in healthcare informatics, in *Machine Learning in Healthcare Informatics* (Springer , Berlin, 2014), pp. 1–23
8. S. Ellahham, N. Ellahham, Use of artificial intelligence for improving patient flow and healthcare delivery. J. Comput. Sci. Syst. Biol. **12**(3) (2019)
9. S.M. Kulkarni, G. Sundari, A review on image segmentation for brain tumor detection, in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (IEEE, 2018)s, pp. 552–55
10. H. Liyanage, S.-T. Liaw, J. Jonnagaddala, R. Schreiber, C. Kuziemsky, A.L. Terry, S. de Lusignan, Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearb. Med. Inform. **28**(1), 41–46 (2019)
11. M. Marwan, A. Kartit, H. Ouahmane, Security enhancement in healthcare cloud using machine learning. Procedia Comput. Sci. **127**, 388–397 (2018)
12. N. Mehta, A. Pandit, S. Shukla, Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study. J. Biomed. Informatics **100**, 103311 (2019)
13. R. Miotto, F. Wang, S. Wang, X. Jiang, J.T. Dudley, Deep learning for healthcare: review, opportunities, and challenges. Brief. Bioinform. **19**(6), 1236–1246 (2018)
14. A. Panesar, *Machine Learning and AI for Healthcare.* (Springer, Berlin, 2019)
15. P. Sanjay, C. Meng, Z. Che, Y. Liu, Benchmark of Deep Learning Models on Large Healthcare Mimic Datasets (2017). ArXiv Preprint ArXiv:1710.08531
16. L. Shinners, C. Aggar, S. Grace, S. Smith, Exploring healthcare professionals' understanding and experiences of artificial intelligence technology use in the delivery of healthcare: an integrative review. Health Inf. J. **26**(2), 1225–1236 (2020)
17. S. Siddique, J.C.L. Chow, Machine learning in healthcare communication. Encyclopedia **1**(1), 220–239 (2021)
18. T.Q. Sun, R. Medaglia, Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Gov. Inf. Q. **36**(2), 368–383 (2019)
19. J. Waring, C. Lindvall, R. Umeton, Automated machine learning: review of the state-of-the-art and opportunities for healthcare. Artif. Intell. Med. **104**, 101822 (2020)

Exploiting Tacit Knowledge: A Review and Possible Research Directions

Pawankumar Saini and Pradnya Chitrao

Abstract In the corporate world, tacit knowledge is becoming recognised as a differentiator. The value and necessity of focusing on leveraging tacit knowledge are highlighted via a literature study. The important themes covered in this review are the knowledge generation process, information sharing, behaviours that influence tacit knowledge sharing, sharing strategies, and so on. The research examines the different factors that lead to the industrial exploitation of tacit knowledge. To address the research gaps identified, the authors attempt to propose possibilities for future research in this area of discourse. The authors feel that the provided issues for future research will aid academics in further exploring the potential of tacit knowledge exploitation in the sectors, in addition to the research work done by many scholars in the tacit knowledge domain. The authors are conscious that this article has certain limitations because it focuses on tacit knowledge in the industry, although the impact of tacit knowledge in other sectors needs to be investigated as well. This study could lead to new approaches for businesses to tap into the potential of tacit knowledge held by their staff through knowledge production and sharing.

Keywords Knowledge management · Tacit knowledge · Explicit knowledge · Knowledge sharing · Knowledge creation

1 Introduction

The markets are in a volatile state and everything, including the competition, is shifting [1]. Companies that continually develop new information, disseminate it extensively within the firm, and promptly incorporate it into new technologies and

P. Saini (✉)
Symbiosis International University, Pune, Maharashtra, India
e-mail: pawankumarsaini09@gmail.com

P. Chitrao
Symbiosis Institute of Management Studies, A Constituent of Symbiosis International University, Pune, Maharashtra, India
e-mail: pradnyac@sims.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_73

products are successful. If companies want to improve the efficiency of their resource (input) utilisation and the productivity of their manufacturing processes, they must acquire knowledge [2]. The knowledge-based perspective of the firm sees a company as a knowledge-creating entity, arguing that knowledge, and the ability to develop and use it, is the most essential source of a company's long-term competitive advantage [3]. Employee participation, both willing and active, will determine the effectiveness of knowledge management methods [4]. A key element that we discuss in this review is tacit knowledge (TK) management which starts with creating TK and then sharing it for further use. Acquisition, management, and reuse of TK are more difficult than explicit knowledge (EK). While a lot of research has been done around EK areas, there is very little attention paid to TK [5]. This research will help people from the industry to apply the findings of the research.

2 Literature Review

TK, according to Michael Polanyi, is knowledge that cannot be stated or verbalised. He went on to say that one of the most important aspects of TK is that "we know more than we can tell" [6]. There is the issue of difficulty in articulating what we know, but according to Polanyi, we may not even be aware of what we know or how our TK relates to what we can demonstrate. Because TK is such a personal knowledge, it is only shared when individuals meet face to face and take action. TK is converted to EK, which is then combined with experiences to revert to tacit. To develop new knowledge, there must be a constant upward spiral of information [6]. Nonaka stated that TK is difficult to codify, making it difficult to communicate with others [7]. Because of our deeply formed mental models, beliefs, and viewpoints, we take TK for granted. As a result, articulating TK is challenging [7]. TK can also be defined as a skill acquired by actions and experience, both formally and informally. The main challenges with TK diffusion, according to Haldin-Herrgard T, are the TK's unconsciousness and difficulty in utilising it [8]. As non-codified and disembodied know-how, Howells claims that TK can be obtained by informal take-up of learned behaviour and procedures [9]. Learnings in an unstructured or semi-structured manner characterise TK acquisition and transfer. According to Jones and Leonard, we must address various factors of transforming TK to organisational knowledge in order for many more people to gain and use the knowledge [10]. According to Paolino, Lizcano, Lopez, and Lloret, TK expands with use and so is limitless [11]. According to Mohajan, TK is a dynamic process that is heavily influenced by an individual's social ties and characteristics [12].

3 Objectives of the Study

The purpose of this study is to highlight the various aspects that are significant when it comes to exploiting TK for organisational economic growth. The goal of the research is to discuss the features of TK, TK in the context of organisations,

knowledge creation, sharing, and the challenges of managing TK in the workplace. The study's goal is to look at the many aspects of TK management in companies and to identify the practices that are currently being used to reap the benefits of TK. If organisations can fully utilise the potential of TK, they will benefit. We hope this article will help to open the possibilities of better TK management, its use, and to further the discourse of TK in industries.

4 Methodology of the Study

This is a review article and is prepared based on secondary data. A review of existing studies brings out the work that has so far been done and the gaps that remain in the literature on the role of TK in industries. This is followed by questions for future research.

5 Tacit Knowledge

There are two sorts of knowledge: EK and TK [3]. EK is a type of knowledge that includes facts, propositions, symbols, and so on [13]. TK refers to knowledge that is difficult to formalise, such as belief, viewpoint, mental models, concepts, and ideals [3]. While TK and EK are distinct, they are at opposite ends of a knowledge continuum and should not be viewed as antagonistic; rather, it is vital to recognise that we embrace them in various ways [14]. The shift on the spectrum happens based on the changes of the individual's ability to express and formulate knowledge tacitness [15]. Even if TK is embedded in the individual, EK, according to Polanyi, must be grasped tacitly. Polanyi concludes that there is no pure EK and that the tacit element is ingrained in all forms of knowledge [16]. When learning based on knowledge transfer takes place, both TK and EK are utilised. In this context, they are inextricably linked, and this is required to achieve specified learning outcomes [17]. There is a dynamic interaction with one another that allows for the creation of new knowledge and the development of creative individual and communal activity outcomes [18].

6 Knowledge Creation Process

Considering knowledge as a resource, it becomes important that organisations make optimum use of it. They must raise awareness of the process of knowledge generation and the importance of continual TK sharing among individuals, as well as assist in nurturing, articulating, and amplifying it [7]. Nonaka and Takeuchi proposed the SECI (socialisation, externalisation, combination, internalisation) four-step conver-

sion model, which consists of a continuous process of knowledge creation spiral [1]. Socialisation (from TK to TK), externalisation (from TK to EK), combination (from EK to EK), and internalisation (from EK to TK) are the four modes of conversion between TK and EK [3]. At each level of the conversion, different stakeholders are involved. The organisation is responsible for developing knowledge visions, defining objectives, facilitating dialogue at the appropriate time and place, and ensuring a conducive environment for the SECI model's effective and efficient conversions [19].

7 Knowledge Sharing

Knowledge sharing is a critical link that allows new knowledge, products, and services to be created. This fact is only now beginning to be grasped by the organisation's people [14]. Other knowledge management (KM) processes and practises are brought to life by knowledge sharing [20]. When the agents share a same background and experience, mutual relationships make it easier for knowledge transfer to occur [21]. Organisations must create an environment that encourages social connections so that knowledge, experiences, and abilities may be shared across all departments [22]. Given the choice between TK and EK transfers, firms prefer TK transfers as it is difficult to replicate [23]. Knowledge is the engine that drives organisation's performance. Every organisation is trying to grow knowledge or acquire it if needed, and then work on transferring it [24, 25]. Firms that focus on knowledge transfer have a better possibility of developing more knowledge and, as a result, differentiators. These businesses are more likely to expand and become productive [26, 27].

7.1 Methods of Knowledge Sharing

TK could provide a vast array of chances and potentials for discovery and innovation. Regardless of whether it is TK or EK, firms are putting a renewed emphasis on knowledge exchange. As a result, businesses are devising new techniques to enhance knowledge sharing [14]. Some of the ways that TK can be or is being shared in organisations has been shared by researchers. Having co-working spaces can promote inter-domain learning by facilitating TK exchange, synthesising, and sharing of domain-related ideas [28]. Using metaphors and analogies can help individuals externalise their TK and convey it more easily, making it easier for the TK user [2, 29]. Having knowledge containers in place, which is the capture of knowledge in the documented form of values, rules, and procedures can help in knowledge sharing [30]. Knowledge sharing could be ingrained in everyday work routines and informal gatherings [20]. Sharing personal experiences with decisions made, difficulties addressed, and so on can help others to use the same [31, 32]. E-Mentoring (sharing TK through social media) is a great tool to pass on the tacit knowledge [33]. Putting

systems in place that allow for face-to-face contacts, like as coaching and mentoring, can considerably aid TK transfer [34]. The complexity of coding and recording is eliminated by bringing individuals together to share their experiences, narratives, stories, and observations [35]. New techniques are shared through demonstrating them in pilot training courses. This aids in the conversion of TK to EK [36]. Learning by emulating the activities of professionals is the simplest, most cost-efficient, and most effective method of passing on TK to others [36–38]. Using the right questioning technique to extract tacit knowledge and then codifying it for sharing with others is of great help [34, 39–42].The method for conveying knowledge will be determined by the setting, audience, and learning domain, among other factors.

8 Tacit Knowledge in Industries

Because EK is knowledge that everyone knows and is public by nature, it is TK that distinguishes the organisation and offers it a competitive advantage [43]. Employee TK is used to drive service organisations, and TK becomes more strategic in terms of delivering performance [44]. Workers learn a lot of TK from their mentors and experienced employees by observing, mimicking, and practising, rather than through procedural manuals. This also aids in the preservation of knowledge within the organisation [1, 45, 46]. The mentor or the experienced individual plays a very important role in transferring the tacit knowledge to the newcomer and guiding him or her [47]. However, little is done to capture workers' TK, and there is a significant loss of skills when employees quit or retire [48]. It becomes vital to have a solid structure in place to capture organisational memory via KM systems, and a dedicated manager is required for this. The manager has operational experience as well as technical knowledge in managing the KM system [49]. Because TK is unevenly distributed, it is critical to set the stage and align collective action, all while looking for methods to improve the quality of TK on all levels [50].

9 Challenges in Managing Tacit Knowledge

The most difficult aspect of managing TK is keeping people who have acquired it because the TK is lost when they leave the company. The loss is further multiplied when the competitive edge is lost due to these individuals joining competing organisations. Previous study has shed light on the various problems that must be overcome. Organisations generally are not able to determine the sort of information leaving the organisation. Organisations are unable to appraise the importance and value of knowledge loss, and hence overlook the seriousness of the situation [51]. The increased use of online transactions is isolating workers who can only work with what they already know and are unable to learn new TK. It becomes challenging to share knowledge across many organisational units due to this [33]. When a large

amount of TK is concentrated in the hands of a few individuals, the organisation becomes vulnerable [30]. Depending on people-based approaches alone will always pose a risk of TK loss [30]. When information is lost, it causes challenges such as duplication of effort, the need to find new qualified personnel, and the risk of not regaining all the lost knowledge [48]. When an individual departs the company, critical relationships created with external partner networks are damaged or lost [51]. When top executives leave an organisation, they take with them their working methods and knowledge of the procedures they oversaw. As a result of their lack of basic working expertise, the subordinates are susceptible [32, 48]. Because TK can only be shared through example, the number of possibilities to share it is limited [52]. Organisations are still not geared up to have a proper knowledge transfer and capture mechanism [15]. Organisations have focussed solely on EK, while TK management has gone almost unnoticed [53, 54]. There is a lack of knowledge sharing due to a lack of trust, social networking, self-awareness, and training [55].

10 Research Directions

Future study can be focussed on some of the questions raised as a result of this literature review. Is it true that organisations lose money due to a lack of TK capture? Is it tough to draw top management's attention to TK because of its difficult to express nature? How high on an organisation's strategic objective is TK capture? To what extent does an organisation's TK be captured? When new experienced people join a business, to what extent does TK flow into the organisation? When an employee quits an organisation, how much of the tacit knowledge is lost? To what extent can TK be passed on from one person to the next? To what extent can TK to EK conversion be quantified? To what extent does TK transfer differ between online and in-person working systems? To what extent does tacit knowledge transfer get hampered when units are located geographically apart? To what extent can organisations become vulnerable because of individual TK accumulation? If individuals leave the firm, how will this affect business with external partners? What impact does the departure of a senior leader have on the organisation?

11 Conclusions

One of the key duties of senior management in an organisation, according to researchers, is to create an environment in which people are motivated to share their tacit knowledge and convert it to explicit knowledge for the business's growth. Organisations are ignorant to the financial loss that results from a lack of proper TK management. There are not enough metrics for calculating the amount of TK and EK produced. For the institutionalisation of the knowledge generation process, organisations must create an organisational structure with clear duties. In organisations,

there is definitely a disparity in the level of information exchange. There are a lot of strategies for knowledge sharing that have been compiled that the organisation might apply. The literature review shows that businesses are still far from obtaining an ideal level of knowledge management. To complete the TK and EK KM realm, there is still a lot of study and implementation to be done, as well as a lot of questions to be asked and addressed.

References

1. I. Nonaka, H. Takeuchi, *Knowledge-Creating Company* (Knowledge-Creating Company, 1991, 1995)
2. J.G. Woods, From decreasing to increasing returns: the role of Tacit knowledge capital in firm production and industrial growth. J. Knowl. Econ. **10**(4), 1482–1496 (2019). https://doi.org/10.1007/s13132-016-0351-2
3. I. Nonaka, R. Toyama, A. Nagata, *A Firm as a Knowledge-creating Entity: A New Perspective on the Theory of the Firm* (Japan, 2000)
4. D. Chawla, H. Joshi, Knowledge management practices in Indian industries—a comparative study. J. Knowl. Manag. **14**(5), 708–725 (2010). https://doi.org/10.1108/13673271011074854
5. J. Hao, Q. Zhao, Y. Yan, G. Wang, A review of tacit knowledge: current situation and the direction to go
6. M. Polanyi, The tacit dimension, knowledge in organizations, ed by L. Prusak (1966)
7. I. Nonaka, *The Knowledge-Creating Firm* (Harvard Business Review, 1991)
8. T. Haldin-Herrgard, Difficulties in diffusion of tacit knowledge in organizations. J. Intell. Capital **1**(4.), 357–365 (2000). https://doi.org/10.1108/14691930010359252
9. J. Howells, Tacit knowledge, innovation and technology transfer. Technol. Anal. Strate. Manag. **8**(2), 91–106 (1996). https://doi.org/10.1080/09537329608524237
10. K. Jones, L.N.K. Leonard, *From Tacit Knowledge to Organizational Knowledge for Successful KM* (2009), pp. 27–39. https://doi.org/10.1007/978-1-4419-0011-1_3
11. L. Paolino, D. Lizcano, G. López, and J. Lloret, A multiagent system prototype of a tacit knowledge management model to reduce labor incident resolution times. Appl. Sci. (Switzerland) **9**(24) (2019). https://doi.org/10.3390/app9245448
12. H.K. Mohajan, Sharing of tacit knowledge in organizations: a review (2016). [Online]. Available http://www.openscienceonline.com/journal/ajcse
13. M. Polanyi, Tacit knowing: its bearing on some problems of philosophy. Rev. Mod. Phys. **34**(4) (1962). https://doi.org/10.1103/RevModPhys.34.601
14. M.A. Fauzi, N. Paiman, A critical review of knowledge sharing in various industries and organizations. Int. J. Sci. Technol. Res. **8** (2020) [Online]. Available www.ijstr.org
15. C.H. Chuang, S.E. Jackson, Y. Jiang, Can knowledge-intensive teamwork be managed? Examining the roles of HRM systems, leadership, and tacit knowledge. J. Manag. **42**(2), 524–554 (2016). https://doi.org/10.1177/0149206313478189
16. D. Lesjak, S. Natek, Knowledge management systems and tacit knowledge. Int. J. Innov. Learn. **29**(2), 166 (2021). https://doi.org/10.1504/ijil.2021.10034239
17. J. Roberts, From know-how to show-how? Questioning the role of information and communication technologies in knowledge transfer. Technol. Anal. Strat. Manag. **12**(4), 429–443 (2000). https://doi.org/10.1080/713698499
18. I. Nonaka, G. von Krogh, Tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory. Organ. Sci. **20**(3), 635–652 (2009). https://doi.org/10.1287/orsc.1080.0412
19. I. Nonaka, R. Toyama, T. Hirata, Managing flow a process theory of the knowledge-based firm-managing flow. https://doi.org/10.1057/9780230583702preview

20. A. Abdelwhab Ali, D.D.D. Panneer Selvam, L. Paris, A. Gunasekaran, Key factors influencing knowledge sharing practices and its relationship with organizational performance within the oil and gas industry. J. Knowl Manag. **23**(9), 1806–1837 (2019). https://doi.org/10.1108/JKM-06-2018-0394
21. D.J. Teece, The market for know-how and the efficient international transfer of technology. Ann. Am. Acad. Polit. Soc. Sci. **458**(1), 81–96 (1981). https://doi.org/10.1177/000271628145800107
22. H.F. Lin, Knowledge sharing and firm innovation capability: an empirical study. Int. J. Manpower **28**(3–4), 315–332 (2007). https://doi.org/10.1108/01437720710755272
23. D.R. Williams, Knowledge transfers in the US biopharmaceutical market during a time of transition. J. Pharma. Innov. **15**(3), 445–454 (2020). https://doi.org/10.1007/s12247-019-09395-3
24. R. Teigland, M. Wasko, Knowledge transfer in MNCs: examining how intrinsic motivations and knowledge sourcing impact individual centrality and performance. J. Int. Manag. **15**(1), 15–31 (2009). https://doi.org/10.1016/j.intman.2008.02.001
25. R. van Wijk, J.J.P. Jansen, M.A. Lyles, Inter-and intra-organizational knowledge transfer: a meta-analytic review and assessment of its antecedents and consequences (2008)
26. T. Kostova, Transnational transfer of strategic organizational practices: a contextual perspective (1999)
27. A.C. Inkpen, E.W.K Tsang Wayne, Knowledge transfer (2005)
28. R. Bouncken, M.M. Aslam, Understanding knowledge exchange processes among diverse users of coworking-spaces. J. Know. Manag. **23**(10), 2067–2085 (2019). https://doi.org/10.1108/JKM-05-2018-0316
29. P. Busch, *Tacit Knowledge in Organizational Learning* (IGI Pub, 2008)
30. P. Wethyavivorn, W. Teerajetgul, Tacit knowledge capture in Thai design and consulting firms. J. Constr. Developing Countries **25**(1), 45–62 (2020). https://doi.org/10.21315/jcdc2020.25.1.3
31. B. Saeed, A. Mahmood, and A. Saeed, Tacit knowledge sharing in technology-based firms: role of organization citizenship behavior and perceived value of knowledge. Int. J. Sci. Technol. Res. (2020) [Online]. Available www.ijstr.org
32. P. Dogra, A.E. Sparkling, Supervisors' reliance on tacit knowledge and impediments to knowledge sharing in trades
33. Z. Chen, D. Vogel, T. Yang, J. Deng, The effect of social media-enabled mentoring on online tacit knowledge acquisition within sustainable organizations: a moderated mediation model. Sustainability (Switzerland) **12**(2) (2020). https://doi.org/10.3390/su12020616
34. K. Plangger, M. Montecchi, I. Danatzis, M. Etter, J. Clement, Strategic enablement investments: exploring differences in human and technological knowledge transfers to supply chain partners. Ind. Mark. Manag. **91**, 187–195 (2020). https://doi.org/10.1016/j.indmarman.2020.09.001
35. A.C. Inkpen, A. Dinur, Knowledge management processes and international joint ventures. Organ. Sci. **9**(4), 454–468 (1998). https://doi.org/10.1287/orsc.9.4.454
36. D. Asher, M. Popper, Tacit knowledge as a multilayer phenomenon: the 'onion' model. Learning Organ. **26**(3), 264–275 (2019). https://doi.org/10.1108/TLO-06-2018-0105
37. A.S. Reber, An evolutionary context for the cognitive unconscious. Philosophical Psychol. **5**(1), 33–51 (1992). https://doi.org/10.1080/09515089208573042
38. G. Csibra and G. Gergely, Natural pedagogy as evolutionary adaptation. Philosophical Trans. Roy. Soc. B: Biolog. Sci. **366**(1567). 1149–1157 (2011). (Royal Society). https://doi.org/10.1098/rstb.2010.0319
39. M. Bhardwaj, J. Monin, Tacit to explicit: an interplay shaping organization knowledge. J. Knowl. Manag. **10**(3), 72–85 (2006). https://doi.org/10.1108/13673270610670867
40. T. Gavrilova, T. Andreeva, Knowledge elicitation techniques in a knowledge management context. J. Knowl. Manag. **16**(4), 523–537 (2012). https://doi.org/10.1108/13673271211246112
41. M. Sh, Al-Qdah, J. Salim, A conceptual framework for managing tacit knowledge through ICT perspective. Procedia Technol. **11**, 1188–1194 (2013). https://doi.org/10.1016/j.protcy.2013.12.312

42. M. Supanitchaisiri, O. Natakuatoong, S. Sinthupinyo, The innovative model for extracting tacit knowledge in organisations (2020)
43. R. Seidler-de Alwis and E. Hartmann, The use of tacit knowledge within innovative companies: knowledge management in innovative enterprises. J. Knowl. Manag. **12**(1), 133–147 (2008). https://doi.org/10.1108/13673270810852449
44. J.-N. Ezingeard, S. Leigh Ernst, Y. Rebecca Chandler-Wilde, J.-N. Ezingeard, S. Leigh, R. Chandler-Wilde, Knowledge management at Ernst & Young UK: getting value through knowledge flows (2000). [Online]. Available http://aisel.aisnet.org/icis2000/93
45. I. Nonaka and A. Y. Lewin, A dynamic theory of organizational knowledge creation
46. P. Letmathe, M. Rößler, Tacit knowledge transfer and spillover learning in ramp-ups. Int. J. Oper. Prod. Manag. **39**(9–10), 1099–1121 (2019). https://doi.org/10.1108/IJOPM-08-2018-0508
47. P.M. Hildreth, C. Kimble, The duality of knowledge (2002)
48. S.M. Ferdous Azam, J. Tham, A. Albattat, Psycho-social perspectives of knowledge sharing and job performance in Malaysia: conceptual articulation. [Online]. Available www.ijstr.org
49. J.Cárcel-Carrasco, J.A. Cárcel-Carrasco, E. Peñalvo-López, Factors in the relationship between maintenance engineering and knowledge management. Appl. Sci. (Switzerland) **10**(8) (2020). https://doi.org/10.3390/APP10082810
50. I. Nonaka, R. Toyama, Strategic management as distributed practical wisdom (phronesis). Ind. Corp. Change **16**(3), 371–394 (2007). https://doi.org/10.1093/icc/dtm014
51. A. Daghfous, O. Belkhodja, Managing talent loss in the procurement function: insights from the hospitality industry. Sustainability (Switzerland) **11**(23) (2019). https://doi.org/10.3390/su11236800
52. O. Ibert, Towards a geography of knowledge creation: the ambivalences between 'knowledge as an object' and 'knowing in practice'. Reg. Stud. **41**(1), 103–114 (2007). https://doi.org/10.1080/00343400601120346
53. J.R. Gamble, Tacit vs explicit knowledge as antecedents for organizational change. J. Organ. Change Manag. **33**(6), 1123–1141 (2020). https://doi.org/10.1108/JOCM-04-2020-0121. (Emerald Group Holdings Ltd.)
54. K.E. Sveiby, Disabling the context for knowledge work: The role of managers' behaviours. Manag. Decis. **45**(10), 1636–1655 (2007). https://doi.org/10.1108/00251740710838004
55. R. Anwar, M. Rehman, K.S. Wang, M.A. Hashmani, Systematic literature review of knowledge sharing barriers and facilitators in global software development organizations using concept maps. IEEE Access **7**, 24231–24247 (2019). https://doi.org/10.1109/ACCESS.2019.2895690

An Efficient Approach to Stamp Verification

Ha Long Duy, Ha Minh Nghia, Bui Trong Vinh, and Phan Duy Hung

Abstract Stamps have become one of the most important security features in big companies where huge amounts of documents need processing daily. The stamp attached to a document is used to determine the authenticity of that document so that it is necessary to identify whether a stamp is forged or genuine. However, nowadays, it is easier for the general public to forge stamps. This paper presents a practical approach for stamp verification, based on the three stages process similar to some previous work: stamp segmentation, classification stamp or non-stamp, and stamp authenticity verification. In each stage, this work tries and tests new algorithms/methods to give a new way of solving the problem in each stage. Firstly, in our approach, an unsupervised learning machine method is implemented to detect all the objects in the input image, so all the regions including stamps and text are extracted. Next, two separate models of support vector machine classification are constructed. The first one is to distinguish between stamps and other objects in a document. The second model will determine the object which was classified as stamps in the first model whether it is genuine or not. The results show that this approach can perform the stamp verification tasks effectively.

Keywords Stamps verification · Image segmentation · Support vector machine

H. L. Duy · H. M. Nghia · P. D. Hung (✉)
FPT University, Hanoi, Vietnam
e-mail: hungpd2@fe.edu.vn

H. L. Duy
e-mail: duyhlhe141012@fpt.edu.vn

H. M. Nghia
e-mail: nghiahmhe140299@fpt.edu.vn

B. T. Vinh
Hanoi Procuratorate University, Hanoi, Vietnam
e-mail: vinhbt@tks.edu.vn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_74

1 Introduction

Nowadays, security is an indispensable aspect in the operation of big companies, especially in banks, insurance companies, and financial companies. Every day, thousands of invoices, contracts, certificates, and documents are handled. Therefore, it is necessary to have a way to guarantee the authenticity of the content. The stamp is one of the most widely used security methods, along with handwritten signatures. However, with the popularity of the Internet these days and the availability of technology, it is easier for the general public to forge stamps. Big companies process a huge volume of documents everyday so that the risk of forged stamps is very big. Companies can lose hundreds of thousands or even millions of dollars just by errors caused by forged stamps. Moreover, forged stamps can affect the company's reputation and the customer's trust [1]. In some banks, stamp verification relies on manual handling work instead of using software [2]. Therefore, there are many cases where the forged document is accepted as genuine. In order to solve this problem, an automated stamp verification method is needed to be developed.

In this paper, the process of stamp verification inherits the process by many previous works which divide it into three stages. This work follows this three stages process closely but for each stage, the way of implementation is different with different methods and algorithms being applied. The goal is to present a new practical method for this stamp verification problem. First, the stamp must be identified and extracted from the document image. By using color space transformations and k-means clustering, the scanned color image is split according to colors and then used the XY-cut algorithm; all the components in the image are segmented into candidates which contain both stamps candidates and other elements like logos and text. The second stage is to classify the extracted candidates to identify which candidates are stamps, which are not stamp. Each candidate image is passed through a support vector machine (SVM) model to identify whether that candidate is a stamp or not. After getting all the stamps, in the final stage, another SVM model will handle the task of classifying whether the stamps are genuine or forged.

2 Related Works

Unlike handwritten signatures, stamp verification is quite limited in terms of number of published works. Many previous works only focus on the task of detecting and segmenting stamps from a document. The majority of these works applied many different color-based, [3] shaped-based features [4], or used geometric features with key point descriptors to detect and segment stamps [5]. Their ultimate goal is to separate stamps from logos, text, and other information in original document images. Chen et al. [6] developed a method to detect stamps on checks of Chinese banks with a region-growing algorithm. Micenková el at [7] presented a new method for detecting and extracting stamps of various colors, even stamps that overlap with a signature or

text of a different color. There are two steps in this work, first is to separate every part of the image from text and background with color clustering; then, all the candidates obtained are classified to be whether stamp or not by using a set of features. In many works, which are dedicated to stamp detection and segmentation, the work that is worth paying attention to is the work by Younas et al. [8] with the name "D-StaR: A Generic Method for Stamp Segmentation from Document Images." They proposed a brand-new approach called D-StaR based on deep learning, capable of handling stamp segmentation in any color, shape, size, orientation, and this approach can even detect overlapping stamps. They used fully convolutional networks (FCNs) to segment stamp masks from scanned document images. Moreover, for pixel-based evaluation and reforming the original stamps, contour refinement is performed to the expected masks. D-Star is the first method to use deep learning to segment stamps and get amazing results. It is also the basis for other studies such as [7, 9, 10].

Many other previous works solved the stamp verification entirely, both stamp segmentation and stamp authentication verification. Chung et al. [11] focused on a dataset of Chinese antique seal/stamp; in the stamp segmentation step, they proposed a method based on geometric transformation and geometric transformation to find borders and align the perspective for two imprints. Then, to verify the authenticity of stamps, they calculated the similarity by PSNR and SSIM indexes as the detection metrics. Chung et al. improved their method in the next work [12]. The work by Takahiko Horiuchi [13] brings up a problem that many techniques for seal/stamp verification only solve this problem as a general pattern matching problem; therefore, the paper presented a new approach to deal with this problem. In the experiment step, an interesting way to test the method is used; the paper used both binary reference images and 3D reference to test the accuracy of the method. Works [14–17] presented some different ways in the stamp verification problem which include: using edge difference histogram, neural network, based on average relative error, and judging the percent of difference inside and difference outside to determine whether a stamp is genuine or not.

Finally, there are two works that inspired us to develop three stages of stamp verification in this paper, all by the author Micenkov et al. The first work [10] introduced three main stages of the problem. First, they segmented all every component in the original document image. Then, they used SVM to classify whether that component is stamp or not. Finally, they used features like color, shape, and print to verify stamps. With the same process, the work by Micenková el at [9] inherited two of their previous work [7, 10], and their work is improved by using k-means clustering; additionally, in the segmentation stage, they still used features to classify genuine and forgered stamps. In this study, we try and test new algorithms/methods to propose a best way for solving the problem in each stage.

3 Methodology

In this work, the hardcopy of the document is scanned in color, and the image is segmented completely. Candidates for solving the stamp identification problem are identified as rectangular segments, from which features are extracted. Candidates are classified as stamps or non-stamps using a binary classification system (logos, text etc.). Segments designated as stamps are further categorized to distinguish between legitimate and fake stamps; the process is shown in Fig. 1. Stamps are considered single-color (blue, green, red, etc.) items in this study. As a result, the primary principle behind detecting them in a picture is to group components that are the same color and are close to one another. Special examples of multiple-color stamps are detected as several stamps, which are subsequently combined.

This work identifies candidate segments by using this principle. One candidate segment is formed by each stamp in the image. If the image contains color logos or pictures, these (or single-color parts of them) are also identified as candidate segments. Then determine which parts match the genuine stamp and which are forged. The following section will give detail of this approach.

3.1 Segmentation

In the segmentation stage, the goal is to detect all the components in the original document image and crop each of those components into a new image. These new images will serve as a new material for the next classification stages. Because of the significant correlation among the channels, the RGB color model is not suitable for image segmentation. In order to work with color stamps, separate the backdrop, black text, and other roughly achromatic (white, black, and gray) components of the image

Fig. 1 Stages of the stamp verification

first, so the image is converted from an RGB image into a grayscale image. After that Otsu's binarization [18] with a threshold value of 200 is applied to split the image into background and object. The image is also blurred with Gaussian blur pixels of similar features are desired to be assigned the same label, spatially continuous pixels are desired to be assigned the same label, and the number of unique labels is desired to be large. With the image processed and ready to go, the next step is to find every cluster in the image; once again, the Otsu's binarization [18] is applied along with k-mean clustering is also applied. For each cluster, canny edge detection [19] of OpenCv is implemented to detect every edge on that cluster. Also, more OpenCv operations are applied including: cv2.dilate to make the object thicker, morphologyEx to close the cluster to the object, and then, finding the contour can be easy. Last, the largest contour is selected then crop each object detected and save it as a new image and move it to a different folders.

In order to segment every candidate properly, the XY-cut algorithm [20] is used to recursively partition the page into rectangles, resulting in candidate solutions with the smallest bounding boxes. In conclusion, the method in the first stage, we introduced above has four steps: (1) Preprocess image by resize, binarize, bur; (2) Find cluster; (3) For each cluster, detect edges, close figures, find contours; (4) Select largest contours. This method can be considered to be a good and simple way to deal with stamp segmentation problem.

3.2 SVM Classification: Stamp/non-Stamp

Now, all the candidates are segmented from the last step; those candidates contain all kinds of non-stamp candidates such as logos, parts of picture, or even forged stamps. First, the task is to classify whether that candidate is stamp or not, and finally, with all the candidates which were identified as stamp, another classification is needed to determine that stamp candidate is genuine or fraudulent. This process is clarified in Fig. 2.

This paper uses support vector machine for both classification models in each of two stages presented above. It is a supervised learning algorithm that is primarily used to categorize data. In this work, Linear SVM is used; the regularizer is applied in the final output layer, and the hinge loss is used as a loss function.

3.3 SVM Classification: Genuine Stamp/forged Stamp

Our final stage, also the most important stage, the task is to distinguish between genuine and forged stamps. In Fig. 3 is a sample of genuine and forged stamp, the difference can be recognized in hue and sharpness. Therefore, feature extraction is a suitable method in this case. This work uses several feature extractors from the scikit image library for feature extraction. These extractors help to create feature vectors,

Fig. 2 Two models for detection and verification

Fig. 3 Sample genuine and copied stamp after segmentation

(a) Genuine stamp (b) Forged stamp

which are then fed into the SVM classifier. The purpose is to test various scikit image feature extractor and select the ones that produce the best results. The following is the list of extractors used in this work: oriented FAST and rotated BRIEF (ORB), binary robust independent elementary features (BRIEF), center surround extremas for real-time feature detection and matching (CENSURE), histogram of the grayscale image, histogram of oriented gradients, and corner peaks.

For stamps, there are many different features such as hue, uniformity, and shape. In many cases, using a certain features extractor doesn't help to improve or even make classification performance worse. Therefore, it is necessary to determine which features extractors to combine to achieve the best result. A process needs to be developed to find out which features should be utilized. A sample of 100 images is extracted from the training set and then fed into the SVM model to correctly estimate which features reduce model accuracy. The accuracy is returned, and the features list with the highest accuracy is chosen. It was discovered that including the HOG and grayscale histogram extractors always produced more positive results, which is why they are used automatically to extract their respective features and feed them into the model. For the classifiers, the following models are chosen and implemented: logistic regression, SVM adj, SVM linear, SVC, k-nearest neighbors, decision tree, random forest, AdaBoost, Naive Bayes, gradient boosting, latent Dirichlet allocation. The feature vectors obtained from extractors are used as the input for these classifiers.

In the SVM classifier, it is necessary to choose two parameters C and gamma; these parameters have a big influence on the resulting accuracy. For this problem, a SkLearn machine learning library method called "GridSearchCV" [21] which can compute the best possible parameter is applied. With C, the value is determined from 1.0e-02 to 1.0e10. With gamma, the value is determined from 1.0e-09 to 1.0e3. This helps to increase the accuracy significantly compared to the default parameter C and gamma.

This testing process to find the best performing feature extractor is an improvement in this work. It helps to use feature extractor and classification algorithm and get the best performance in stamp verification possible.

4 Experiments and Results

4.1 Dataset and Evaluation

Evaluation is performed separately for stages: detection, classification, and verification. The public dataset of 400 document pages is used and available at http://madm.dfki.de/downloads-ds-staver. Stamped invoices with color logos and text are included in this dataset. Stamps are frequently layered with text or other objects. It was created by printing automatically generated invoices, manually stamping them, and scanning them in color at 600dpi. Lower resolutions were obtained by downscaling to reduce the effort of ground-truth labeling.

To evaluate the performance of verification, the dataset is split into training set and test set with the ratio 8:2. All of the models in this work are trained and tested on these two sets.

4.2 Result

The results obtained from two classification models are very positive. For the first SVM model, classify stamp or non-stamp, after training on a full dataset, the mean accuracy is 0.9. For the second classification, several extractors are used and tested to find the best combination of features among every possible combination of four extractors: ORB, corner peaks, BRIEF, and CENSURE. As presented in the previous section, histogram of grayscale image and HOG always produced better results so that always have to include them. After implementing with many classify algorithms, the best result of the mean accuracy is 0.96 as shown in Table 1.

The results show that this approach can be considered to be a good option when it comes to stamp verification problems. The process of determining which is the best way to utilize features extractor is a good improvement and can be applied in many other classification problems.

Table 1 Result by classifiers

Classifier name	Accuracy
Logistic regression	0.93
SVM, adj	0.82
SVM, linear	0.94
SVC	0.96
K-nearest neighbors	0.89
Decision tree	0.91
Random forest	0.97
Random forest 2	0.82
AdaBoost	0.93
Naive Bayes	0.84
Gradient boosting	0.93
Latent Dirichlet allocation	0.94

5 Conclusion and Future Works

This paper introduces stamp verification methods and algorithms. In each stage, we try, test, and recommend the best model for the process. Hence, this approach can be considered a simple and effective way to handle stamp verification problems. Other works can apply this approach, especially the works involving image segmentation or image classification.

In future work, we would like to develop this approach in more kinds of stamps. Right now, this paper mainly focuses on stamps from the public dataset that consist of entire stamps from foreign countries. The idea is to shift the focus to Vietnamese stamps, which contain Vietnamese letter patterns, symbols on them. This would make it easier to apply stamp verification methods for Vietnamese companies and organizations.

References

1. SBA Stone Forest.: Mitigating Risks Associated with Seals. Stone Forest Business Advisors. https://www.stoneforest.com.sg/business-advisors/Articles/Article/mitigating-risks-associated-with-seals. Accessed 1 Oct 2021
2. W. Gao, S. Dong, X. Chen, A system for automatic Chinese seal imprint verification, in *Proceedings of 3rd International Conference on: Document Analysis and Recognition, Montreal, Quebec* (vol. 2, 1995), pp. 660–664
3. B. Micenkov, J.V. Beusekom, Stamp detection in color document images. ICDAR 1125–1129 (2011)
4. Y. Bhalgat, M. Kulkarni, S. Karande, S. Lodha, Stamp processing with examplar features. arXiv:1609.05001 (2016)
5. P. Forczmanski, A. Markiewicz, Low-level image features for stamps detection and classification. CORES **2013**, 383–392 (2013)

6. L. Chen, T. Liu, J. Chen, J. Zhu, J. Deng, S. Ma, Location algorithm for seal imprints on Chinese bank-checks based on region growing. Optoelectron. Lett. **2**, 155–157 (2006)
7. B. Michenkova, J.V. Beusekom, Stamp detection in color document images. *Document Analysis and Recognition (ICDAR)* 1125–1129 (2011)
8. J. Younas, M.Z. Afzai, M.I. Malik, F. Shaifat, P. Lukowícs, S. Ahmed, D-StaR: a generic method for stamp segmentation from document images, in *Proceedings of 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (2017). https://doi.org/10.1109/ICDAR.2017.49
9. B. Michenkova, J.V. Beusekom, F. Shafait, Stamp verification for automated document authentication. Comput. Forensics 117–129 (2015)
10. B. Michenkova, Verification of authenticity of stamps in documents.http://acmbulletin.fiit.stuba.sk/vol3num4/micenkovaSPY.pdf. Accessed 1 Oct 2021
11. W.H. Chung, M.E., Wu, Y.L. Ueng, Y.H. Su, Forged seal imprint identification based on regression analysis on imprint borders and metrics comparisons, in *Proceedings of IEEE Conference on Dependable and Secure Computing* (2018)
12. W.H. Chung, M.E. Wu, Y.L. Ueng, Y.H. Su, Seal imprint verification via feature analysis and classifications. Futur. Gener. Comput. Syst. **101**(12), 458–466 (2019)
13. T. Horiuchi, Automatic seal verification by evaluating positive cost, in *Proceedings of Proceedings of Sixth International Conference on Document Analysis and Recognition* (2002), pp. 572–576
14. H. Jin, Z. Hao, L. Tiegen, Seal imprint verification using edge difference histogram, in *Proceedings of Optoelectronic Imaging and Multimedia Technology II* (2012), p. 855804
15. J. Hong-yu, G. Zhen, Research of the seal imprint time verification method based on neural network, in *Proceedings of International Conference on Intelligent System Design and Engineering Application* (2010), pp. 170–173
16. J. Liang, X. Tong, Z. Yuan, The circular seal identification method based on average relative error. Appl. Mech. Mater. 513–517 (2014)
17. W. Shuang, L. Tiegen, Research on registration algorithm for check seal verification, in *Proceedings of Electronic Imaging and Multimedia Technology V* (2007), p. 68330Y
18. N. Otsu, A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics vol. 9, no. 1, pp. 62–66 (1979) https://doi.org/10.1109/TSMC.1979.4310076
19. J.F. Canny, A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **8**(6), 679–698 (1986)
20. P. Sutheebanjard, W. Premchaiswadi, A modified recursive X-Y cut algorithm for solving block ordering problems, in *Proceedings of 2nd International Conference on Computer Engineering and Technology* (2010) , pp. V3-307–V3-311
21. G. Ranjan, A.K. Verma, R. Sudha, K-nearest neighbors and grid search CV based real time fault monitoring system for industries, in *Proceedings of IEEE 5th International Conference for Convergence in Technology* (2019), pp. 1–5

Fragmented Central Affinity Approach for Reducing Ambiguities in Dataset

Tanvi Trivedi and Mahipal Singh Deora

Abstract It is always a known fact that the role of data and its purity is very crucial in the data mining. The key role of data in the data mining is related from decision-making. It is well-known fact that if data are impure, then result will be a false picture. This crucial stage is also known as the ambiguities in datasets. Anomalous or irregular value in database is solitary of the biggest problems faced in data analysis and in data mining applications. Data preprocessing for the data mining is a key phase which is crucial place where ambiguities of database can be reduce or remove. The present study proposed an algorithm which tries to solve the problem related to an anomalous and irregular values, i.e., outliers, inliers, and missing values from a real-world imbalanced database. The study projected is based on the fragmented central affinity approach for reducing ambiguities in dataset.

Keywords Central affinity · Ambiguities · Dataset · Anomalous values · Missing values · Attribute

1 Introduction

Data in the dataset and database are denoted as the base unit. But, it is one of the most important segment of the whole data mining. Because the entire result is depends on the completeness of the data as well as the dataset. In the age of data science and analytics, the role of data is too crucial. Now, it becomes backbone of the information technology age. In the data mining or knowledge discovery in the database, the role of data is very crucial as it is base for decision taking. Data in the form of consolidated result of the analysis are more important to take decision. Likewise, the completeness of data as well as dataset is critical for any result or decision formations. Thus, the role of data preparation for data mining is the subject of key consideration.

This research paper is associated with an anomalous values which are identified and get a treatment to overcome from the characteristics of the anomalous values to

T. Trivedi (✉) · M. S. Deora
Bhupal Nobel's University, Udaipur, India
e-mail: tanvitri@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_75

remove ambiguities. The approach given here is an accommodation of several logics which need to reduce the ambiguities in respect of humanoid data entry approach and prone failures as well as improper evaluations. Consequently, it is valued to reconnoiter with anomalous data on or after the data sources used.

2 Methodology

The study tries to address the given problem in two consecutive section in which first section is related to identify the problem area. This section is completely associated with the process of pointing the anomalous values. The second area is associated with recovery of the missing values by the algorithm proposed for the study.

As the expansion point of view, both the stages are manage in the single elaborative algorithm which had a systematic approach to address the problem and produce conclusions to reduce the ambiguities.

The approach of reducing ambiguities covers several phage. These phage covers reading and importing dataset where missing/anomalous data are available. Afterward, the algorithm identifies the point of anomalous values. The major task which is taken in the consideration here is as follows:

- Reading and importing the dataset
- Deliberation of anomalous values
- Recover anomalous/missing values
- Test the significance.

3 Algorithm

The proposed approach in the form of algorithm is given below; this approach gives a picture to recover the missing values with the help of attribute vicinities. This approach developed the view of predicted data from the same attribute, by utilizing new concept of fragmented central affinity approach for reducing ambiguities in dataset. We can consider the details of the steps in the regular flow to reduce the ambiguities in the dataset.

Attribute/Variable/Dataset:

$V = (v_1, v_2, v_3 \ldots\ldots\ldots v_n)$.

where $V = V_{alb} + V_{aml} + V_{mis}$.

$V_{abl} = (v_1, v_2, v_3 \ldots\ldots\ldots v_k)$,

$V_{aml} = (v_{k+1}, v_{k+2}, v_{k+3} \ldots\ldots\ldots v_m)$.

$V_{mis} = (v_{m+1}, v_{m+2}, v_{m+3} \ldots\ldots\ldots v_n)$.

// after conversion of anomalous values in null values……..

Import: $V = (v_1, v_2, v_3 \ldots\ldots\ldots v_n)$.

for (i = 1to n),

If (value (v_i) = = NULL), then.

v_{a2} = value (v_{i-2})// Value of second above from v_i.
v_{a1} = value (v_{i-1})// Value of first above from v_i.
v_{b1} = value (v_{i+1}) // Value of first below from v_i.
v_{b2} = value (v_{i+2})// Value of second below from v_i.
V_{temp} =0 // Initialization of temporary variable
$V_{temp}(SUM) = \sum_{i=i-2}^{i+2} vi$
$V_{temp} = V_{temp}(SUM)/n$ where n=no. of values in fragmented array
Dv1, Dv2,Dv3 =0// Initialization of temporary variable
Dv1= (val (v_{i-1}) –val (v_{i-2}))
Dv2= ((val (v_{i+1}) –val (v_{i+2})) /2)
Dv3 = (val (v_{i+2}) –val (v_{i+1})).
Dv =0 // Initialization of variable
V_{pred} =0 // Initialization of variable
Dv = ((Dv1 + Dv2 + Dv3) / 3).
$V_{pred} = Dv + V_{temp}$.
val (v_i) = V_{pred} // Assigning estimated value to missing value place.
$i = i + 1$.
repeat until ($i >= n$).
Stop.

4 Dataset

The dataset consists of detail about Global Carbon Dioxide Emission from fossil burning by fuel type (Million tons) from the year 1964 to 2013. Thus, it is set of 50 records in the every column. The study includes three variables/attributes, namely year, coal, and natural gas on which study is conceded.

5 Operation Procedure

The dataset with diverse variables was occupied for the study. In the strategic way, after conversion of anomalous values in the missing values, the same is executed, the ratio of the deliberated missing values is 15%.

The projected algorithm alters the anomalous values of the dataset into the null values that is finally assumed as the missing values; after that these cases are resolved through the statistical inference.

The dataset of the study has been divided into three main classes as standard datasets, second missing attribute values, and lastly dataset with predicted or recovered values.

The consequence and data consolidation acknowledged from the dataset with 15% missing values. To get the solo value for projecting whole attribute, some statistical

consolidation is taken in the deliberation; these are arithmetic mean, standard deviation, and coefficient of variation. To test the significant of the result, analysis of variance ANOVA is utilized.

6 Data Analysis

Analysis with Mean: As per the Mean Values of variable of coal in different datasets are as mean received from the normal dataset about variable coal is 2351. After considering the 15% missing values in the same attribute, the mean values goes to 2049. Then, after approximation, the mean value comes on 2358.

For variable natural gas, the mean received from normal dataset is 994. Further, by including 15% of missing or null values, it goes to 845. After approximation, the consolidated mean values received are as 994. Therefore, it is observed that after approximation of missing values, the mean value of normal attribute is similar to approximated attribute for all three variables.

Standard Deviation: According result of standard deviation for variable coal, standard deviation for the normal dataset is received as 764. Similarly, the values of standard deviation for the 15% missing values of dataset are received 797 which is slightly differ from the after approximation values standard deviation is received as 766.

For variable natural gas, the values of standard deviation for the normal dataset for the variable natural gas are acknowledged as 429.07. Whereas, the standard deviation values with 15% missing values are acknowledged as 430; after approximation, value of standard deviation is observed as 429.

Analysis (Co-efficient of Variation): As per the values observed for the co-efficient of variation, it is observed that variables independently show similar nature. The likeness in the CV values for the entire datasets shows that there is homogeneousness in the variable values before and after the approximation.

ANOVA for Variable Coal

As per the hypothesis considered for the analysis of variance, it is assumed that as:

H0: $\mu 1 = \mu 2 = \mu 3$.

The alternate hypothesis assumed as at least one of the μ equivalences not kept which means one set mean value is not equal.

The test of ANOVA is applied on the dataset of coal (Table 1).

Table 1 Analysis of variance for coal

Source of variation	SS	df	MS	F
Between groups	24,232.6092	2	12,116.3	0.020196
Within groups	83,990,998.1	140	599,935.7	
Total	84,015,230.7	142		

Critical/tabulated value: F (2, 140: 0.5%) = 3.060

Table 2 Analysis of variance for natural gas

Source of variation	SS	df	MS	F
Between groups	6549.09295	2	3274.546	0.017763
Within groups	25,808,554.1	140	184,346.8	
Total	25,815,103.2	142		

Critical/tabulated value: F (2, 142: 0.5%) = 3.0608

7 Discussion of Result

According to the result, it is found that F (calculated) 0.020 < 3.060 F (critical). Hence, it shows that null hypothesis (H_0) is accepted at the 0.5% level of the significance.

ANOVA for Variable Natural Gas

As per the hypothesis considered for the analysis of variance, it is assumed that as:

H0: $\mu 1 = \mu 2 = \mu 3$

The alternate hypothesis assumed as at least one of the μ equivalences not kept which means one set mean value is not equal.

The test of ANOVA is applied on the dataset of natural gas (Table 2).

As per the result, it is observes that F (calculated) 0.01776 < 3.060 F (Tabulated). Therefore, it can be commit that null hypothesis (H_0) is accepted at the 0.5% level of the significance. It shows that there is no significant difference amid the various set of particularly variable natural gas in concerning to the mean values.

8 Conclusion

As per the result observed from the analysis the test of measurement of central tendency, this is noted that projected algorithm gives vital role to recover the missing values, and this is also observed from the result of the SD and values of CV. Thus, we can say that the result received from the algorithm supports dataset to reduce the ambiguities. According to the ANOVA result, the recommendation for the algorithm is very strong, and algorithm is very much suitable for small dataset till 300 to 500 records of time series.

References

1. A. Duraj, A. Niewiadomski, P.S. Szczepaniak, Outlier detection using linguistically quantified statements. Int. J. Intell. Syst. **33**(9), 1858–1868 (2018)
2. D.D. Pandya, S. Gaur, *Detection of Anomalous Value in Data Mining*, vol. 2 (Kalpa Publications in Engineering, 2018), pp.1–6
3. J. Lee, Y. Wonpil, Concurrent tracking of inliers and outliers. ArXiv E-prints. Retrieved 5 Oct 2015
4. B. Pratima, K. Muralidharan, Some inferential study on inliers in Lindley distribution. Int. J. Stat. Reliab. Eng. **3** (2), 108–129 (2016). ISSN (Print) 2350-0174. ISSN (online) 2456-2378
5. G. Sanjay, M.S. Dulawat, A perception of statistical inference in data mining. Int. J. Comput. Sci. Commun. **1**(II) (2010)
6. G. Sanjay, Estimation of missing value at extremes in data mining. Int. J. Adv. Found. Res. Comput. **14**(3), 13–19 (2014)
7. Y.H. Dovoedo, S. Chakraborti, Boxplot-based outlier detection for the location-scale family. Commun. Stat. Simul. Comput. **44**(6), (1492) (2015)

Correction to: Human Activity Recognition Using LSTM with Feature Extraction Through CNN

Rosepreet Kaur Bhogal and V. Devendran

Correction to:
Chapter "Human Activity Recognition Using LSTM with Feature Extraction Through CNN" in:
Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396,
https://doi.org/10.1007/978-981-16-9967-2_24

The original version of the book was inadvertently published with an incorrect line in the Abstract in Chapter 24 (Human Activity Recognition Using LSTM with Feature Extraction Through CNN). This has now been rectified and the line has been removed.

The chapter and book have been updated with the changes.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-16-9967-2_24

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2_76

Author Index

A
Aadityan Sridharan, 701
Adedeji Tomide Akindadelo, 369
Ajai Kumar, 453, 523
Ajay Kumar Sood, 179
Akshat Maheshwari, 179
Aliasgar Chatiwala, 683
Ambuja Salgaonkar, 337
Amit Mittal, 85
Amit Singh Patel, 741
Anjani Gupta, 751
Ankita Das, 533
Annie Rajan, 337
Anupam Kumar Sharma, 751
Aparna Hambarde, 497
Arabinda Rath, 417
Arman Rasool Faridi, 577
Arshad Shaikh, 337
Arshpreet Kaur, 655
ArunKumar Heddallikar, 51
Asfak Ali, 475
Asha Choudhary, 711
Ashish Kumar, 711
Ashtha Karki, 299
Ashwin R. Kubasadgoudar, 543
Atrayee Gupta, 267
Atreya Rastradhipati, 393
Atul Rana, 623
Axita Shah, 551
Ayodeji Olalekan Salau, 369

B
Babita Yadav, 611
Baig, M. M., 665
Baig, S. Q., 417
Bandari Pranay Kumar, 21
Bhoomi Gupta, 623, 633
Bhople Yogesh Jagannath, 447
Bidisha Banerjee, 267
Bınod Kumar, 197
Bisakha Biswal, 417
Bui Trong Vinh, 75, 781

D
Debesh Kumar Das, 475
Deepa Thilak, K., 137
Dessalegn Bitew Aeggegn, 369
Devendran, V., 245
Dhrubajyoti Baruah, 125
Dhyanendra Jain, 751
Dipesh Kamdar, 691

E
Esha Bhagat, 63

G
Gannera Mamatha, 169
Gaurav, A., 29
Gautham Rampalli, 21
Gayatri Devi, 417
Geethu, N., 323
Giriprasad, M. N., 169

H
Ha Long Duy, 781
Ha Minh Nghia, 781
Hardik Narendra, 223

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications*, Lecture Notes in Networks and Systems 396, https://doi.org/10.1007/978-981-16-9967-2

Harini Ragu, 91
Harshal Arolkar, 721
Harshal Shah, 683
Harshini Pothina, 205
Himadri Shekhar Roy, 267

I
Ismail Bin Musirin, 197

J
Jahnavi Bodhankar, 453
Jane Doe, 277
Jatin C. Modh, 355
Jatinderkumar R. Saini, 355
Jeyalakshmi, J., 257
John Doe, 277
Joydip Sarmah, 125
Jyoti Pareek, 551

K
Kalaiselvi, K., 137
Kamal Gayan, 125
Kaushal Shah, 347
Kavita Ajay Joshi, 85
Keerthana Prasad, 39
Kirtan Pathak, 63
Kirtirajsinh Zala, 149
Kirubakaran, M. K., 675
Komal Mehta Bhagat, 645
Krish Dadwani, 683
Krishna Sree, V., 29
Kumar Shashvat, 655
Kumaresan, K., 137

L
Lakshmikanthan, C., 485, 509
Lenali Singh, 453, 523

M
Madhu Shukla, 149
Mageshwari, M., 437
Mahesh B. Shelke, 187
Mahipal Singh Deora, 791
Malathi, M., 137
Malavika, R., 381
Manasi Khanvilkar, 393
Mandar Anil Chitre, 599
Mandeepsingh Mandla, 223
Manjula K. Pawar, 741
Manuj Joshi, 599
Meenakshi Agarwal, 179
Meet Amin, 683
Michael Shell, 277
Mohammad Luqman, 577
Monika Saini, 711
Mou Chatterjee, 761

N
Nagaraja, K. V., 205
Nalini C. Iyer, 543, 565
Nandini M. Chaudhari, 467
Nandini Mukherjee, 267
Naresh, R., 437
Navya, K. T., 39
Nguyen Huu Tien, 75
Nidhay Pancholi, 347
Nidhi, 589
Nikita, P., 565
Nilesh Dubey, 63
Nishant Desai, 683

P
Padmavathi, K., 197
Panchal, D. S., 187
Parul Patel, 159
Pattanashetty, B., 11
Pavan Kurariya, 453
Pawankumar Saini, 771
Pham Son Bach, 75
Phan Duy Hung, 75, 781
Pille Kamakshi, 21
Poongothai, T., 675
Poonkuzhali, S., 257
Pradeep Kumar Juneja, 85
Pradnya Chitrao, 771
Prakash Garia, 85
Prakashgoud Patil, 741
Pranali Nitnaware, 665
Pranoti Mane, 1
Prashant Chaudhary, 453, 523
Prashant Singh, 751
Prashant Vats, 751
Pratiksha Vallapure, 497
Preeti Singh, 289
Priti Ranjan Sahoo, 101, 299, 761
Priya Swaminarayan, 223
Pulkit Joshi, 523

R
Rachit Adhvaryu, 691

Raghavendra Shet, 543
Rajesh, M., 323
Rajeshkumar, T., 137
Rajkumar Mali, 497
Rakesh J. Ramteke, 467
Ram Sarkar, 475
Ranjan, 655
Rasika Kumbhar, 497
Rathod, D. P., 51
Ravi Shankar Jha, 101
Ravivanshikumar Sangpal, 497
Rejikumar, G., 405
Remya Vivek Menon, 405
Rosepreet Kaur Bhogal, 245

S
Saahith Reddy, M., 29
Sachin Gupta, 611, 623, 633
Sachin N. Deshmukh, 187
Sakthi, U., 675
Saleh Nagi Alsubari, 187
Samiran Chattyopadhyay, 267
Sandeep Kumar Sunori, 85
Santosh L. Deshpande, 733
Saranya, S., 137
Sarmistha Neogy, 267
Sarvpal Singh, 289
Sasikumar, P., 91
Satya Ranjan Phukon, 125
Saumitra Vatsal, 425
Sawarkar, S. D., 311
Sayema Kausar, 1
Senthil Murugan, T., 21
Senthilkumar, A. V., 197
Shalini Agarwal, 425
Shambhavi Kumar, 347
Shamshuddin K, 11
Sharan Kumar, 169
Shashi Pal Singh, 453, 523
Shashidhar, N., 565
Shila Jawale, 311
Shivoham Singh, 599
Shraddha Amit Kalele, 509
Shraddha S. Shinde, 51
Shraddha, B. H., 565
Shreeya Mishra, 645
Shrikant Sonekar, 1
Shrikant V. Sonekar, 665
Shuchita Upadhyaya, 589
Shweta Arora, 85
Shyam Sundar Meena, 235
Simranjitsingh Pabla, 223
Smrutirekha, 101, 299, 761
Snehal Tembhurne, 665
Somesh Sharma, 85
Sowmia, K. R., 257
Sravani Kumari, K., 29
Sreenivasulu, G., 111
Sriyanjana Adhikary, 267
Subhraneil Das, 39
Suchita A. Chavan, 467
Sudhanshu Maurya, 85
Suhas Khot, 497
Sundararaman Gopalan, 701
Suparna Dutta, 533
Supriya Sawwashere, 665
Suresh, A., 381
Swapnil Rana, 63
Swasti Patel, 223

T
Takele Ferede Agajie, 369
Tamizharasan, M., 509
Tanvi Trivedi, 791
Tatini Shanmuki Krishna Ratnam, 485
Thangaraj, K., 675
Thulasi Chitra, N., 111

U
Uday Khokhariya, 347
Umesh Kumar, 645
Urja Mankad, 721

V
Valerie David, 91
Vartika, 655
Ved Prakash Sharma, 751
Vemu Nikhil, 91
Venkatesh Mane, 543, 565
Venu Madhav Kuthadi, 111
Vidya S. Handur, 733
Vijay Anand, M., 137
Vineet Singhal, 179
Viral Parekh, 691
Vishal, P., 565
Viswanadha Raju, S., 111
Vrinda Tokekar, 235
Vu Thu Diep, 75

W
Wricha Mishra, 393

Y
Yalew Werkie Gebru, 369

Advances in Intelligent Systems and Computing 1386

Pushparaj Shetty D.
Surendra Shetty *Editors*

Recent Advances in Artificial Intelligence and Data Engineering

Select Proceedings of AIDE 2020

Springer

Advances in Intelligent Systems and Computing

Volume 1386

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Advisory Editors

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India

Rafael Bello Perez, Faculty of Mathematics, Physics and Computing, Universidad Central de Las Villas, Santa Clara, Cuba

Emilio S. Corchado, University of Salamanca, Salamanca, Spain

Hani Hagras, School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

László T. Kóczy, Department of Automation, Széchenyi István University, Gyor, Hungary

Vladik Kreinovich, Department of Computer Science, University of Texas at El Paso, El Paso, TX, USA

Chin-Teng Lin, Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan

Jie Lu, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia

Patricia Melin, Graduate Program of Computer Science, Tijuana Institute of Technology, Tijuana, Mexico

Nadia Nedjah, Department of Electronics Engineering, University of Rio de Janeiro, Rio de Janeiro, Brazil

Ngoc Thanh Nguyen, Faculty of Computer Science and Management, Wrocław University of Technology, Wrocław, Poland

Jun Wang, Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/11156

Pushparaj Shetty D. · Surendra Shetty
Editors

Recent Advances in Artificial Intelligence and Data Engineering

Select Proceedings of AIDE 2020

Springer

Editors
Pushparaj Shetty D.
Department of Mathematical
and Computational Sciences
National Institute of Technology Karnataka
(NITK)
Mangalore, India

Surendra Shetty
Department of MCA
NMAM Institute of Technology
Karkala, India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-16-3341-6 ISBN 978-981-16-3342-3 (eBook)
https://doi.org/10.1007/978-981-16-3342-3

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This volume of *Artificial Engineering in Data Engineering* contains the proceedings of the 2nd International Conference on Emerging Trends in Engineering—ICETE 2020 being held in NMAM Institute of Technology, Nitte, Karkala, Udupi, India, during December 22 and 23, 2020. Some of the best researchers in the field delivered keynote address on the theme areas of the conference. This gave an opportunity to the delegates to interact with these experts and to address some of the challenging interdisciplinary problems in the areas of artificial intelligence and data engineering.

AIDE 2020 attracted over 120 submissions. Through rigorous peer review, 41 high-quality papers were recommended by the international program committee to be presented at the conference and included in this volume. The AIDE 2020 conference is the second series of the international conference in artificial intelligence and data engineering that focuses on the recent technologies and trends in artificial intelligence and data engineering field in general.

AIDE 2020 was devoted to novel methods in the fields of machine learning, IoT, artificial intelligence, deep learning, computer networks, cloud computing, and data mining. The conference features several keynote addresses in the area of advanced computer science and information technology processing. These areas have been recognized to be the key technologies poised to shape modern society in the upcoming decade.

On behalf of the organizing committee, we would like to acknowledge the support from sponsors who helped in one way or the other to achieve our goals for the conference. We wish to express our appreciation to Springer for publishing the proceedings of AIDE 2020. We also wish to acknowledge the dedication and commitment of the AIDE Editorial Staff. We would like to thank the authors for submitting their work, as well as the technical program committee members and reviewers for their enthusiasm, time, and valuable suggestions. The invaluable help of the members from the organizing committee and volunteers in setting up and maintaining the online

submission systems, assigning papers to the reviewers, and preparing the camera-ready version of the proceedings is highly appreciated. We would like to profusely thank them for making AIDE 2020 a success.

Mangalore, India
Karkala, India
January 2021

Dr. Surendra Shetty
Dr. Pushparaj Shetty D.

Contents

Smart Environment and Network Issues

Machine Learning-Based Ensemble Network Security System 3
Prashanth P. Wagle, Shobha Rani, Suhas B. Kowligi, B. H. Suman, B. Pramodh, Pranaw Kumar, Srinivasa Raghavan, K. Aditya Shastry, H. A. Sanjay, Manoj Kumar, K. Nagaraj, and C. Subhash

Machine Learning-Based Air Pollution Prediction 17
Sheethal Shivakumar, K. Aditya Shastry, Simranjith Singh, Salman Pasha, B. C. Vinay, and V. Sushma

Crop and Fertilizer Recommendation System Based on Soil Classification 29
Pruthviraj, G. C. Akshatha, K. Aditya Shastry, Nagaraj, and Nikhil

Human Activity Classification Using Deep Convolutional Neural Network 41
Aniket Verma, Amit Suman, Vidyadevi G. Biradar, and S. Brunda

Intelligent IoT-Based Product Dispenser and Billing System 51
Roshan Fernandes, Anisha P. Rodrigues, Anusha Rani, Rachana Pandit, Vijaya Padmanabha, and B. A. Mohan

Models Predicting PM 2.5 Concentrations—A Review 65
Anusha Anchan, B. Shabari Shedthi, and G. R. Manasa

Performance Analysis of Modified TCP New Reno for MANETs 85
Sharada U. Shenoy, Udaya Kumar K. Shenoy, and M. Sharmila Kumari

Smart Health and Pattern Recognition

Identification of Helmets on Motorcyclists and Seatbelt on Four-Wheeler Drivers 99
Divyansh Saini, Vedashree Arundekar, K. V. Priya, and Divya Jennifer D'Souza

Prediction of Autism in Children with Down's Syndrome Using Machine Learning Algorithms 109
D. N. Disha, S Seema, and K. Aditya Shastry

Speech Emotion Recognition Using K-Nearest Neighbor Classifiers 123
M. Venkata Subbarao, Sudheer Kumar Terlapu, Nandigam Geethika, and Kudupudi Durga Harika

Object Detection and Voice Guidance for the Visually Impaired Using a Smart App 133
Ramya Srikanteswara, M. Chandrashekar Reddy, M. Himateja, and K. Mahesh Kumar

Application to Aid Hearing and Speech Impaired People 145
Akshatha Patkar, Steve Martis, Anupriya, Rakshith, and Deepthi G. Pai

Variants of Fuzzy C-Means on MRI Modality for Cancer Image Archives 161
C. K. Roopa, B. S. Harish, and R. Kasturi Rangan

A Review on Effectiveness of AI and ML Techniques for Classification of COVID-19 Medical Images 171
M. J. Dileep Kumar, G. Santhosh, Prabha Niranjajn, and G. R. Manasa

Medical Image Encryption Using SCAN Technique and Chaotic Tent Map System 181
Kiran, B. D. Parameshachari, and H. T. Panduranga

Utilization of Dark Data from Electronic Health Records for the Early Detection of Alzheimer's Disease 195
Sonam V. Maju and O. S. Gnana Prakasi

Brain Tumor Segmentation Using Capsule Neural Network 205
Jyothi Shetty, Shravya Shetty, Vijaya Shetty, and Chirag Rai

Forgery Detection and Image Recommendation Systems

Using Machine Learning for Image Recommendation in News Articles 215
Rohit Jere, Anant Pandey, Hasib Shaikh, Sulochana Nadgeri, and Pragati Chandankhede

An Approach to Noisy Synthetic Color Image Segmentation Using Unsupervised Competitive Self-Organizing Map 227
P. Ganesan, B. S. Sathish, L. M. I. Leo Joseph, B. Girirajan, P. Anuradha, and R. Murugesan

Building Dataset and Deep Learning-Based Inception Model for the Character Classification of Tigalari Script 239
Sachin S. Bhat, Alaka Ananth, Rajashree Nambiar, and Nagaraj Bhat

Handwritten Character Recognition Using Deep Convolutional Neural Networks 253
R. Shashank, A. Adarsh Rai, and P. Srinivasa Pai

Implementing Face Search Using Haar Cascade 263
Ramyashree and P. S. Venugopala

Deep Learning Photograph Caption Generator 277
Savitha Shetty, Sarika Hegde, Saritha Shetty, Deepthi Shetty, M. R. Sowmya, Reevan Miranda, Fedrick Sequeira, and Joyston Menezes

Streaming of Multimedia Data Using SCTP from an Embedded Platform 289
E. S. Vani and Sankar Dasiga

A Fast Block-Based Technique to Detect Copy-Move Forgery in Digital Images 299
Vaneet Kour, Preeti Aggarwal, and Ravreet Kaur

Bottlenecks in Finite Impulse Response Filter Architectures on a Reconfigurable Platform 309
Kunjan D. Shinde and C. Vijaya

Copy-Move Image Forgery Detection Using Discrete Cosine Transforms 327
R. P. Vandana and P. S. Venugopala

Sentiment Classification and Data Analysis

A Detail Analysis and Implementation of Haar Cascade Classifier 341
Gaurav Ghosh and K. S. Swarnalatha

Sentiment Analysis of Twitter Posts in English, Kannada and Hindi languages 361
Saritha Shetty, Sarika Hegde, Savitha Shetty, Deepthi Shetty, M. R. Sowmya, Rahul Shetty, Sourabh Rao, and Yashas Shetty

An Efficient Algorithm for Fruit Ripeness Detection 377
Sharath Kumar and Ramyashree

Kannada Document Classification Using Unicode Term Encoding Over Vector Space 387
R. Kasturi Rangan and B. S. Harish

Determining Stock Market Anomalies by Using Optimized *z*-Score Technique on Clusters Obtained from *K*-Means 401
Bibek Kumar Sardar, S. Pavithra, H. A. Sanjay, and Prasanta Gogoi

Data-Driven Strategies Recommendation for Creating MOOCs for Effective Online Learning Experience 415
Tanay Pratap and Sanjay Singh

A Neural Attention Model for Automatic Question Generation Using Dual Encoders 427
Archana Praveen Kumar, Gautam Sridhar, Ashlatha Nayak, and Manjula K Shenoy

A Comparative Study of Efficient Classification Models 441
Roopashri Shetty, M. Geetha, Dinesh U. Acharya, and G. Shyamala

Sentiment Classification on Twitter Media: A Novel Approach Using the Desired Information from User 449
B. Shravani, Chandana R. Yadav, S. Kavana, Dikshitha Rao, and Sharmila Shanthi Sequeira

About the Editors

Dr. Pushparaj Shetty D. is working as a faculty at National Institute of Technology Karnataka (NITK) Surathkal since 2000. He is presently holding the position of Associate Professor in department of Mathematical and computational sciences at NITK Surathkal. He obtained his B.E. in Computer Science and Engineering in 1999 from Mangalore University and M.E. in Computer Science and Engineering in 2005 from Indian Institute of Engineering Science and Technology, Shibpur, Kolkata (formerly Bengal Engineering and Science University). He obtained his Ph.D. from the Computer Science and Applications group, Department of Mathematics at the Indian Institute of Technology Delhi (July 2014). His research interests are in the area of wireless sensor networks, graph algorithms and their applications in engineering, cloud computing and high performance computing. He has co-authored more than 25 research papers in peer-reviewed journals and conferences including few book chapters. He is a senior member of IEEE, senior member of ACM, and life member of Computer Society of India (CSI) and Institution of Engineers India (IEI). Dr. Pushparaj is volunteering as the Chair of IEEE Mangalore Subsection for the year 2021.

Dr. Surendra Shetty is currently a Professor and Head, Department of Master of Computer Applications, NMAM Institute of Technology, Nitte. He had completed his B.Sc. in 2001 and he obtained his Master of Computer Applications in 2004 from Visvesvaraya Technological University, Belagavi. Dr. Surendra Shetty had been awarded his doctoral degree for his research work "Audio Data Mining Using Machine Learning Techniques" in 2013 from University of Mangalore. He has published more than 30 research papers in different international journals and conferences. He is currently guiding six research scholars. Dr. Surendra Shetty authored two book chapters in different publications entitled *Machine Learning Approach for Carnatic Music Analysis* and *Applications of Unsupervised Techniques for Clustering of Audio Data.* He has received research grant of 20 lakhs from VGST (GoK) for carrying out research on "Automatic Natural Language Processing and Speech

Disorder Problems in Kannada Language". He has 16 years of teaching experience. His research areas of interest are cryptography, data mining, pattern recognition, speech recognition, MIS, software engineering and testing.

Smart Environment and Network Issues

Machine Learning-Based Ensemble Network Security System

Prashanth P. Wagle, Shobha Rani, Suhas B. Kowligi, B. H. Suman, B. Pramodh, Pranaw Kumar, Srinivasa Raghavan, K. Aditya Shastry, H. A. Sanjay, Manoj Kumar, K. Nagaraj, and C. Subhash

Abstract A vital issue related to security for systems that are interconnected is the undesirable intrusions by agents. Various types of attacks are possible by the intruders. The attacks are associated with network security since the entry to the system can be accomplished through a network. Hence, in a typical organization, attacks can be external or internal attacks. In this work, the possibility of using Machine Learning (ML) techniques to detect network intrusions in combination with other techniques is explored. The outcome of this research will be the automatic application of firewall rules according to the present state of the network to prevent intrusions. Various classifiers are explored and then the Decision Tree (DT) classifier is implemented that functions as a signature-based system. For classification, the network traffic and logs of system in a UNIX based machine are captured. The frequent connections are identified, so that any anomaly in resource usage can be detected. This along with the classifier functions as an ensemble system to prevent network intrusions.

P. P. Wagle · S. Rani · S. B. Kowligi · B. H. Suman · K. A. Shastry (✉) · H. A. Sanjay · M. Kumar
Nitte Meenakshi Institute of Technology, Banglore 560064, India
e-mail: adityashastry.k@nmit.ac.in

H. A. Sanjay
e-mail: sanjay.ha@nmit.ac.in

M. Kumar
e-mail: manoj.kumar@nmit.ac.in

B. Pramodh · P. Kumar · S. Raghavan · K. Nagaraj · C. Subhash
UNISYS, Banglore 560035, India
e-mail: Pramodh.Bettadapura@in.unisys.com

P. Kumar
e-mail: pranaw.kumar@in.unisys.com

S. Raghavan
e-mail: srinivasa.Raghavan@in.unisys.com

K. Nagaraj
e-mail: Nagaraja.k@in.unisys.com

C. Subhash
e-mail: subhash.chanda@in.unisys.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_1

Keywords Network intrusion detection · Signature based · Decision tree · Firewalls · Parameters

1 Introduction

An intrusion detection System (IDS) represents a software application/device used for supervising a network/system to detect anomalies/violations of policy. Decision Tree algorithm denotes a technique for predictive modeling. Classification trees are the tree models in which the target variable is having discrete values. Historic information is taken by the predictive models for forecasting future values in an inexpensive manner. These models assist humans in taking critical decisions, making them more effective or in some instances the whole process of taking decisions can be automated [1].

The motivation behind this work is to explore predictive modeling techniques to accurately classify the malicious packets from the non-malicious ones and to choose one model out of Decision Tree (DT), Naive Bayes (NB), K-Nearest Neighbor (k-NN), and Logistic Regression (LR) with appropriate reasoning. After trying out the various models, the DT models were found to perform better with respect to accuracy and precision.

In this paper, the NSL-KDD dataset (NKD) [2] is used that denotes the improvised version of the KDD CUP 1999 dataset which is a database of connections. This database comprises of standard dataset that needs to be examined for several simulated intrusion types. A connection represents TCP packet sequence that begins and ends at certain definite times, amid which flow of data occurs to and from source IP address to a target IP address through a standard protocol. The connection is categorized as belonging to normal class or attack class. No connection is labeled with multiple attack types. Every connection instance comprises of approximately 100 bytes. Out of all the connections, exploratory data analysis (EDA) is done to choose only 10 attributes out of the 41 attributes of each connection.

The paper contents are as follows: Related work is discussed in Sect. 2. The data source with the associated parameters is described in Sect. 3. Section 4 examines the EDA with the implementation and the reasoning for choosing the DT algorithm over the others. Section 5 demonstrates the experimental setup and results. Section 6 concludes the research work done followed by references.

2 Related Work

The previous works for classifying packets as malicious or not using different ML models are mentioned in this section.

Kumar et al. [3] proposed a ML based IDS, mNIDS (Mobile-Network IDS) using tree and rule-based algorithms such as J48, Random Forest, RIDOR, JRIP, and PART.

Although mNIDS produced high accuracy, it was found that it can be further enhanced when used in conjunction with the traditional IDS such as knowledge-based anomaly detection systems and hybrid systems based on networks.

Yang Xin et al. [4] proposed a review of related work in ML and deep learning (DL) techniques for applications related to cybersecurity. Several commonly used datasets like DARPA, KDD Cup 99, NSL-KDD, and ADFA datasets were recommended. The authors also summarized outcomes of applying various ML and DL algorithms on these datasets.

Rishabh Das et al. [5] evaluated four ML models based on MODBUS data gathered from a gas pipeline. Numerous attacks have been categorized utilizing the ML algorithms, and lastly, the algorithm's performance has been measured. It was concluded that the RF algorithm might be more suitable as a core IDS algorithm for its optimal real-time performance.

Nutan Farah Haq et al. [6] conducted a survey on the designs of solitary, hybrid, and ensemble systems. The survey comprised of comparison of classification techniques in a statistical fashion. Along with the analysis of datasets, feature selection (FS) was also performed. It was found that utilizing diverse FS techniques and operating with the superlative technique was helpful for the classification and enhanced the deliberation of FS steps in identifying the intrusions. Additional examination demonstrated that solitary / baseline classifiers in measurement of performance can be substituted by ensemble classifiers.

Almseidin et al. [7] performed and evaluated several experiments to assess various ML classification algorithms on KDD intrusion dataset. They mainly focused on false negative (F_N) and false positive (F_P) to enhance the detection rate of the IDS. Experiments by authors revealed that the RF algorithm attained the highest rate of average accuracy while the DT algorithm obtained the minimum value for F_N.

The works mentioned above give a general idea about the type of dataset to use and the corresponding ML techniques to apply in order to increase accuracy of classification. Drawing from their conclusions and experimenting with other techniques, the DT classifier and NKD dataset is selected in this work. In this work, the built-in firewall, iptables, of Unix systems coupled with the proposed DT classifier (hybrid IDS) is utilized to block possible malicious traffic via whitelisting, inspired by [8], which was not implemented in the above-mentioned works. This work also deals with FS to filter irrelevant attributes and noise removal from the NKD dataset for obtaining an unbiased classifier as proposed by [2].

3 Dataset Description

This section describes the source of data from the KDD dataset [2] which is used to perform the experiments. It also describes the 42 parameters considered for class prediction. The NKD dataset [2] which is an improvised version of the KDD CUP 1999 dataset is employed in this work. Four classes are present in the dataset,

namely denial-of-service (DOS) like SYN flood attack, R2L (accessing in an unauthorized manner from a machine located remotely) attack like password guessing, U2R (accessing the privileges of super users in an unauthorized manner) attack like overflowing the buffers, and probing—surveillance, and other probing like scanning of ports. NSL-KDD dataset with 42 attributes is used in this work. This dataset represents an improvised version of the KDD CUP 1999 dataset in which redundant records have been deleted for achieving unbiased outcomes. Each database record represents a normal connection or an attack.

4 Proposed Work

This section describes the devised model for performing intrusion detection. Figure 1 demonstrates the big picture of the incoming network connection.

From Fig. 1, one can get to know the lifecycle of each incoming connection in the network. During the life cycle end, a connection is going to be classified as either malicious or normal. All incoming connections are subjected to the first and second line of defense. External connections may or may not be subject to the first line of defense (LOD).

Fig. 1 High-level design framework

4.1 First LOD (LOD-1)

The LOD-1 for protecting from network attacks is whitelisting specific IP addresses and ports on a network with the default behavior of "DROP." The iptables are made use of as the firewall on a Linux machine. Whitelisting has been implemented in the following manner, concerning a single node: Initially, all incoming and outgoing traffic is allowed by setting default policy to "ALLOW" on the host node's iptables INPUT and OUTPUT chains. After getting a list of frequently used IP addresses and their respective ports, the default policy of iptables is modified back to "DROP" and the obtained IP addresses and ports are whitelisted by changing their policy to "ALLOW".

The threshold value of the frequency can be selected as the mean of the total packets captured by all the peer machines. This threshold is found to exhibit a positive correlation with the number of packets captured by each machine. Also, the history of averages can be considered to further improve the value of the threshold. To accommodate new traffic rules, it is recommended to apply whitelisting in regular intervals, so that any changes in traffic flow will be reflected in the firewall rules. Hence, any unusual connections and surge of connections will be prevented ensuring high availability.

4.2 Second LOD (LOD-2)

The LOD-2 involves blacklisting of IP addresses showing malicious behavior. Firstly, the model is trained and deployed on the cloud. The connections are then subjected to the action of the model and then classified as malicious or not. Figure 2 depicts the architecture of LOD-2 used in this work.

Given the dataset, meaningful insights from the data need to be obtained for creating a model that classifies the connections as malicious or not. Data preprocessing techniques like EDA, data scaling, data encoding, data sampling, FS, dataset partition, along with prediction for model training has been performed in this work. The data is split into test and train datasets, and on that data, then, EDA is done. Based on the EDA, Fig. 3 shows the insights on the distribution of classes.

The next step involves the feature scaling of numerical attributes. It essentially aids in data normalization, so that the range is not exceeded. Occasionally, it assists in faster computations in the algorithm. The standard scaler where the values are replaced by the z-scores are used.

$$\text{Standardization, } z = \frac{x - \mu}{\sigma} \tag{1}$$

Fig. 2 Architecture of LOD-2

Fig. 3 Attack class distribution of the NSL-KDD dataset

where μ is the mean given by Eq. (2), σ is the standard deviation given by Eq. (3)

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{2}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{3}$$

The Min-Max scaling is given by Eq. (4)

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{4}$$

The next step would be encoding of the categorical values. Label encoding is done in which the classes are coded between 0 and n. After this, random oversampling is done for reducing the bias of majority classes over the minority classes. Overfitting of models results due to oversampling the redundant records in a random manner from the minority class of the training dataset. The results of random over sampling are as follows:

- Original dataset shape counter ({1: 67,343, 0: 45,927, 2: 11,656, 3: 995, 4: 52}).
- Resampled dataset shape counter ({0: 67,343, 1: 67,343, 2: 67,343, 3: 67,343, 4: 67,343}).

Out of the 42 features, certain features are insignificant. Most of them may be very interrelated among them, and hence, it causes redundancy. Hence, the features that have the highest impact on the accuracy of the ML model is chosen. The matrix or a correlation heatmap for finding the correlation between various values and eliminating one out of every pair of most correlated values. However, the RF Classifier is utilized as a means of assessing the feature importance of the dataset.

RFE involves feature ranking with recursive feature elimination. Out of these features, the features which impact accuracy of the ML model is chosen. The dataset is then partitioned into equal samples, so that there are only two classes which are normal and malicious. This helps in generalizing the proposed model and adding more attack classes in the future. Moreover, it is better to abstract the classes into normal and malicious as it is one of the goals of the research. After the data preprocessing is done, the model building is done. The three ML models, namely NB, DT, and LR, are implemented and compared. The DT model demonstrated higher accuracy. An improvised version of the CART algorithm was implemented in this work. The scientific design is as follows:

Given training vectors $x_i \varepsilon R^n$, $i = 1, \ldots, n$ and a vector of class labels $y \varepsilon R^l$, the DT algorithm splits the space in a recursive fashion, so that the instances possessing the similar classes are classified together. Algorithm-1 shows the DT for detecting the network intrusions.

Algorithm-1: Decision Tree (DT) for intrusion detection

INPUT: *S*, the set of classified instances
OUTPUT: Tree of Decisions
Require: $S \neq \emptyset$, *num_attributes* > 0
1: **procedure** BUILDTREE
2: **repeat**
3: *max-G* ← *0 //Maximum-Gain*
4: *divideA* ← *null*
5: e ← *Entopy(Features) // Entropy*
6: **for all** *Features s in F* do
7: *gain* ← IG(s,e)
8: **if** *gain* > *max-G* then
9: *max-G* ← *gain*
10: *divideA* ← s
11: **end if**
12: **end for**
13: Partition(*S,divideA*)
14: **until** all splits are processed
15: **end procedure**

Suppose the node at m is denoted as Q. For every split of candidate $\theta = (j, t_m)$ comprising of an attribute and threshold, the data is divided into $Q_{left}(\theta)$ & $Q_{right}(\theta)$ subsets.

$$Q_{left}(\theta) = (x, y)|x_j \leq t_m \tag{5}$$

$$Q_{right}(\theta) = Q|Q_{left}(\theta) \tag{6}$$

The *H()* computes the impurity at m. This selection relies on the job being performed (regression/classification).

$$G(Q, \theta) = \frac{n_{left}}{N_m} H\big(Q_{left}(\theta)\big) + \frac{n_{right}}{N_m} H\big(Q_{right}(\theta)\big) \tag{7}$$

Select the parameters that minimize the impurity as shown in Eq. (8).

$$\theta^* = argmin_\theta G(Q, \theta) \tag{8}$$

Recurse for subset $Q_{left}(\theta^*)$ & $Q_{right}(\theta^*)$ till the maximum permissible depth is obtained, $N_m < min_{samples} or N_m = 1$

The classification criterions are as follows. Suppose the target feature to be classified possesses the values 0, 1, …, K-1, for node m, signifying a region R_m with N_m records. Equation (9) represents the distribution of class k records in nodes

$$p_{mk} = 1/N_m \sum_{x_i \in R^m} I(y_i = k) \tag{9}$$

Gini and entropy are the two standard impurity measures. Equation (10) gives the formula for entropy a node in DT.

$$H(X_m) = -\sum_k p_{mk} \log(p_{mk}) \tag{10}$$

The information gain (IG) with entropy measure of impurity is chosen as the classification criteria in this research as shown in Eq. (11).

$$\text{IG} = Entropy(before) - \sum_{j=1}^{K} Entropy(j, after) \tag{11}$$

5 Experimental Setup and Results

This segment presents the outcomes of the classifiers NB, LR, k-NN, and DT. The confusion matrix (CM) and metrics of performance namely accuracy, precision, F1-score, support, and recall were used to evaluate the classifiers.

Table 1 depicts the matrix of confusion used for evaluating the classifiers [7, 8]

5.1 NB Classifier

For NB classifier, the cross-validation mean score was 0.973776072139, and the final model accuracy was found to be 0.973768617377. Tables 2 and 3 show the CM and the classification report of the NB classifier.

Table 1 CM [7, 8]

Forecasted class	Actual class	
	Positive	Negative
Positive	True +ve (T_p)	False +ve (F_p)
Negative	False –ve(F_N)	True –ve (T_N)

Table 2 NB-CM

Predicted values	Actual values	
	Positive	Negative
Positive	65346	1997
Negative	1536	65807

Table 3 NB-classification report

	Precision	Recall	F1-score	Support
0.0	0.98	0.97	0.97	67343
1.0	0.97	0.98	0.97	67343
Avg/total	0.97	0.97	0.97	134686

Table 4 LR CM

Predicted values	Actual values	
	Positive	Negative
Positive	65527	1816
Negative	773	66570

Table 5 LR-classification report

	Precision	Recall	F1-score	Support
0.0	0.99	0.97	0.98	67343
1.0	0.97	0.99	0.98	67343
Avg/total	0.98	0.98	0.98	134686

5.2 LR Classifier

For the LR classifier, the cross-validation mean score was 0.980822083372, and the model accuracy was 0.980777512139. Tables 4 and 5 depict the CM and classification report of the LR classifier, respectively.

5.3 K-NN Classifier

For the k-NN classifier, the cross-validation mean score was 0.996569816347, and the model accuracy was 0.99775774765. Tables 6 and 7 illustrate the confusion matrix and classification report of the k-NN classifier, respectively.

Table 6: k-NN CM

Predicted values	Actual values	
	Positive	Negative
Positive	67287	56
Negative	246	67097

Table 7: k-NN classification report

	Precision	Recall	F1-score	Support
0.0	1.00	1.00	1.00	67343
1.0	1.00	1.00	1.00	67343
Avg/total	1.00	1.00	1.00	134686

Table 8 DT CM

Predicted values	Actual values	
	Positive	Negative
Positive	67343	0
Negative	7	67336

Table 9 DT-classification report

	Precision	Recall	F1-score	Support
0.0	1.00	1.00	1.00	67343
1.0	1.00	1.00	1.00	67343
Avg/total	1.00	1.00	1.00	134686

5.4 DT Classifier

For the DT, cross-validation mean score was 0.999769836897, and the model accuracy was 0.999948027263. Tables 8 and 9 show the CM and the classification report of the DT classifier, respectively.

From Fig. 4, it was observed that the DT classifier outperformed all other classifiers based on the metrics evaluated.

The trained ML model was deployed in the cloud after converting the trained model into pickle format as illustrated in Fig. 5.

The machines which must be protected will be continuously sniffing the network and will be storing the connection data in a pcap file format. The abro(zeek) script and C script were employed to convert the pcap data into the NKD format and hence to collectively extract more information about the connections. The collected data was then sent in the NKD format to the cloud, so that the trained ML algorithm acts on the data sent by the various machines and then acts upon the data to generate iptables rules specific to a machine. The whole process is serialized by using a message queue which will do the asynchronous processing where multiple nodes will send the collected network data in real time to one processing machine. Hence, each machine is protected by the dynamic firewall rules that will be generated by the ML model deployed in the cloud.

Fig. 4 Comparison of models

Fig. 5 Classification in real time

6 Conclusion and Future Scope

In this research, several methods to limit the damage caused by the unknown connections are explored. This was done by whitelisting the frequently used connections by calculating the threshold dynamically in LOD-1. This will account for the internal attacks as the frequent connections can be identified and whitelisted. Various ML algorithms like k-NN, DT, NB, and LR are explored, and DT was chosen as it obtained the most accuracy among the others. The data in a typical UNIX machine

was collected using the packet sniffers, and the system logs were utilized to extract metadata of the connections. The ML algorithms were applied after storing the data in the database, so that the connections which are malicious will be blacklisted. In future, for LOD-1, a linear regression algorithm can be installed to calculate the value of the threshold instead of calculating the averages. Novel attacks can be detected by utilizing ML techniques. The data collected in the database can be used for mining and generating rules to prevent novel attacks.

References

1. M. Almseidin, M. Alzubi, S. Kovacs, M. Alkasassbeh, Evaluation of machine learning algorithms for intrusion detection system, in 15th International Symposium on Intelligent Systems and Informatics (SISY), pp. 277–282. IEEE (2018)
2. KDD CUP 1999 Dataset, https://kdd.ics.uci.edu/databases/kddcup99/task.html. Last accessed 3 July 2019
3. S. Kumar, A. Viinikainen, T. Hamalainen, Machine learning classification model for Network based Intrusion Detection System, in 11th International Conference for Internet Technology and Secured Transactions (ICITST), pp. 242–249. IEEE (2016)
4. Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, C. Wang, Machine learning and deep learning methods for cybersecurity. IEEE Access **6**, 35365–35381 (2018)
5. R. Das, T. H. Morris, Machine learning and cyber security, in International Conference on Computer, Electrical & Communication Engineering (ICCECE), pp. 1–7. IEEE (2017)
6. N. Haq, Md. Avishek, F. Shah, A. Onik, M. Rafni, Md. Dewan, Application of machine learning approaches in intrusion detection system: A Survey. Int. J. Adv. Res. Artificial Intelligence **4**(3) (2015)
7. M. Almseidin, M. Alzubi, S. Kovacs, M. Alkasassbeh, Evaluation of machine learning algorithms for intrusion detection system, in 15th International Symposium on Intelligent Systems and Informatics (SISY), pp. 277–288. IEEE (2017)
8. R. Patgiri, U. Varshney, T. Akutota, R. Kunde, An investigation on intrusion detection system using machine learning, in Symposium Series on Computational Intelligence (SSCI), pp. 1684–1691. IEEE (2018)
9. Z. Ahmad, S. Khan, C. Shiang, F. Ahmad, Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Trans. Emerging Telecommunications Technologies (2020). https://doi.org/10.1002/ett.4150

Machine Learning-Based Air Pollution Prediction

Sheethal Shivakumar, K. Aditya Shastry, Simranjith Singh, Salman Pasha, B. C. Vinay, and V. Sushma

Abstract Due to urbanization, air pollution (AP) has become a vital issue. AP is adversely impacting the humanity by causing asthma and other air-borne diseases. The accurate prediction of AP can aid in ensuring public health. Assessing and maintaining the quality of air has become one of the vital functions for the governments in several urban and industrial areas around the world today. There are a lot of parameters that play a significant role in increasing the air pollution like gases released from the vehicles, burning of remains of fuels, industrial gases, etc. So, with this increase in pollution, there is an increasing requirement of devising models which would record the information about the concentration of the pollutants. The deposition of the injurious air gases is adversely impacting people's life, especially in urban areas. Machine learning (ML) is a domain that has recently become popular in AP prediction due to its high accuracy. There are, however, many different ML approaches, and identifying the best one for the problem at hand is often challenging. In this work, different ML techniques such as linear regression (LR), decision tree regressor (DTR), and random forest regressor (RFR) algorithms are utilized to forecast the AP. Results revealed that the RFR performed better than LR and DTR on the given data set.

Keywords Air pollution · Prediction · Machine learning

1 Introduction

Globally, the rapid increase in population and economic upswing in cities has led to environmental problems such as AP, contamination of water, noise pollution, and many more. AP is a mix of solid elements and atmospheric gases. Emissions from vehicles, industrial chemicals, dirt, etc. are a major part of AP in metropolises. So,

S. Shivakumar · K. A. Shastry (✉) · S. Singh · S. Pasha · B. C. Vinay · V. Sushma
Nitte Meenakshi Institute of Technology, Bangalore 560064, India
e-mail: adityashastry.k@nmit.ac.in

V. Sushma
e-mail: sushma.v@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_2

keeping a check on the quality of air has developed into a major issue in many of the cities, as the increase in air pollution is directly proportional to the risks of having airborne diseases and skin-related diseases. AP has direct impact on health of humans. There is an increased public awareness about the same worldwide. AP is found to cause global warming, and rains constituting of acids among other climatic disasters. Asthma patients are found to suffer a lot due to AP. Precise air quality forecasting can reduce the effect of maximal pollution on the humans and biosphere as well. Therefore, better and improved forecasting of AP forms one of the critical factors for preventing major climatic disasters from taking place [1].

Currently, three approaches used for forecasting AP are statistical models, numerical methods, and ML. The first approach is based on statistical models. Statistical models are built on LR possessing single variable. They have shown negative correlation among the atmospheric variables (wind, precipitation, and temperature). The second approach is based on numerical techniques like transport of chemicals and diffusion of atmospheric components. Even though these techniques are used for forecasting AP, the accuracy of their prediction is very much reliant on updated sources that are very hard to obtain or generate. Furthermore, the characteristics of geography vary from location to location along with terrain conditions that are very complex in nature. These factors pose serious challenges in developing the numerical models for AP prediction, mainly because of the air flow complexity such as the speed of wind and its direction about the topographic feature. In contrast to pure models of statistics, ML techniques are able to consider numerous parameters in a single model. Lately, ML algorithms are becoming popular for predicting air pollution. They have the potential to predict the pollution in air more accurately than the traditional methods [2].

This work deals with the application of three ML algorithms, namely LR, DTR, and RFR on the data set downloaded from Kaggle [3] for AP prediction. The dataset comprised of a total of 14 input attributes described later in the paper. The target attribute was the air humidity (AH). Exhaustive trials were performed for the prediction of AP. Results revealed that the RFR with grid search performed accurately with lowest root mean square error (RMSE) when compared to RFR without grid search, DTR, and LR.

The paper contents have been organized into five main sections. Section 2 discusses the related work in the area of air pollution prediction using ML techniques. The proposed work is deliberated in Sect. 3, and the results are discussed in Sect. 4. The conclusion and future scope are provided at the end.

2 Related Work

This segment deliberates certain relevant works done in the area of air pollution prediction using ML techniques.

In [4], urban pollution was analysed and mapped as per the terrestrial areas. Tehran data was considered for experimentation. Moreover, authors compared the

accuracy of forecasts of LR and Naive Bayes (NB) algorithm. NB forecasts information more precisely than the other ML techniques for categorizing AQ classes which are not known. Methods related to regularization and optimization were used to forecast air pollutants' level (particulate matter, sulphur dioxide). Authors projected the values utilizing the data sets of two regions. One region predicted the O_3 and SO_2 values, while the second region forecasted O_3 and PM2.5 values. Data modelling was performed using similarity, and authors employed LR for classification. The assessment criteria utilized was RMSE. The major drawback of [4] is the inability of the LR techniques in forecasting unexpected events. Furthermore, the authors utilized data from only two regions that limited the model's generality.

In the work [5], a new deep learning (DL) technique was proposed for analysing IoT smart city data. Authors have proposed a new technique centred on long short-term memory (LSTM) networks to forecast future AQ values in an urban city. The authors demonstrated that the LSTM performed better than support vector regression (SVR). The work [6] proposes a method to predict the concentration of PM2.5 by utilizing the recurrent neural network (RNN) with LSTM. The results of [6] revealed that the devised approach was able to successfully estimate the PM2.5. During forecast in the long term, the value of PM2.5 was precisely forecasted to elevate. For predicting PM2.5 values that are distant via diverse climatic data related to flow of air, the progressing features of air contamination can be learnt more precisely. Precise PM2.5 forecast values can be provided by the model for lengthier periods of time. The LSTM is embedded in the RNN, where a directed graph is formed by the connections between units. This model has exhibited remarkable performance in several forecasting techniques that depend on historic data. Consequently, authors applied LSTM for solving such problem types. The work [7] mainly used the data gathered from IoT devices for the AQ forecast in smart cities. Data forms a major criterion for developing innovative solutions for smart cities. Regardless of the data prospects of smart cities, developing these solutions and applications forms a daunting task. The exposure to particulate matter which is a category of AP has triggered millions of deaths.

As observed from the above works, accurate AP prediction is the need of the hour as it can make the world healthier and safer. In this regard, ML techniques are being used effectively for AP prediction. Hence, this work focuses on accurate AP prediction using ML techniques.

3 Framework and System Design

This section discusses the framework and the system design of the proposed system for predicting the air pollution. The whole procedure is split into two segments, training and testing, as represented in Fig. 1.

In the first part, the training data (TD) was utilized to train the chosen algorithm. In the second part, the model is subject to testing. During testing, the model uses the previously generated results and predicts the value of target variable. The trained

Fig. 1 Model design using testing and training data

model is employed to assess the accuracy of the trained model. The same procedure is repeated using all algorithms under consideration, namely LR, DTR, and RFR algorithms. Figure 2 illustrates the design of the devised technique.

The air pollution data set with 9358 records and 15 attributes was obtained from Kaggle [3]. The analysis of the data set involved estimating the types of attributes. Most of the attributes were numeric in nature. Preprocessing the data set involved estimating the missing values that were denoted as −200 in the data. These missing values (−200) were replaced by zero. Since the data set was small, deleting the instances whose values were missing would make the data set even smaller. Hence, we applied mean imputation method for predicting the missing values. In this method, the missing values were substituted by the mean of the column. This step was repeated for every attribute (column) in the data set.

The heatmap was generated for the data set. The heatmap gave the correlation between the features of the data set. The two features having negative values indicated negative correlation between them. The data was then partitioned into training and test sets. The TD comprised of 70% (6550 records) of the total records. The remaining data (30%) of 2808 records comprised the test data. In this work, three ML algorithms, LR, RFR, and DTR models, were applied on the air pollution data set for predicting the air pollution. In DTR model, the maximum depth (d) of the tree was varied from 2 to 5. Better result was obtained for DTR with $d = 5$. For the RF algorithm, the parameters were searched using grid search as well as without grid search (parameters were applied randomly). Better result was obtained for RFR with grid search.

As the data set consisted of only numerical values, the RMSE was used for measuring the error rate of the prediction. Mean of the target attribute (AH) was taken as the threshold value. If the predicted value of the model goes above this threshold, then we can infer that the air pollution is high, else it is less.

Fig. 2 Design of the proposed system

3.1 Implementation Details

Stepwise procedure of the developed system is summarized below:

(i) Data set collection: A data set that is relevant to the problem statement is collected from a trusted source and cleaned to facilitate analysis. Cleaning involves dealing with missing values in any method best suited to the future steps. In our work, the data set of air pollution was collected from Kaggle [3].

(ii) Feature selection: The selected attributes possess divergent significance levels to the problem statement. This technique is to modify the crucial feature which the confirmation technique relies on. In this work, the correlation among the variables was obtained using the heatmap. The features having negative values were removed.

(iii) Exploratory data analysis (EDA): It represents a procedure in which preliminary examinations are performed on data set for identifying missing values, redundant records, and other data inconsistencies. This technique is utilized in the data preprocessing phase for preparing the data before the ML model is applied. In this work, EDA was done using the heatmap that gave the relationship among the features.

(iv) Training phase: A TD is a set of data that is used for learning the model and fitting the relevant parameters. Most techniques search through the TD for identifying realistic relationships and tend to overfit the data, meaning they can recognize and exploit obvious relations in the TD that do not embrace in general. In this work, the training was done using three ML models, namely LR, DTR, and RFR on a TD of 6550 records.

(v) Testing phase: A test data set is independent of the TD, but it possesses the same probability as that of the TD. Overfitting happens when the ML model delivers accurate predictions for TD but exhibits poor performance on test data. Therefore, a test set is utilized for performance assessment of ML models. In this work, the LR, DTR, and RFR models were evaluated on the test data of 2808 records.

The work was implemented in two phases. The first phase was data cleaning. The data set acquired from Kaggle [3] had 9358 entries. As the data set is already small, deleting rows consisting missing After applying various cleaning techniques as well as feature selection algorithms, we were left with a data set consisting of 10 attributes and 9358 records. Air Humidity represents the target feature. Missing values are tagged with −200 value. This set of data can be utilized solely for the purpose of research.

Table 1 Data set description

Attribute name	Attribute description
Date	Date of air quality measurement in DD/MM/YYY format
Time	Time of measuring the air quality in HH.MM.SS format
CO Conc	Concentration of carbon oxide on an hourly basis measured in mg/m^3
PT08.S1	Tin oxide measured hourly in microg/m^3
Benzene	Concentration of benzene on an hourly basis measured in mg/m^3
PT08.S2	Titania measured hourly
NOx	Hourly concentration of nitrogen oxide measured in ppb
PT08.S3	Tungsten oxide measured hourly
NO_2	Hourly measurement of nitrogen dioxide in microg/m^3
PT08.S4	Tungsten oxide measured hourly
PT08.S5	Indium oxide measured hourly
T	Temperature measured in degree per Celsius
RH	Relative humidity measured in percentage
AH	Target attribute representing absolute humidity

3.2 Dataset Details

The data set for experimentation was taken from the source: UCI data set type: xlsx with data set size: 9358 * 15 from [3]. Table 1 shows the description of the attributes of the data.

3.3 Algorithms

In this work, three ML algorithms were employed for the prediction of AP. They are discussed below.

(i) DTR: Decision tree (DT) learning represents a ML technique for modelling forecasts. The DT is employed as a predictive model in which the data records are represented in the intermediate nodes, while the class to be predicted will be in the leaf. The target attributes in DTR are continuous values representing real numbers. Due to their simplicity and fast learning, DT has become popular in recent times. A typical DTR used in our work is depicted in Fig. 3.

Fig. 3 DTR for AP prediction

(ii) RFR: RFR represents a learning technique based on ensembles of regression trees. Here, several regression trees are constructed during training, and the mean prediction of each tree is taken to give the final output. A RFR used in this work is illustrated in Fig. 4.

(iii) LR model: LR models the association among scalar target and 1/more independent variables (explanatory variables). The LR model utilizes linear predictor functions for estimating unknown parameters from the data. LR is a type of supervised ML learning technique that does a regression task. The target feature is forecasted based on the independent features. Its primary usage is in determining the associations among the variables along with prediction of continuous values. In this effort, the target feature representing AH is predicted using the LR model.

4 Experimental Set-up and Results

The trials were performed on Windows 10 OS with 1 TB HDD and 8 GB RAM. Python 2.7 was used as the programming language along with the libraries Pandas, NumPy, Scikit-Learn, Matplotlib, and Seaborn.

The heatmap in Fig. 5 shows correlation between variables.

Figure 6 depicts the comparison between the LR, DTR, and RFR algorithms.

The RMSE value shows that the model can relatively predict the data accurately. After speaking to certain industry experts, it is concluded that the least RMSE value shows that the model is fit and predicts accurately. Considering the above statement, it is observed that RFR without grid search gives the least RMSE value of 0.1019.

Fig. 4 RFR for AP prediction

5 Conclusion and Future Scope

In this effort, three ML algorithms, viz. LR, DTR, and RFR, have been employed for AP prediction. For modelling independent variables, we assess the interaction terms to assess whether the outcome of one independent feature relies on the values of other features. The RMSE value shows that the model can relatively predict the data accurately. The RMSE values obtained for the ML algorithms were between 0.2 and 0.5 indicating higher accuracy in AP prediction. Exhaustive trials demonstrated that RFR without grid search was the best-suited algorithm in comparison with LR and DTR models. As the data was numerical in nature, RFR provided the best results.

Fig. 5 Heatmap of the air quality dataset

Fig. 6 Comparison of algorithms for AP prediction

In future, this work can be extended to include real-time data to give more realistic prediction. A user interface can also be created to tell if the place is highly polluted or not. To improve the accuracy of the results, larger data set can be used. More robust ML algorithms can be utilized to perform better predictions.

References

1. V.R. Pasupuleti, P. Uhasri, S. Kalyan, H.K. Reddy, Air quality prediction of data log by machine learning, in 6th International Conference on Advanced Computing and Communication Systems

(ICACCS), Coimbatore, India, pp. 1395–1399 (2020)
2. G. Sakarka, S. Pillai, C.V. Rao, A. Peshkar, S. Malewar, Comparative study of ambient air quality prediction system using machine learning to predict air quality in smart city, in Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India. Lecture Notes in Networks and Systems, ed. by M. Dutta, C. Krishna, R. Kumar, M. Kalra, vol 116. Springer (2020)
3. https://www.kaggle.com/sayakchakraborty/air-quality-prediction-of-relative-humidity. Accessed online on 24 December 2019
4. S. Ameer, M. Ali Shah, A. Khan, H. Song, C. Maple, S. Ul Islam, M. Nabeel Asghar, Comparative analysis of machine learning techniques for predicting air quality in smart cities. IEEE Access **7**, 128325–128338 (2019)
5. I. Ko¨k, M. Ulvi Simsek, S. Ozdemir (2017) A deep learning model for air quality prediction in smart cities, in International Conference on Big Data (Big Data), pp. 1983–1990. IEEE (2017)
6. R.O. Sinnott, Z. Guan, Prediction of air pollution through machine learning approaches on the cloud. In IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT), pp. 51–60. IEEE (2018)
7. Y-T. Tsai, Y-R. Zeng, Y-S. Chang, Air pollution forecasting using rnn with lstm, in IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech‘), pp. 1074–1079. IEEE (2018)

Crop and Fertilizer Recommendation System Based on Soil Classification

Pruthviraj, G. C. Akshatha, K. Aditya Shastry, Nagaraj, and Nikhil

Abstract Agriculture forms a major occupation in countries like India. More than 75% people rely on farming for their daily wages. Hence, achieving good yield in the crops grown by farmers is the major concern. Various environmental factors have a significant impact on the crop yield. One such component that contributes majorly to the crop yield is soil. Due to urbanization and enhanced industrialization, the agricultural soil is getting contaminated, losing fertility, and hindering the crop yield. Machine Learning (ML) is employed for agricultural data analysis. The proposed ML based model aims at classifying the given soil sample datasets into four different classes, namely very high fertile, high fertile, moderately fertile, and low fertile soil utilizing support vector machine (SVM) technique. It also predicts the suitable crops that can be grown based on the class which the soil sample belongs to and suggests the fertilizers that can be used to further enhance the fertility of soil. Using proposed model, farmers can make decisions on which crop to grow based on the soil classification and decide upon the nitrogen–phosphorous–potassium (NPK) fertilizers ratio that can be used. Comparison of the SVM algorithm with k-nearest neighbor (k-NN), and decision tree (DT) has shown that SVM performed with a higher accuracy.

Keywords Machine learning · Soil classification · Support vector machine · Crop yield

1 Introduction

Data analysis is a procedure which involves data inspection, cleaning the data, and modeling the data with the aim of extracting useful information and conclusions. It is a technique of analyzing, extracting, and predicting meaningful information from vast data. Companies make use of this process to turn their client's raw data to

Pruthviraj · G. C. Akshatha (✉) · K. A. Shastry · Nagaraj · Nikhil
Nitte Meenakshi Institute of Technology, Banglore 560064, India
e-mail: akshatha.gc@nmit.ac.in

K. A. Shastry
e-mail: adityashastry.k@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_3

useful information. This process of data analysis can also be used in the agriculture sector. Most of the farmers have been relying on their long-term field experience on specific crops to expect a higher yield during the next harvesting period. But they still do not get the crops' worth price. It is mostly because of selection of inadequate crops or improper irrigation. Hence, agricultural researchers stress on the need for an efficient mechanism to predict and improve the crop growth. Majority of research works carried out in agriculture field, focusing on biological mechanisms to identify crop growth and improve its yield. The crop yield primarily depends on various factors like diversity of crop, quality of seed, and various environmental factors like sunlight, temperature, soil (pH), water (pH), rainfall, humidity, and variety of soil nutrients. By analyzing soil of a particular region, best crop to obtain more yield can be predicted. This prediction will help farmers to identify the appropriate crops that is grown in their farm according to the soil classification. The estimation of yield harvest is a tedious task because it is affected by various factors such as soil productiveness, weather conditions, and practices followed for cultivation such as amount of irrigation and date of sowing [1–3].

Soil classification is one of the interesting areas of research in the engineering field. The traditional methods such as cone penetration test (CPT) consume lots of time and need experts for extracting accurate results. Latest approach to soil classification is to use ML techniques. Several ML algorithms such as k-NN, DT, and SVM can be used. SVM is suitable to solve classification problems [1–3]. Hence, in our proposed work, we are using SVM for classification of soil nutrition level into different classes like class 0(Low fertile), class 1 (moderately fertile), class 2 (high fertile), and class 3 (very high fertile). We have used SVM algorithm to classify the given soil data samples founded on numerous measurable attributes of soil. Similarly, based on these features, the soil can be classified as class 0 class 1, class 2, and class 3. And also, with the same information, suitable crop can be grown which will give more yield that will be predicted.

The remainder of this manuscript is comprised of four sections. Section 2 converses about the related work carried out in the field of soil classification. Section 3 deliberates the methodology followed for classifying the given soil sample datasets based on the various nutrients in the soil. Section 4 confers about the experimentation carried out on the sample soil dataset and validates the efficiency of the methodology used. The paper ends with conclusion section.

2 Related Work

Extensive amount of study is being done for classifying the soil samples using different dataset. Zaminur et al. [4] presented a hybrid model that can predict soil series based on land type, and as per the prediction, they will suggest suitable crops but do not specify the fertilizers that can be utilized to improve the soil fertility.

Ashwini et al. [5] proposed the classification technique centered on SVM. The soil samples are graded using different scientific features such as texture, drainage,

color, and terrain. These soil samples are classified into red, black, clay, alluvial, etc., categories but do not specify about fertility. Jay Gholap et al. [6] proposed a classification approach using classification algorithms such as JRip, Naïve Bayes, J48, and regression algorithms like least median square regression and linear regression. The classification is not based on soil nutrients and does not recommend crop and fertilizer ratio. Ramesh et al. [7] have proposed a classification model which is based on Naïve Bayes classifier. In their work, the classification is not based on soil nutrients and do not propose a new approach.

Vibha et al. [8] designed a hybrid model based on clustering and classification. They mainly focus on soil nutrient mining. The different soil parameters are not considered for classification. The authors do not recommend how to improve the fertility of the soil. Sofianita Mutalib et al. [9] in their model use the techniques k-means and self-organizing map (SOM) for classifying the soil. It is basically comparing the performance of two existing algorithms, and the classification is not done based on the soil nutrients. It is mainly based on color, texture, etc.

The proposed model mainly focuses on classifying the soil records gathered from GKVK UAS, Bangalore, Karnataka, India. The dataset has 1550 soil samples obtained from Tarikere, Kadur, Sringeri, and Koppa, places of Chikamagaluru district, of Karnataka. Our proposed approach based on SVM is applied to the soil samples, and the classification is done based on various metal compositions such as calcium, magnesium, potassium, lime, and soil moisture content, which denote the soil parameters. It also aims at solving farmer problems by predicting the effective crops that can be grown in the given class of soil and also recommends NPK fertilizer ratio to increase the nutrients in soil based on the class to which it belongs, which will help farmers to get more crop yield.

3 Proposed Work

Motivated from the work done by various researchers in the domain of soil classification, we proposed an SVM based classifier for soil classification. The key idea behind the SVM algorithm is to generate the hyper plane that can partition n-dimensional space into classes such that a new data point can be placed in the appropriate class. SVM selects the extreme point or vectors called support vectors which aid in creating the hyper plane [8].

3.1 SVM Algorithm

SVM is the ML algorithm which determines the decision planes that creates decision boundaries. A decision boundary divides the class of objects from one another. Support vectors have highest prominence in ML field. The main objective of the Kernel function is to abstract the data and transform to the required form. SVM is

an employed method both for classification and regression tasks and used widely for classification. SVM algorithm takes data as input and creates a hyper plane that splits the information into classes. Let us consider the subsequent case, let us assume we have two objects: circle and triangle, and our data has two key features let us denote it as x and y. What is that we need is a classifier that, which is given a pair of (x, y) coordinates, it outputs 1 if it is either circle or triangle. Labeled training data is plotted on a plane as shown in Fig. 1 [9].

SVM reads the data records and generates the hyper plane that best splits the labels. Multiple hyper planes may exist, among which it is necessary to identify the best hyper plane that splits the labels in many accurate ways shown in Fig. 2. In SVM, best hyper plane is the one that maximizes the margins from both objects [10].

In the above example, the data could be separated linearly; hence, we were able to draw a hyper plane to separate circles from triangles. Usually things are not that simple. Sometimes we must deal with nonlinear data as demonstrated in Fig. 3 [7].

From Fig. 3, it is clear that there is no linear decision boundary that divides both objects. To resolve it, a third dimension is added as shown in Fig. 4 [11].

In 3D, a hyper plane is parallel to x-axis at some z. Remapping it back to two dimension, a decision boundary is obtained which is a circumference of some radius that separates the objects triangle and circle using SVM as demonstrated in Fig. 5 [11].

Fig. 1 Illustration of two classes; blue dots depicts class 1 samples; red triangles depicts class 2 samples [9]

Fig. 2 Optimized hyper plane separating classes of two-dimensional data using SVM [10]

Fig. 3 Illustration on nonlinear data [7]

The proposed SVM based classifier is built by analyzing the soil sample training dataset. Class labels for given data are predicted during the classification step. The dataset records and their related categories of classes under analysis are split into training set and test set. 70% of the soil sample data are utilized for model training, and remaining 30% of the sample data are used for testing purposes. Individual tuples that constitute the training data are sampled randomly from the information under

Fig. 4 Application of kernel function in SVM for separating nonlinear data [11]

Fig. 5 Nonlinear data separated by hyper plane [11]

examination. The remaining tuples are independent of training tuples and constitute the test data, which means that the test set will not be utilized to construct the classifier. The test set is utilized to obtain the accuracy of classifier. The classifier accuracy is based on the proportion of test records that are classified correctly by the classifier. Based on the threshold values of the nutrients that we have used in our datasets, we will be able to classify them into different classes like class 0 (low fertile), class 1 (moderately fertile), class 2 (high fertile), and class 3 (very high fertile), and based

on these classes, we will be able to predict the suitable crops and NPK fertilizers ratio. The proposed approach is represented in Fig. 6.

Description of the soil dataset [12]: The dataset used in our proposed work is obtained from GKVK UAS which include samples from different area of Chikkamagaluru district. It includes samples from various taluk of Chikkamagaluru district like Tarikere, Kadur, Sringeri, and Koppa. There are 1550 soil samples in the dataset, and the attributes of datasets are as follows: N, P, K, Ca, Mg, Lime, C, S, and moisture.

Crops' health depends on sufficient supply of soil nutrients and moisture. As the supply of moisture decreases, plant normal function and development are disturbed, and crop yields are reduced. Few soil sample data records used in our proposed approach is as demonstrated in Table 1.

The threshold nutrient values for the soil samples that we have used in our proposed approach is demonstrated in Table 2.

Fig. 6 Proposed workflow

Table 1 Soil sample dataset

Ca	Mg	K	S	N	Lime	C	P	Moisture
9.653	6.585	142	108	226.05	5.83	1.29	18	0.9
19.88	22.2	339.35	77	308.25	6.45	2	298	0.8
2.931	41.22	514.29	108	277.42	6.43	0.74	48	0.6

Table 2 Threshold values

	Low	Medium	High	Very high
Ca	0–1.5	1.5–11.2	11.2–33.6	33.6–45.1
Mg	0–18	18–36	36–54	54–72
K	0–150	150–300	300–450	450–620
S	0–35	35–80	80–120	120–170
N	0–200	200–300	300–400	400–1430
Lime	0–6	6–7	7–8	8–14
C	0–0.6	0.6–1.4	1.4–2.2	2.2–3.0
P	0–85	85–170	170–255	255–350

4 Experimental Results and Analysis

Experimental results show that the proposed SVM based classifier was able to classify 845 soil samples in to correct classes and misclassified 240 soil samples resulting in 77.85% accuracy. The results are demonstrated in Table 3.

Analysis: The overall performance of a proposed SVM based classifier is described using the confusion matrix as shown in Fig. 7. The rows indicate actual class labels, and columns indicate the predicted labels. According to the matrix, 58% of the class-0 labels are also predicted as class-0 labels. 27% of the class-0 labels are misclassified as class 1. 1% of the class-0 labels are misclassified as class 2. 14% of the class-0 labels are misclassified as class 3. 3.8% of the class-1 labels are misclassified as class 0. 91% of the class-1 labels are classified as class 1 only. 0.4% of the class-1 labels are misclassified as class 2. 0.46% of the class-1 labels are misclassified as class 3. 0.12% of the class-2 labels are misclassified as class 0. 0.8% of the class-2 labels are misclassified as class 1. 74% of the class-2 labels are classified as class 2. 5.9% of the class-2 labels are misclassified as class 3. 16% of the class-3 labels are misclassified as class 0. 0.02% of the class-3 labels are misclassified as

Table 3 Results of proposed approach

Correctly classified soil sample instances	Incorrectly classified soil sample instances	Total number of testing records	Accuracy of proposed approach
845	240	1085	77.85%

Fig. 7 Confusion matrix for proposed classifier

class 1. 0.9% of the class-3 labels are misclassified as class 2. 63% of the class-3 labels are classified as class 3.

Classification of soil is done based on the level of fertility. The values which are the features of the tested soil samples are given as input to the system. It is classified as to which class it belongs. There are four classes, namely class 0, class 1, class 2, and class 3. Class 0 refers to low fertile soil, class 1 refers to moderately fertile soil, class 2 refers to high fertile soil, and class 3 refers to highly fertile soil. Based on the classification, the crops that can be grown are suggested. This helps the farmers to grow the best-suited crop that is adaptable to their soil conditions. If the soil belongs to class 0, then crops like beans, green peas, carrot, and onion can be grown. If the soil belongs to class 1, then crops like radish, cow pea, cabbage, cucumber, and sweet potato can be grown. If the soil belongs to class 2, then crops like sugar cane, paddy, bajra, guava, acacia, etc., can be grown. If the soil belongs to class 3, then crops like barley, cotton, tobacco, sunflower, etc., can be grown.

The experimental results are depicted as shown in Figs.8 and 9.

The snapshot demonstrated in Fig. 8 shows the classification of given input soil sample dataset into class 0 and suggests the suitable crop such as beans and chickpea that can be grown, along with the fertilizer ratio.

The snapshot shown in Fig. 9 shows the classification of given input soil sample dataset into class 1 and suggests the suitable crop that can be grown, along with the fertilizer ratio.

Fig. 8 User interface for classifying the soil sample and suggestion of crop and fertilizer ratio

Fig. 9 User interface of the classified soil result of the proposed classifier and suggestion of crop and fertilizer ratio

4.1 Comparison of SVM with k-NN and Decision Tree

The soil sample dataset is verified against three classification algorithms like k-NN, DT, and SVM. Initially, the k-NN algorithm (K-nearest neighbor) is used for classification. For k-NN, the training and test data has been loaded. Then, we were supposed to choose the k value (which is nearest data point). K will be the new point. After that, for each point in test data, we computed the distance between data points and each row of training data with the help of Euclidian distance from new data point. We have included the test record for the given class label. But for the dataset chosen, k-NN gives less accuracy when compared to the other algorithms as it is susceptible to imbalanced dataset. Accuracy obtained using k-NN algorithm is 72.04%. Later DT algorithm has been used. In DT algorithm, soil parameters was classified based on the threshold value in each node. Based on the threshold value, soil samples are classified into different class level at the leaf or the last node. But it gave less accuracy when compared with SVM since it is prone to overfitting. Accuracy we got using DT algorithm is 69.46%. Finally, our proposed model was built using the SVM algorithm by using the training and testing dataset. Our dataset has 1550 records where we are using 70% data for training and remaining 30% of data for testing. SVM obtained higher accuracy because it captures nonlinearity in dataset (Table 4).

Table 4 Comparison of ML algorithms

Algorithm	Accuracy
SVM	77.85
k-NN	72.04
DT	69.46

Fig. 10 Accuracy comparison between SVM, k-NN, and DT

Figure 10 shows the comparison of accuracies between the SVM, k-NN, and DT algorithms.

As demonstrated in Fig. 10 for the SVM algorithm, we have got 5% to 8% of improvement in accuracy level when we compared with k-NN and DT algorithms. Accuracy we got using SVM algorithm is 77.85. From these results, we can infer that the SVM algorithm performed better classification when compared to k-NN and DT.

5 Conclusion

The proposed model aims at classifying the given soil sample datasets into four different classes, namely low fertile, moderately fertile, high fertile, and very high fertile soil, using SVM algorithm. It also predicts the suitable crops that can be grown based on the class which it belongs to and also suggests the NPK fertilizers ratio that could be employed to obtain better yield of the crops grown in case if the fertility of the soil is not up to the mark. Using this model, farmers can find out the fertility class of their field soil and can make decisions on which crops to grow, so that they will get better yield and fertilizer ratio to be used to further improve the fertility of the soil.

In the proposed classification model, linear SVM is used for classifying the given soil sample dataset. In future, the same work can be extended using kernel SVM on the same soil dataset to see how well it performs in terms of classifying soil type, given N, P, K, Ca, Mg, Lime, C, S, and moisture parameters as input. Hence, kernels

like radial basis kernel, Gaussian kernel, etc., can be explored as a part of future work. Moreover, future work may also include incorporation of neural networks on the same training dataset to see if the classification accuracy improves any further.

References

1. K.A. Shastry, H.A. Sanjay, H. Kavya, A novel data mining approach for soil Classification, in 9th International Conference on Computer Science & Education, Vancouver, BC, pp. 93–98 (2014)
2. K. Aditya Shastry, H.A. Sanjay, G. Deexith, Quadratic-radial-basis-function-kernel for classifying multi-class agricultural datasets with continuous attributes. Appl. Soft Comp. **58**, 65–74 (2017)
3. S. Panchamurthi, Soil analysis and prediction of suitable crop for agriculture using Machine Learning. Int. J. Res. Appl. Sci. Eng. Technol. **7**(3), 2328–2335 (2019)
4. S. Al Zaminur Rahman, K. Chandra Mitra, S.M. Mohidul Islam, Soil classification using ML methods and crop suggestion based on soil series, in 21st International Conference of Computer and Information Technology (ICCIT), pp. 1–4. IEEE (2018)
5. A. Rao, A. Gowda, R. Beham, Machine Learning in soil Classification and crop detection. Int. J. Sci. Res. & Dev. **4**(1), 792–794 (2016)
6. J. Gholap, A. lngole, J. Gohil, S. Gargade, V. Attar, Soil data analysis using classification techniques and soil attribute prediction. Int. J. Comput. Sci. Issues **9**(3), 1–4 (2012)
7. R. Vamanan, K. Ramar, Classification of agricultural land soils: A data mining approach. Agricultural J. **6**(3), 379–384 (2011). https://doi.org/10.3923/aj.2011.82.86
8. L. Vibha, G.M. HarshaVardhan, S.J. Prasanth, P. Deepa Shenoy, K.R.L. VenuGopal, M. Patnaik, A hybrid clustering and classification technique for soil data mining, in Proc. ICTES. IEEE, pp. 1090–1095 (2007)
9. S. Mutalib, S-N Fadhlun Jamian, S. Abdul-Rahman, A. Mohamed, Soil classification: an application of self organizing map and k-means, in: Proc 10th IEEE-ISDA. IEEE, pp 439–444 (2010)
10. G. Yi-Yang, Ren, Nan-ping, Data mining and analysis of our agriculture based on the DT, in Proc. IEEE-ISECS, pp 134–138 (2009). 10.1109 ICCCM. 5267962
11. N. Jain, A. Kumar, S. Garud, V. Pradhan, P. Kulkarni, Crop selection method based on various environmental factors using ML, February 2017.
12. K. Srunitha, S. Padmavathi, Performance of SVM classifier for image-based soil classification, in 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 411–415. IEEE (2016)

Human Activity Classification Using Deep Convolutional Neural Network

Aniket Verma, Amit Suman, Vidyadevi G. Biradar, and S. Brunda

Abstract The human activity classification is an identification of person's actions from an image or video frames captured by a camera. This helps in tracking person movements which are useful for various purposes. The human activity classification is a challenging task due to factors like complex background captured in the image, occlusion, and camera view angle. Therefore, there is a need for development of a robust human activity classification system. In this paper, a classification model based on deep convolutional neural network is proposed which extracts coordinates of human joints and parts affinity field from an image for predicting human pose and classification of activity. The model is designed to carry out classification of six major activities, and they are raising hand, pick-up an item, squats, namaste greeting, on-phone, and pointing a gun. These activities represent a class of suspicious and non-suspicious activities. The activity "namaste greeting" is included considering the COVID-19 issue. The proposed deep convolutional neural network comprises of two branches for joints coordinates estimation and parts affinity field estimation. The model is trained on MPII human pose dataset and tested on live videos captured from a webcam to evaluate its performance. The model gives highest training accuracy of 99.7 and 99.2% test accuracy.

Keywords Deep convolutional neural network · Pose estimation · Activity recognition · Parts affinity field

1 Introduction

In this era of online media, the amount of images and video databases are increasing having unbelievable progress. Using the data, there is a huge scope for computer vision community to come up with algorithms that can build smarter surveillance systems by validating the human activity, in order to prevent the occurring of criminal activities. Human activity recognition is a method in which one can predict the

A. Verma · A. Suman · V. G. Biradar (✉) · S. Brunda
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: vidyadevi.g.biradar@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_4

behavior of a person based on the trace of their movements. Basically the idea behind is that once the activity of the subject is identified and known, then a smart computer system model can provide assistance. In recent studies, human activity recognition has seen success by using deep learning techniques to solve problems.

Human activity classification aims to classify activities of a person from a still image or video frames. Image-based activity classification plays a significant role in various applications in developing healthcare systems, identification of suspicious activities, robotic applications, and event analysis. Human activity classification is a challenging task due to factors like complex background in the video or image, interaction of people leading to overlapping of body parts, occlusions, and more. A 2D human pose estimation is the core task in identification and classification of human activities. Human pose estimation techniques exploit the information on human body joints coordinates and their limbs associations.

Human activities are identified by observing movements in body joints and main joints of human skeleton. The traditional methods have limitations to accurately classify and identify human activity for scenarios where people in the scene are interacting in some way. Figure 1a and b show images with persons that are close to each other, and there is an overlap of limbs belonging to different person. It is evident from the stick figure as highlighted in Fig. 1a and b that their joints are overlapped. Deep learning models have been excelled in solving most of the computer vision problems due to their learning capability when subjected to large dataset. Deep convolutional neural networks are the major category of deep learning models, and these models are used in solving many problems like object detection, face detection, and image classification. This paper implements a model for human activity classification using deep convolutional network with two branches.

Fig. 1 Part affinity issue **a** persons hands joints are overlapping with body joints, **b** hand joints of middle person occluded

Human pose estimation is carried out using a CNN-based regression model for the extraction of body joints. This model uses multistage cascaded regressors, and results indicate that CNN regression model is efficient in localization of hidden joints. The research work in [1] presents a joint model consisting of CNN and graph model for prediction of human pose. A two stage is presented which comprises of two models, first is a multistage CNN model designed for extraction of required features from the image and the second is a prediction model that predicts the pose from the features extracted by CNN model [2].

The paper is organized into different sections, Section 2 gives a brief survey on human activity classification methods. Section 3 presents the methodology of proposed model for human activity classification. Section 4 demonstrates results, and conclusion and scope for future enhancement are given in Section 5.

2 Literature Survey

Literature review has been carried out to understand the methodology and limitation of various existing methods for human pose estimation. The advantages and disadvantages are summarized in Table 1.

The study of existing methods indicates that there are still challenges in human pose estimation, and there is a need for robust algorithm. This paper proposed a human activity classification model based on deep learning technique which is discussed in the next section.

3 Proposed Model

In this work, a double-branch CNN model is used for estimation of human pose and human activity classification. The upper branch is for detection of joints coordinates, and lower branch is for computation of parts affinity field as shown in Fig. 2. This work utilizes images from MPII human pose dataset for training the deep learning model. The images for activities like namaste greeting are locally generated as images in this category are not in plenty in MPII database, which are also used for training and testing. Human pose estimation requires estimation of joints coordinates and parts affinity field. This has been carried out using a double-branch CNN model as shown in Fig. 2.

The image is fed into double-branch CNN in which the upper branch estimates the coordinates of body joints and the lower branch estimates parts affinity fields. This information is combined to estimate human pose, and it is fed to dense network for classification. The model is tuned to accurate classification of six activities, and they are raising hand, pick-up an item, squats, namaste greeting, on-phone, and pointing a gun. These activities represent the categories of suspicious and non-suspicious activities.

Table 1 Methods for human activity classification

Paper	Techniques	Advantages	Disadvantages
Hao Du et al. [3]	This paper presents a method to classify human activity using transfer-learned residual network based on micro-doppler spectrograms. The performance of the model was evaluated on CMU Mocap dataset	Accuracy obtained 97% Jitter is less, results' deflection is less	Requires more computational capacity to train CMU Mocap dataset
Xiaoran Shi et al. [4]	Authors of this paper have used deep learning techniques on spectrograms data. The method uses convolutional generative adversarial networks and DCNN	It offers good scalability Requires less time to train	Accuracy is less than 82% Jitter is high, results deflect too often
Zhe Cao et al. [1]	This paper presents an effective method to detect 2D pose from images containing multiple people The method uses greedy bottom-up parsing strategy	It offers high accuracy, Jitter is very less, results rarely deflect Suitable for real-time application	Requires more computation power
Cheng Ning et al. [5]	This paper proposes a method to increase the accuracy of classification by improving Single Shot Detection (SSD) algorithm. The method is designed by incorporating architectural features of pretrained inception model	Algorithm is robust	Jitter is too high; results deflect too often Less accuracy for small objects

(continued)

Table 1 (continued)

Paper	Techniques	Advantages	Disadvantages
Fabian Caba et al. [6]	This paper introduces three possible applications where ActivityNet can be used. They are untrimmed video classification, trimmed activity classification, and activity detection	In terms of richness of taxonomy and activity diversity, more varieties are presented using ActivityNet	The obtained accuracy is less
Jaeyong Sung et al. [7]	In this paper, an inexpensive sensor called RGBD was used to take the input. To model different properties of human activities, a two-layered maximum entropy Markov model (MEMM) was used	The algorithm showed an accuracy of 84.3% on the training dataset (what activities)	The algorithm showed an accuracy of 64.2% on test dataset
Andrej Karpathy et al. [8]	In this paper, UCF-101 video dataset was used; a slow fusion network using CNN was developed	The slow fusion network out performed over early fusion and late fusion network; it showed an accuracy of 80% on sports data	The algorithm did not show much difference for human object interaction, body-motion, human–human interaction, and playing instruments
Ivan Laptev et al. [9]	This paper uses local space–time features, space–time pyramids, and multichannel nonlinear SVM to improve state-of-the-art results on standard KTH action dataset	The proposed algorithm achieves 91.8% accuracy on std. KTH action dataset	This algorithm need to be improved in the script-to-video alignment and extending the video collection to a much larger dataset

Fig. 2 Human pose estimation mode

The MPII human pose dataset is used in this work, and it has 25,000 images with annotated body joints. The dataset covers 410 human activities which is a benchmark dataset for validation of deep learning algorithms. This research work focuses on classification of six activities, and they are person raising hand, person picking up an item, person on-phone, and person pointing a gun. Therefore, to enhance the dataset and avoid model overfit, additional human pose dataset to represent these activities is created using local Web camera.

The architecture of a two-branch multistage CNN model shown in Fig. 3 is used for 2D human pose estimation. This model makes prediction approximations and computed the confidence map and affinity fields of points and associates each joint with another. Confidence maps are probability distribution that represent the confidence of the joint location at every pixel in given image and part affinity filed associates body joints of person. The images are resized to 224 × 224 and fed to the double-branch CNN. The input image is analyzed using VGG-19 pretrained model, and the resulting feature image is fed to multistage double-branch CNN. The predictions from multistage network are concatenated with image features which are fed to the next stage.

A set of 18 confidence map for each joint is generated by the upper branch, and the lower branch predicts a set of 38 parts affinity fields and represents the degree of association between joint parts. Part affinity fields help in pruning the graph that represents stick figure. The double-branch CNN model used for pose estimation gives the coordinates of all the body joints as its output. These coordinates are written to a text file. The text file is then converted into a csv file and that file is fed to the DNN.

Fig. 3 Architecture of two-branch multistage CNN [1]

4 Results and Discussion

The DCNN model has been trained on MPII pose data and custom generated dataset which comprises of actions corresponding to namaste and point a gun. The dataset is further balanced by duplicating pose images for namaste. This model runs under Windows 10 working framework, and the PC setup is Intel Core i7 9750H CPU, Nvidia GTX 1650 4 GB GPU. The frame rate per second of estimating human pose in 2D is 6 FPS for VGG 16 model using GPU (GTX 1660).

The model is trained with batch size 20 and learning rate 0.001 on TensorFlow. The testing of the model is carried out on downloaded videos and live videos captured from webcam. The human joints coordinates are extracted from these video frames which are stored in a csv file. Sample images are shown in Fig. 4.

The research work focus on classification of six activities in one-person image or video frame. The motivation behind choosing raising hand, person picking up an item, person on-phone, and person pointing a gun is that these represent suspicious and non-suspicious activities. The activity 'namaste' is chosen keeping present COVID-19 pandemic in view where instead of handshake greeting, namaste is recommended. The activity 'pointing-gun' is for suspicious activity, and other activities represent social behavior.

Figure 5a–f shows the results of activity classification and stick figure of human pose for videos downloaded from YouTube. The activity is indicated by a caption on input images.

Fig. 4 Human pose and skeleton **a** person **b** skeleton of person in **a**

Fig. 5 Results showing body joints of a person and the actions performed videos collected from YouTube **a** Namaste, **b** pointing gun, **c** on-phone, **d** raising hands, **e** squats, and **f** picking an item

Figure 6a–f shows the results for videos captured on fly using Web camera. The model shows robust performance for both the cases.

The model is trained by increasing number of epochs with adam optimizer and categorical cross entropy loss function. Figure 6 shows the performance of the model for six activities. Figure 6a shows increase in the model accuracy with increase in number of epochs, and Fig. 6b shows decease in the model accuracy with increase in number of epochs. The model achieves accuracy of 99.7% training accuracy and 99.2% of test accuracy. The model shows satisfactory performance (Fig. 7).

Fig. 6 Results showing body joints of a person and the actions performed from live videos **a** Namaste, **b** pointing gun, **c** on phone, **d** raising hands, **e** squats, and **f** picking an item

Fig. 7 Model performance **a** graph of accuracy vs epochs and **b** graph of loss vs epochs

5 Conclusion

In this paper, a deep convolutional neural network model is proposed for human activity classification for six activities, they are raising hand, pick-up an item, squats, namaste greeting, on-phone, and pointing a gun. The MPII human pose dataset is used for evaluation of model. The model has given encouraging results for still images, YouTube videos, and for the videos captured using webcam on fly. The model gives training accuracy of 99.7% and 99.2% test accuracy which is satisfactory. The future work may be to classify images with multiple persons in the image.

References

1. Z. Cao, T. Simon, S. Wei and Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, in Computer Vision and Pattern Recognition (CVPR), IEEE Conference, pp. 1302–1310, 2017
2. H.D. Mehr, H. Polat, Human activity recognition in smart home with deep learning approach, in 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG). Istanbul, Turkey **2019**, 149–153 (2019)
3. H. Du, Y. He, T. Jin, Transfer learning for human activities classification using micro-doppler spectrograms, in IEEE International Conference on Computational Electromagnetics (ICCEM), Chengdu, pp. 1–3, 2018
4. X. Shi, Y. Li, F. Zhou, L. Liu, Human activity recognition based on deep learning method, in International Conference on Radar (RADAR), Brisbane, pp. 1–5, 2018
5. C. Ning, H. Zhou, Y. Song, J. Tang, Inception single shot multiBox detector for object detection, in International Conference on Multimedia & Expo Workshops (ICMEW), pp. 549–554, 2017
6. F. Caba Heilbron, V. Escorcia, B. Ghanem, J. Carlos Niebles, Activitynet: A large-scale video benchmark for human activity understanding, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 961–970
7. J. Sung, C. Ponce, B. Selman, A. Saxena, Human activity detection from RGBD images, plan, activity, and intent recognition, 64 (2011)
8. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-scale video classification with convolutional neural networks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1725–1732 (2014)
9. I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, Learning realistic human actions from movies, in 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1–8). IEEE, June 2008

10. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, SSD: Single shot multibox detector, in Computer Vision and Pattern recognition, pp.1–7, 2015
11. P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1627–1645 (2010)
12. Y. Zhang, J. Liu, K. Huang, Dilated Hourglass Networks for Human Pose Estimation, in Chinese Automation Congress (CAC), pp. 2597–2602, 2018
13. G. Gkioxari, B. Hariharan, R. Girshick, J. Malik, Using k-Poselets for Detecting People and Localizing Their Keypoints, in IEEE Conference on Computer Vision and Pattern Recognition, pp. 3582–3589, 2014
14. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016
15. E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, DeeperCut: A deeper, stronger, and faster multi-person pose estimation model, IEEE transactions of Pattern Analysis and Machine Learning, Vol. No.9910 pp. 34–50, 2016
16. S. Johnson, M. Everingham, Clustered pose and nonlinear appearance models for human pose estimation, in British Machine Vision Conference, pp.1–11, 2010
17. W. Ouyang, X. Chu, X. Wang, Multi-source Deep Learning for Human Pose Estimation, in IEEE Conference on Computer Vision and Pattern Recognition, Columbus, pp. 2337–2344, 2014
18. Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, Y. A. Sheikh, OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. In IEEE transactions on pattern analysis and machine intelligence, pp.1–14, 2019
19. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
20. Z. Zhigang, D. Guangxue, L. Huan, Z. Guangbing, W. Nan, Y. Wenjie, Human behavior recognition method based on double-branch deep convolution neural network, in Chinese Control And Decision Conference (CCDC), Shenyang, pp. 5520–5524, 2018

Intelligent IoT-Based Product Dispenser and Billing System

Roshan Fernandes, Anisha P. Rodrigues, Anusha Rani, Rachana Pandit, Vijaya Padmanabha, and B. A. Mohan

Abstract The rise of wireless technology and other communication techniques has made electronic commerce extremely famous. Buying at broad malls is becoming a regular occurrence in metro towns. For these areas, there is a massive rush during weekends and holidays. People buy different things and place them in a trolley. After purchases are completed, one has to go to the payments billing counter. The cashier checks the bill at the billing counter using a barcode reader which is a very time-consuming procedure which results in a long queue at the billing counter. A smart product is the one that helps make daily life easy, convenient, and secure. In this paper, we present an innovative shelf-checkout concept that helps smart shopping and billing. The main concept here is to support a person shopping on a regular basis in terms of the reduced amount of time spent when buying a product. The proposed work also implements an analysis of the products to gain a best-buy option. To achieve this feat, we have incorporated a recommendation system along with the general product display. The main objective of this work is to provide a technology-oriented, cost-effective, easily scalable, and robust system for personal shopping.

Keywords RFID · Smart shopping · E-commerce · Cosine similarity · Smart recommendation system · Arduino microcontroller

R. Fernandes (✉) · A. P. Rodrigues · A. Rani · R. Pandit
Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, Udupi, India
e-mail: roshan_nmamit@nitte.edu.in

A. P. Rodrigues
e-mail: anishapr@nitte.edu.in

V. Padmanabha
Department of Mathematics and Computer Science, Modern College of Business and Science, Baushar, Muscat, Sultanate of Oman

B. A. Mohan
Department of Computer Science and Engineering, NMIT, Bengaluru, India
e-mail: mohan.ba@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_5

1 Introduction

Due to the advent of wireless technology and other networking innovations, e-commerce has become more and more trendy. Shopping has become a regular occurrence at large mall in major cities. People buy something special and put it in their cart. A billing counter for payment should be installed upon completion of the order. The cashier at the billing counter prepares a bill employing the barcode reader which is a very time-consuming operation and leads to long billing counter queue. The key concept in the proposed work is to help a customer, when they have less time to purchase a product in day-to-day shopping. The customer can sprint away from the shop once he or she has done shopping without waiting in the queue for billing. The idea behind the device is based on RFID technology, where RFID [1] tags are inserted into items and user cards and RFID readers are accustomed to reading this tag for the store's effective, efficient, and theft-controlled service. In using the new method, most of the flaws of the barcode technique are solved. In the proposed system, GSM technology is used to alert the customer for the best goods. Building a small prototype of a smart shopping network that can make the process of buying goods and handling payments makes easier for the customer. The proposed work also keeps track of information about the customer. Smart cards and "shelf check-out" can capture customer buying preferences, provide useful data to help the service providers understand consumer dynamics, maximize inventory rates, and minimize congestion in grocery stores. This is accomplished by using an analytics component in the proposed work.

The proposed work is based on RFID technology, where RFID tags are inserted in the access cards and RFID readers are used to read these tags for safe, secure, and theft-controlled shop service. In the proposed method, GSM technology is used to warn consumers with the products taken. Because of advanced warning systems, it is automatically added to the cart whenever a user picks up an object. If the package is not correctly checked out or the payment fails, the exit is blocked for the object till in possession.

Content-based solution involves a fair deal of knowledge on the correct aspects of the objects, instead of using experiences and input from the consumers. For instance, it may be object attributes such as category, year, manufacturer, price, or other important aspects using natural language processing. The natural language processing involves analyzing the feedbacks given by the customers. Collaborative filtering, on the contrary, requires little more than the historical preference of users over a collection of objects. Since it is based on historical evidence, the main assumption here is that in the future the people who have agreed to purchase the items in the past will continue to do the same in the future too. The proposed work also provides an intelligent recommendation system for the customers, based on the purchase history of the same customer or other customers having similar interests. This helps the customer to make a better choice in terms of the item to be purchased, choice of the brand, price savings, and many more. For this to accomplish, the proposed method uses cosine similarity using matrix factorization. Equation (1) gives the cosine similarity.

$$\mathrm{Sim}(u_i, u_k) = \frac{r_i \cdot r_k}{|r_i||r_k|} = \frac{\sum_{j=1}^{m} r_{ij} r_{kj}}{\sqrt{\sum_{j=1}^{m} r_{ij}^2 \sum_{j=1}^{m} r_{kj}^2}} \tag{1}$$

The reason for using this algorithm despite its shortcomings is because of the limited size of the dataset. Cosine similarity works well in limited size dataset. Also, the assumption is that both the vectors having non-null values. With the growth in size of the dataset, the processing power progressively worsens. Also, the algorithm does not provide accurate results if no nearby projected items are found. Cosine similarity method works with matrix projections of given data objects. The similarity measures ranges between − 1 and + 1, where − 1 is completely dissimilar and + 1 is completely similar. The closer the items are in the projection, the similar they are. The next section discusses the related work on intelligent product dispenser system.

2 Related Work

Eriyeti Murena et al. [2] proposed a control system design for a vending machine. The proposed system handled coin or mobile payment methods. The control system is used for dispensing snacks, newspapers, beverages, and smoking cigarettes. It uses a digital touchscreen–user interface. It also provides refrigeration system by continuously controlling the speed of compressor and fails in recovering after the failure. No intelligent suggestions provided. Wejdan Alsaeed et al. [3] proposed a mobile application to manage the various services in an organization, namely restaurants. It uses the geographical information system to handle the services in a better way. It computes the arrival time more accurately. The main drawback of this system is that it cannot be generalized for any organization or shopping malls. Hariharan Ramalingam et al. [4] have discussed the Internet of things-based application usage in the banking domain. The paper discusses the various challenges involved in developing automated, intelligent shopping, or payment system. The challenges include data privacy, data security risk, and data density. Any automated, intelligent IoT-based system must take care about these three parameters. Kaveh Azadeh et al. [5] presented an overview of the recent trends in automated warehousing by using robotic techniques and intelligent systems. The study shown that the item pick process in a shopping mall with intelligent suggestions have a major effect on the growth of these shopping malls. The proposed system, hence, come up with a study on an intelligent suggestion system. Katerina Berezina et al. [6] have reviewed various applications of robots, artificial intelligence, and service automation in various service organizations, namely in restaurants. The study shows that artificial intelligence-powered devices in the shopping malls increase the personalization of customers and thereby increasing the business. Radoslaw Klimek [7] proposed an intelligent queue management system to serve individual customers. It uses IoT sensor camera network device

system which classifies the customers as elders, people in a hurry, and many more. Accordingly, the system handles the queue. The system is tested using a simulator and not in a real-time scenario. Adilakshmi Satya Sri et al. [8] designed an automated product dispenser system for five specific products. It also enables the customer for the payment. The proposed method does not provide the intelligent recommendations. Monika V N et al. [9] proposed an automated medicine box vending machine. It uses ARM microcontroller, relay, GSM, and RFID technology to accomplish the same. No intelligent recommendation is provided. Chandrasekaran N. et al. [10] discuss how blockchain, artificial intelligence, machine learning and deep learning have helped in increasing the business in various fields. It also helps in performing better predictive analysis. Jinlong Wang et al. [11] proposed a new warehouse integration technology to improve automation in tobacco industry. Authors used IoT on cloud platform. This improved business in tobacco industry and reduced the overall cost. This work is restricted to only the cigarettes and tobacco products. Abhaya Kumar Sahoo et al. [12] proposed a recommendation system based on individual's health profile. This increases the sale of items there by increasing the business. The drawback is to collect the health profile of users as most of hospitals do not share the patient details.

By analyzing the various research articles, we have proposed a system which combines both automatic billing of items and intelligent recommendation for the customers. The proposed system is discussed in the next section.

3 Proposed System

In attempt to connect the consumers with the right options and exclusive discounts based on their previous shopping orders, the proposed system presented users with the latest and highly requested items using the collaborative filtering algorithm, which returns the most often ordered goods depending on the users' purchase history. The proposed system uses various technology to provide the end users with a quick and secure solution.

3.1 System Control Flow

To start with, the system has built a very simple and important interface for the consumers for quick access and to enjoy all the items at the comfort of their homes. To encourage administrators to view both data and related material, a dedicated platform to handle the users' records and documentation has been deployed. Usage of artificial intelligence has been introduced in order to include real-time suggestion framework. Components such as microcontrollers, RFID, LCD, keypad, motors, and several more are at the hardware end to create the shelf. The system also provides a

Fig. 1 Proposed system

personalized IoT-based system in order to allow users access from anywhere. Figure 1 shows the overall flow of the proposed system.

The accompanying procedure explains how the proposed system process operates using various methods. It is a machine bootstrapped with multiple procedures. The various steps are explained below:

Step 1: The required client registers with the Play Store application.

Step 2: The user will be provided with an RFID after the registration process, which allows the system to authenticate the users at any shopping outlet.

Step 3: The consumer can visit the nearest outlet when he / she wants to go and buy goods.

Step 4: Consumers use LCD and keypad to insert their secondary authorization key in order to obtain access to the shelf at the shelf panel in a shopping mall.

Step 5: The customer picks his/her items of their choosing and then clicks on purchase.

Step 6: The user will be asked for the amount on his chosen UPI Payment App until transaction is completed.

Step 7: After conclusion of payment, all goods should be collected by the customer.

Step 8: All items ordered will now be added to the Users App under the group My Orders.

Step 9: The user can also review his/her other information in the app for even richer device experience.

3.2 Recommendation Algorithm

The proposed system essentially aims to predict the "preference" that a consumer will offer to an object. It is seen in a number of industries. The suggestion method in this work indicates when the customer buys peach, which belongs to the fruit category; it advises that the customer buy watermelon which belongs to the same fruit category. Also, customers will score the goods on recommendation. The matrix factorization finally gives us the details about how much a person aligns with a collection of latent characteristics and how much the interest blends with this collection of latent characteristics. The difference of the closest regular neighborhood is that the correlation between them will be identified if they have the same underlying interests or not.

$$min_{q,p} \sum \left(r_{ui} - p_u^T q_i\right)^2 + \lambda\left(p_u^2 + q_i^2\right) \tag{2}$$

Given, the characteristics of the system, matrix factorization (in Eq. 2) provides the output with minimum displacement measure.

3.3 System Architecture

A description of the architecture shows how the concept works. Figure 2 shows the high-level architecture. The user is logging in through the application. The customer will only have access to one account. The app has a page to log in. Users are defined via RFID reader and name. The tag transfers information which is stored in the memory of the device. All tags have their own unique identification number which reuses these tags. The tag creates electromagnetic field in the reader. As a tag moves into the area, the reader interprets the information that is contained in the tag and sends it to the storage server, which in turn stores or retrieves label return information. The approved key which contains four digits is given to the user. The client or the owner should be able to enter the shelf and order the items using approved key. The shelf contains helical spring or coil spring that holds the product in the shelf. When removed this spiral spring returns to its original position. The keypad is used here to enter the quantity as well as the type of products that can be selected. There is also a part of the advice in this. It recommends the product in recommendation part. For example, when the customer buys orange it belongs to the category of fruit so it suggests another fruit in the same category. After scanning the customer card-id tag and product tag, it shows customer and product details taken. It can then perform take/retake operations.

Fig. 2 High-level architecture

The project consists of two modules in it. Figure 3 gives the concept design of the system. They are software and hardware part. In software part, it includes mobile application/Web application, recommendation system, and in hardware, it includes RFID, Arduino micro-controller, LCD, and keypad. The software part includes:

Fig. 3 Concept design

- **Customer:** The customer is allowed to buy the items through the app. The customer can only have one account.
- **Application:** The app contains a login tab. It consists of information about the user/customer, such as name, sex, phone no, address, and gender. It is planned to run on desktop, laptop, or computer handheld devices. It can be downloaded from the mobile operating system owner, namely Google Play Store. On many devices, users can access Web sites such as desktops, laptops, tablets, and smart phones. Customers are required to give the feedback in terms of scores to the goods/items they have purchased.

- The hardware components used are discussed below:
- RFID: RFID consists of TAG and RFID Reader. RFID Reader works with the tags to receive information that identifies them. Within them, RFID Tag has a magnetic coil which generates radio frequency waves. User is detected in this project by RFID giving us the advantage of identifying thefts and unauthorized movement of items.
- Arduino microcontroller: It is an open-source hardware and software project developing and manufacturing single-board microcontrollers and microcontroller kits for digital computer construction.
- LCD: Liquid crystal display (LCD) is a flat panel display which uses the liquid crystal characteristics combined with polarizers. It is programmable with ease.

3.4 Recommendation System

The recommendation system consists of two techniques, namely content-based recommendation and the second one is collaborative filtering recommendation system. Content-based solution involves a fair deal of knowledge on the correct aspects of the objects, instead of using experiences and input from the consumers. For instance, it may be object attributes such as category, year, manufacturer, price, or other important aspects using natural language processing. It gains knowledge from the user feedback written in natural language, namely English. Collaborative filtering, on the contrary, requires little more than the historical preference of users over a collection of objects. It uses historical evidence and more information gathered from the users. Since it is based on historical evidence, the main assumption here is that in the future the people who have agreed in the past will continue to agree. The pseudocode for the recommendation system is given below.

Recommendation System Pseudocode:

Step1:

indices = pd.Series(metadata.index,index = metdata['prod_name']).drop_duplicates().

Here, indices represent how the system links to the cosine similarity matrix.

idx = indices[name].

Here, idx will have indices only of product name.

Step 2:

Get the pairwise similarity scores of all products with that product.

sim_scores = list(enumerate(cosine_sim([idx])).

Here, we have enumerated the cosine similarity scores along with the indices. So, we will link the indices 0,1,2,3 that is our products along with the cosine similarity scores. So that we know that this product has these number of scores. And we store these as a single list.

Step 3:

Now sort the products based on the similarity scores. This step is needed because whatever scores are generated, they need not always be in the same order.

sim_scores = sorted(sim_scores, key = lambda x:x[1], reverse = True.

Here, we have given reverse = True, which means, we want to sort the score in the descending order, that is, to fetch the top highest similarity scores.

Step 4:

Get the scores of the five most similar products.

sim_scores = sim_scores[1:6].

Get the product indices.

prod_indices = [i[0] for i in sim_scores].

Return the top 10 most similar product.

return metadata['prod_name'].iloc[prod_indices].

Now if the system prints the similarity scores, the customer will not understand anything. So, the proposed method links these scores to the product name. Metadata is the original dataset that the proposed system has imported. Iloc will identify the row of index and print the product name.

The experiments were conducted by considering both the approaches, and the results obtained are discussed in the next section.

4 Results and Discussion

The proposed system has been designed to provide easy and fast checkout methods which incorporate cheaper processes and components. Based on the outcomes recorded, the system functions are expected with financial efficiency of up to 20% less than that of conventional dispensing systems. Also, the proposed system involves recommendation algorithms, which have proven to reduce shopping time by 7% on an average and at the same time increase sales of related items. Using one such algorithm "cosine similarity," we see that the recommendations always remain on the same Euclidean level hence proving to be accurate. Figures 4 and 5 show the results of collaborative-based filtering recommendation and content-based filtering recommendation, respectively.

Fig. 4 Collaborative-based filtering recommendation

Fig. 5 Content-based filtering recommendation

The red dots indicate the cosine similarity scores. In collaborative filtering (Fig. 4), the points are joined together which are very close to each other. They are on the same level between the range 15 and 20. But if we observe the graph in Fig. 5, that is, content-based filtering, the cosine similarity scores range from 0 to 23. So, the recommendation system can choose any point in the given slot. The results are

analyzed by only considering the features from the dataset. Collaborative filtering predicts more accurately because it has a lot of data.

Figure 6 shows the results of probability distribution of metadata vs features. The X-axis represents the number of products, and the Y-axis represents the probability. The red dots represent the probability of "features" being matched. The blue dots represent the probability of "metadata," respectively. The overlapping of red and blue dots represents the matching of both metadata and features. The graph is obtained when calculating probabilities, the product features appearing through the product names one after the other in the results. The basic content filtering dataset is indicated by the red dots that uses only the features column from the dataset and as you see they are more scattered. The scattering thus leads to anomalies in the output. The blue dots indicate the metadata used for collaborative filtering which not only includes features but also other factors such as cost, weight, category, and contents. The blue dots are continuous and have consecutive data points in the state space. The more data you feed, more accurate will be your prediction.

Talking about the accuracy of the model, used for recommendations, the easiest algorithm that comes to mind is content-based recommendation/filtering. As we have seen in the previous distribution, only using features/tags or names do not suffice in fast and accurate recommendation of items. Hence, an updated and more accurate algorithm is used, namely collaborative filtering. Collaborative filtering uses past data of a single data producer and generates a hypothesis that this attribute is true for all instances of the data point. Hence, the next time the same data point produces some data, and it is highly likely to choose from its previous dataset or to choose something near to it. Collaborative accuracy (Fig. 4) depicted by yellow increases gradually with the continuous generation of new data objects. The more the

Fig. 6 Probability distribution of metadata vs. features

availability of data, more accurate the prediction. On the other hand, content-based filtering provides a short-time spurt of accuracy, but clearly dies down soon enough.

5 Conclusion and Future Scope

The proposed system enables customers to bypass the queues and confusion by converting their mobile into a point of service (PoS). This reduces the burden of customers to wait in queues and checkouts by turning their mobile into a point of service (PoS). The proposed system saves time and human labor, which if well applied, has the power to transform the lifestyle of the people forever. It is an evolution from the self-checkout version, which is being seen today in several supermarkets. The cost of introducing, upgrading, and improving the program will be balanced by raising checkout clerks to surpass cost-savings. The proposed system can install these devices at a cheaper rate for those that are unable to buy specific things, such as food or clothes or other items purchased on a regular basis. This helps the people who can buy it and do not wish to use it. They can deposit into this system and obtain other incentives that they can use instead. Such items/goods can be offered to disadvantaged or needy citizens to support them. More intelligent and sophisticated customer feedback analysis program, to increase product service that would help provide customers with a better user experience, is an open research problem.

References

1. M.I. Younis, SLMS: a smart library management system based on an RFID technology. Int. J. Reason.-Based Intell. Syst. **4**(4), 186–191 (2012)
2. E. Murena, et al., Design of a control system for a vending machine, in Procedia CIRP 91, 758–763 (2020)
3. W. Alsaeed, K. Alhazmi, An ıntelligent spatial-based queue management system, 2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE, 2019
4. H. Ramalingam, V. Prasanna Venkatesan, Conceptual analysis of Internet of Things use cases in Banking domain, TENCON 2019–2019 IEEE Region 10 Conference (TENCON). IEEE, 2019
5. K. Azadeh, R. De Koster, D. Roy, Robotized and automated warehouse systems: Review and recent developments. Transportation Sci. 53(4), 917–945 (2019)
6. K. Berezina, O. Ciftci, C. Cobanoglu, Robots, artificial intelligence, and service automation in restaurants, in *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and Hospitality* (Emerald Publishing Limited, 2019)
7. R. Klimek, Context-aware and pro-active queue management systems in intelligent environments, in 2017 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2017
8. U. Sri, A. Lakshmi Satya, et al., Design of Automatic 5× 5 Dispenser with three kinds of payment modes using Verilog HDL, in 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2020
9. K.M., Hema, M.M. Savitha MM, An IOT based ıntelligent medicine box using vending machine-medical ATM, Institute of Scholars (InSc) (2019)

10. N. Chandrasekaran, N., et al., Digital transformation from leveraging blockchain technology, artificial intelligence, machine learning and deep learning. İn *Information Systems Design and Intelligent Applications* (Springer, Singapore, 2019), pp. 271–283
11. J. Wang, Z. Li, R. Higgs, The application of dense storage ıntegration technology in Tobacco logistics centers based on the ınternet of things, 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). Vol. 1. IEEE, 2017
12. A.K. Sahoo, et al., Intelligence-based health recommendation system using big data analytics, in *Big Data Analytics for İntelligent Healthcare Management* (Academic Press, 2019), pp. 227–246

Models Predicting PM 2.5 Concentrations—A Review

Anusha Anchan, B. Shabari Shedthi, and G. R. Manasa

Abstract With accelerated seriousness of air contamination, particulate matter with diameter 2.5 μm (PM2.5) is considered as the most significant air pollution. It has adversely impacted on individual's health, life and work and has made damage to individuals' well-being. Quick urban development is causing PM 2.5 contamination. Logical and viable forecast of PM2.5 can empower individuals to play it safe ahead of time to stay away from or diminish damage to the creatures on earth. In this manner, the prediction of PM2.5 has become a subject of extraordinary pragmatic essentialness. This examination assessed the forecast of fine particulate matters and summed up the aftereffects of foreseeing models. This review lights up two categories of models, i.e., machine learning and statistical/mathematical models. In addition, forecasting can essentially be enhanced by joining meteorological factors with ground measured parameters of every locale, alongside aerosol optical depth (AOD) rather than just considering meteorological factors. This work suggests the use of high-level AOD information, and future further works could be centered around creating satellite-related foreseeing models for forecasting PM 2.5 and other air contamination.

Keywords Autoregression · Linear support vector machine · Wavelet transform · XGBoost

1 Introduction

In this era of raising technology and globalization, air contamination has become a genuine danger for personage well-being and communal or public government assistance in ongoing many years as well as it creates negative impact on individual's health life [1]. With the improvement of financial aspects on the planet, the air contamination issue has become increasingly more genuine lately, particularly the particles of diameter under 2.5 μm (PM2.5) [2]. The estimation unit of PM 2.5 is

A. Anchan (✉) · B. S. Shedthi · G. R. Manasa
NMAM Institute of Technology (Visvesvaraya Technological University), Nitte, Karnataka, India
e-mail: anusha@nitte.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_6

by micrograms per cubic meter (μg/m 3). Increased concentrations of PM 2.5 have become a worldwide issue. Different investigations have shown that fine airborne particles have the most grounded well-being impacts. Also, airborne particles are vital in influencing barometrically radiation, cloud development, climatic photo-chemical responses and light elimination impacts that impact worldwide climate changes [3, 4, 41]. The impacts of PM 2.5 on general well-being have been all around reported in the writing.

The current patterns for particulate observing procedures will, in general, screen PM 2.5 in view of the immediate relationship with well-being impacts and the avoidance of regular particulate obstruction. Industrial exercises with high essential particulate outflows, for example, coal, concrete, cement or mining, significantly affect air quality because of their concentrated particulate discharges in 2.5–10 μm extend. Utilization of firewood and charcoal for warming and cooking is unavoidable issue, which is likewise adding to PM 2.5 outflows.

This gives the inspiration to exploring the strategies for assessment of particulate issue in territories having less screens at a unique scale, a methodology presently drawing in extensive scholarly intrigue. The focus of this review is as follows:

- Classify and stipulate frequently used methodologies which assesses the PM 2.5 concentration.
- Study the impact of input features considered with various algorithms.
- Find what are the possible performance metrics on which various models' performances are compared.
- Overview of various prediction modeling method recommendation.

2 Objectives of Various Research Articles

This review covers the studies which were carried out with following objectives: prediction of PM2.5 concentration (hourly/day a head), feature importance for PM 2.5 prediction, assessment probability of PM2.5 surpassing a predefined secure limit and sway of meteorological features up on PM 2.5. Most of study had several algorithms selected for comparison as to decide which among them performed best in predicting the specified contamination. This study has carried out by considering papers published 2010 onwards. In this paper, all the study regarding concerned topic has considered data acquired between 2010 and 2020. Table 1 targets featuring the basic standards of machine leaning methods and about their part in improving the expectation execution. Table 2 shows the consolidated study of statistical models/time series model/chemical models.

Table 1 ML models used for PM2.5 prediction

	Objective w.r.t to PM2.5	Region	Data collection duration	Features/parameters	Algorithms	Performance metrics	Best Model as result of comparison
[5]	Prediction (hourly)	Tianjin city, China	Dec 2016	Hourly PM10, SO_2, NO_2, CO, O_3	XGBoost RF MLR DTR SVM	RMSE, MAE R2	XGBoost RMSE = 17.298 MAE = 11.774 R2 = 0.9520
[6]	Daily estimation	USA	2000–2015	Land-use variables Satellite data. Meteorological variables	NN RF Gradient boosting Ensemble model NN + RF_XGBoost	R2, RMSE	Ensemble model NN + RF_XGBoost R 2 = 0.89 RMSE = 1.26
[8]	Feature importance for PM 2.5 prediction	Tehran, the capital city of Iran	2015–2018	Satellite data, Meteorological data, Ground-measured PM 2.5, Geographical data (23 features)	RF XGBoost Deep learning	R2, MAE RMSE	Ranking of features XGBoost gives the leading performance with 0.81(R2), 0.894(R), 10.0(MAE), 3.62(RMSE)
[9]	Assessment probability of surpassing a pre-defined secure limit	Bangkok, Thailand, Kasetsart University	2015–2019	Meteorological data Air pollutants	Naïve Bays Logistic regression, RF NN	F1 score, confusion matrix	Random forest with 0.82 F1-score

(continued)

Table 1 (continued)

	Objective w.r.t to PM2.5	Region	Data collection duration	Features/parameters	Algorithms	Performance metrics	Best Model as result of comparison
[10]	Predicting and Classification	Quito, Ecuador, Cotocollao Belisario	2007 - 2013	Aerosol data, fine particle concentrations, meteorological data	Boosted Trees (Bts), L-SVM, NN, CGM	TPR/FNR MSE, ROC curves	AUC-BT(0.72) is larger than AUC-L-SVM(0.66) MSE of 22.1% for NN and 15.6% for CGM
[12]	(i)Level detection (ii)Prediction	UCI repository	-	Temperature, dew point, wind speed, PM2.5, pressure,	Logistic regression, autoregression (AR)	Mean accuracy Standard deviation, Mean Squared Error(MSE)	Mean Accuracy 0.998859 (LR) Standard deviation Accuracy 0.000612(LR) MSE 27.00(AR)
[19]	Short period prediction	Beijing	2015	AOD, Meteorological factors Gaseous pollutants	MLR	R2, RMSE	Regression model with yearly data R2(0.766) goodness-of-fit and R2 (0.875)
[21]	Prediction	Delhi	2016–2018	Meteorological data, pollutant parameters	SVM	Prediction graph	Close prediction to real value

(continued)

Table 1 (continued)

	Objective w.r.t to PM2.5	Region	Data collection duration	Features/parameters	Algorithms	Performance metrics	Best Model as result of comparison
[22]	Forecast	Pingtung station, Taiwan	2012–2017	Meteorological data—speed, wind, temperature air contamination data—CO_2, PM2.5, PM10, CO, NO, SO_2, HUl	LR, RF, XGBoost KNR, MLP, DT-CART	RMSE R2	XGBoost outperforms other models R2 = 0.9336 RMSE = 0.1302
[29]	Forecasted	Romania Stations Munchen,(Germany) Ploieste(Romania)	M	PM2.5	ANN, ANFIS	MAE, RMSE IA, R	ANN MAE = 1.9278, RMSE = 3.1931, IA = 0.9804, R = 0.9619
[30]	Estimation	55 countries	1997–2014	Remote sensing observations, meteorological observations, CO NO_2	Multivariate, non linear, nonparametric machine learning	R	R for each of the five training data sets is 0.96
[31]	Prediction	25 stations, Seoul, South Korea	2015–2018	Hourly pm2.5, pm10, meteorological features	LSTM, DAE	RMSE	LSTM with RMSE = 12.174
[33]	Forecast	Jiayuguan, Datong, Fushun, Qiqihar, Weinan, and Xuchang, China	2014–2019	PM2.5	WT, SAE, LSTM SAE based on LSTM	MAE	MAE based on SAE-LSTM is least compared to other models

(continued)

Table 1 (continued)

	Objective w.r.t to PM2.5	Region	Data collection duration	Features/parameters	Algorithms	Performance metrics	Best Model as result of comparison
[34]	Prediction	CPCB online station	–	CO, NO, NO_2, SO_2, O_3, VOC, PM2.5, meteorological parameters	SVM, ANN	RMSE, MAE,R	ANN RMSE = 1.51(compared to SVM)

Methods XGBoost, random forest(RF), multiple linear regression(MLR), decision tree regression(DTR), support vector machine(SVM), neural network (NN), linear support vector machine(L-SVM), convolutional generalization model (CGM), true positive rate(TPR), false negative rate(FNR), Area under curve (AUC), mean-squared error (MSE), autoregression (AR), linear regression (LR), K-neighbor regressor (KNR), multilayer perception (MLP), decision tree (DT), index of agreement (IA), artificial neural networks (ANNs), adaptive neuro-fuzzy inference DSystem (ANFIS), long short-term model (LSTM), deep auto-encoders (DAEs), wavelet transform(WT), stacked auto-encoder(SAE).

2.1 Machine Learning Models

Current examinations in the field of climate science and designing show that deterministic models scuffle to catch the connection between the convergence of environmental contamination and their emanation sources. The ongoing advances in factual displaying dependent on AI/ML approaches have developed an answer to tackle these issues. Input variable sort to a great extent influences the execution of an algorithm, in any case, it is yet to be known why an algorithm calculation is favored over the other for a specific undertaking.

2.2 Statistical Models/ Time Series Model/Chemical Models

See Table 2.

3 Region and Data

Almost 90 percent of study reviewed here used air pollution contamination data and meteorological features as their prime feature for the study they were intended to. Along with that, some studies have got benefit of having access to AOD data [8], geographical data[24] and satellite data. In paper [5], chemical transport model (CTM) expectations of PM 2.5 vertical profiles would help adjust AOD. Along with ground information, this examination additionally gained AOD item information from Aqua-Moderate Resolution Imaging Spectrometer (MODIS) 550 nm collection in paper [19]. In article [39], IRMS was used to even detect the fire emission.

4 Data Processing Method

In paper [8], standard normalization was applied on data before training. The resampling method such as synthetic minority over-examining technique (SMOTE) is utilized in [9], in order deal with an imbalanced data set. Linear interpolation technique was used in study [21], and data were normalized using min–max normalization. Authors in paper [22] used Fourier arrangement and spline multinomial approaches to deal with missing values. In papers [23, 25], spatial and temporal distribution is used. To choose the vital boundaries out of an enormous number of boundaries, the Akaike information criterion (AIC) was assessed in the paper [26]. Summer-time air quality overview, spatial variability and temporal variability were carried out in the work [34].

Table 2 Statistical/time series models/chemical models

	Objective	Region	Duration	Features	Algorithms/Methods	Performance metric	Outcome
[7]	Prediction	Beijing	2018	Meteorological data, Air pollutants	Time series models AR, AR-Kalman filter, IMM	MAE, MAPE, RMSE	IMM with best performance
[11]	Trend prediction, value prediction, hidden factors prediction	Shanghai Beijing Guangzhou,Chengdu	2012–2014	Department of US state air quality files	ARMA Model, SV Model SW Model	MSE	Better prediction related to PM 2.5 concentration(six hours)
[13]	Determination	Tehran, the capital of Iran	2006–2016	Spatial data, temporal data	SVR, GWR, ANN, NARX—external input	R2, RMSE	NARX gives best performance
[14]	Analyzing the impact of features	Fuling District, Chongqing	2016–2017	Air pollutants, AOD data, meteorological data	Mathematical model of BP neural network	RMSE	Effective correlation between AOD and PM2.5 is established by BP NN
[15]	Prediction	China	–	Hourly CO, NO_2, SO_2, PM10	ANN BP NN, genetic algorithm BP NN optimized by genetic algorithm	Average relative error and maximum relative error	NN optimized by genetic algorithm gives best performance
[16]	Forecast	Ningbo, Qingdao, Guiyang, Yinchuan	–	–	Novel multi-resolution ensemble model	MAE, MAPE, IA	Proposed ensemble model gives better performance

(continued)

Table 2 (continued)

	Objective	Region	Duration	Features	Algorithms/Methods	Performance metric	Outcome
[17]	Estimation (daily)	USA consisting of 48 adjoining states and Washington DC	2011	AOD data, meteorological data, land-use data, ground-truth PM 2:5 measurement data	Convolutional neural network (CNN)		CNN shows competitive accuracy with contrast to the lately built models Feature rank
[18]	Forecast		2010–2014	PM 2.5, meterological data and snowfall (Snow)	AE, Bi-LSTM	RMSE	Combination of AE AND Bi-LSTM reduced training time, improved prediction accuracy RMSE-3.091
[20]	Prediction	Wuhan, Hubei Province	-	SO2, NO2, O3, PM10, CO, meterological data	Modified Gaussian smoke plume model	RMSE	Successful simulation
[23]	Impact of meteorological conditions	Nagasaki Japan	2013	Meteorological data: temperature, humidity, wind speed, precipitation, PM 2.5	Linear analysis, Spearman corrlation analysis	R2	Precipitation negative correlation, temp positive correlation, wind speed and humidity threshold

(continued)

Table 2 (continued)

	Objective	Region	Duration	Features	Algorithms/Methods	Performance metric	Outcome
[24]	Contribution	Central Chile, Santiago	2012	PM10, PM 2.5, SO_2, NH_3 NO_2, CO, VOC(volatile organic compounds)	MATCH CTM	Average concentrations of pollutants in Unit: μg m –3	CTM simulates higher concentration inert components, compared to the winter period, elemental carbon, dust
[25]	Prediction [Chem v/s machine learning]	Shanghai, China	2015–2018	Meteorological measurement, air pollutants	GBDT/XGBoost, WRF-Chem model, Lasso regression model	MB, ME RMSE R3	XGBoost out performed with MB = 3.6 ME = 0.3 RMSE = 26.1 R = 0.6
[26]	Prediction	Chengdu Sichuan Basin	2013–2017	Meteorological data, Radiosonde meteorological data	GAM	AIC R2 F-value P-Value	Strongest predictors are **T,** wind speed, precipitation, trajectory direction, distance GAM achieved an R 2 of 0.73
[27]	Determine leading nodes	Taiwan	2015 -2017	PM 2.5, Synoptic weather patterns	BD-oriented SNA, PCA	Z-value , daily average concentration of PM2.5	BD-oriented SNA is reasonable means to determine PM2.5 leading nodes

(continued)

Table 2 (continued)

	Objective	Region	Duration	Features	Algorithms/Methods	Performance metric	Outcome
[28]	Hourly concentration study	Gansu and Lanzhou China	2015	PM 2.5, PM 10, NO_2 SO_2, O_3 CO, meteorological factors	Pearson's correlation coefficient (relationship between factors) multivariate analysis	Average concentration of pollutants, correlation coefficient	Wind speeds—negatively correlated, temperature—negatively correlated, humidity—positively correlated, vaporous, pollutants showed no big effect
[32]	Forecast	Shenzhen, China	2013	Meteorological factors, SO_2, NO_2, CO, O_3	ARIMA, SVM, Ensemble ARIMA + SVM, The hybrid-Garch forecasting model	MAE, RMSE, DA, IA	Precision level of half breed Garch model was steadily higher than that of ARIMA or SVM
[35]	Day-ahead forecast	Wuhan and Tianjin	2014–2016	PM 2.5	WT, DE, BP, DE + BP, Hybrid WT-VMD-DE-BP,VMD,	MAE MAPE, TIC, RMSE,	Hybrid WT-VMD-DE-BP decomposition technique is effective with MAE = 4.05 RMSE = 6.25 MAPE = 8.88 PIC = 0.05

(continued)

Table 2 (continued)

	Objective	Region	Duration	Features	Algorithms/Methods	Performance metric	Outcome
[36]	Forecast	Tehran	2013–2016	T, PBL, P, RH, PM 2.5	MART, DFNN, hybrid model based on LSTM	RMSE MAE	LSTM with lowest RMSE = 8.91 MAE = 6.21 (μg m –3)
[37]	Forecast [ML VS Geographical Model]	Tianjin Hebei, Beijing	–	Meteorological parameters, air pollutants	Air mass trajectory-based geographic model, wavelet transformation ANN, hybrid model trajectory-based geographic parameter—input to ANN	MAE, RMSE, IA	Accuracy of hybrid model was high
[38]	Forecast	Beijing, China	2016		ESN, PSO, PSR, Hybrid IPSO-PSR-ESN	RMSE, MAE, SMAPE	Hybrid IPSO-PSR-ESN has better performance
[39]	Estimation	China	1999–2002	Meteorological, fire AOD	GWR	RMSE APE	RMSE = 17.2 APE = 18.5

Methods Autoregressive moving average (ARMA) model, stochastic volatility (SV), model and Stock–Watson (SW) model, artificial neural network (ANN), auto-encoders (AE), bidirectional long short-term memory (Bi-LSTM) chemical transport model(CTM), multi-scale atmospheric transport and chemistry model (MATCH), volatile organic compounds(VOC), gradient boosting decision tree (GBDT), weather research and forecasting (WRF) model coupled with chemistry, generalized additive models (GAMs), social network analysis (SNA), big data (BD), wavelet transform (WT), variational mode decomposition (VMD), differential evolution-back propagation(DE-BP), planetary boundary layer height (PBL), multiple additive regression trees (MART), deep feedforward neural network (DFNN), particle swarm optimization (PSO), echo state network (ESN).
Metrics Mean bias (MB), mean error (ME), mean absolute error (MAE) μg/m 3 root mean square error (RMSE) μg/m 3, correlation coefficient, Akaike information criterion (AIC), Direction accuracy (DA), mean absolute percentage error (MAPE),Theil's inequality coefficient (TIC), symmetric mean absolute percentage error (SMAPE), geographically weighted regression (GWR), absolute percentage prediction error(APE), temperature (T), surface-level pressure (P).

5 Methodology/Algorithms

5.1 Neural Networks

ANNs are the apparatuses used for applying AI which numerically model the working of a human cerebrum, and it has been utilized to construct answers to various issues from different areas. These organizations are dependent on connected designs, preparing components which can be called as nodes. These nodes are regularly stacked in a completely associated arrangement of at least three layers specifically, input layer, hidden layer(s) and the output layer. In papers [5, 6, 10, 21, 22, 34, 37, 39], authors have considered ANN as one of the models. In paper [8], author has used variant of ANN, i.e., deep learning. The **ANFIS** uses fuzzy inference system (FIS)which is a hybrid architecture enriched with ANN features proposed by author in paper [29]. Deep or stacked auto-encoder (DAE) is one among the incredible neural organization designs, used in papers [31, 33]. In paper [17], CNN works by stacking a convolution three layer and a common completely associated layer. The convolution layer comprises nonlinear activation function and different weight-shared kernels. Two layers fundamentally perform nonlinear changes of the input so the portrayal of information indicators is more sensible to get PM2.5 estimated. Authors of paper [14, 15, 35] have used BP NN including at least one shrouded layers including one or more hidden layers which was used to connect AOD and PM2.5, and it reduces overfitting.

5.2 XGBoost and Random Forest

Extreme gradient boosting (XGBoost) is a triumphant ML library emanated from gradient boosting algorithm. It uses more regularized model formalization to have good hold on overfitting. Presented works in papers [5, 6, 8, 22, 25] utilize this model, and out of these five studies, four studies proved that this model is better compared to considered benchmark models. Random forest is emanated from decision trees which fall under the category of supervised ensemble learning methods. Works carried in papers [5, 6, 8, 9, 22] have utilized this model.

5.3 Support Vector Machines

SVM converts the input space into a multi-dimensional in character space; hence, it is used as a discriminative classifier technique. This model can be used for application of classification and regression. In papers [5, 10, 21, 32, 34], SVM was utilized for regressing required pollutant.

5.4 Decision Tree Regression and MLP Regression

Decision tree (DT) algorithm comes under umbrella of ML supervised learning models. In paper [23], DT was used to choose the property for the root node in each level. Multilayer Perception (MLP) belongs to a family of ANNs. The nodes in this model are named as neurons. MLP was utilized in paper presented [24].

5.5 LSTM

Recurrent neural networks (RNN) are emanated from DNN which had cyclic connection between nodes. This helps in persisting the information. Generally, inclination of gradient in the RNN results in gradient explode or shrink exponentially. One solution is to utilize the long short-term memory (LSTM). In paper [18], the presented forecast model comprises two layers: the auto-encoder layer and the Bi-LSTM layer. Mix of auto-encoder and Bi-LSTM models reduces the planning time and improves conjecture accuracy. In articles [31, 33], RNN is along with LSTM to establish a climate for the calculation cycle, weighted lattice, get include and make yield.

5.6 Time Series Model

It is a statistical model. This arrangement incorporates autoregressive (AR) model, moving normal (MA) model and autoregressive and moving normal (ARMA) model. In paper [12], autoregression is applied on time series information so as to anticipate the PM 2.5 which estimates seven days preceding the current date.

Autoregressive–Kalman Filter Method: In article [7], AR model was set up; afterward, AR model is changed into state condition for Kalman filtering. This work presents three models: stochastic volatility (SV) model, autoregressive moving average (ARMA) model and Stock–Watson (SW) Model. Brand new model is obtained joining SW model with time series architecture.

ARIMA: In article [32], ARIMA model and ensemble of ARIMA and SVM were experimented. These were compared with **hybrid Garch forecasting model**.

Interactive Multiple Model Approach: The intuitive numerous model (IMM), which utilizes at least two models, i.e., soft switching algorithm and effective weighted fusion was used in work carried out in [7]. The examination demonstrates that the intuitive numerous model (IMM) be up to anticipate more precisely than alone AR model or alone AR-Kalman model.

5.7 NAXRX, GWR, EZN

In paper [13], NARX strategy had least error plus the best assurance coefficient, trailed by SVR along with preferable execution over GWR and trailed by ANN. Paper [36] introduces air mass trajectory-based geographic model which is been a useful tool to find direction if pollutants as well as source location of them. The paper [38] proposed a new a hybrid model combining echo state network (ESN) and an improved particle swarm optimization (IPSO) algorithm. Phase space reconstruction (PSR), a time series model, constructs the group of coordinate components that characterized the behaviors of original system. In this study, IPSO-PSR-ESN hybrid method was introduced.

5.8 Weather Research and Forecasting Model with Chemistry (WRF-Chem)

The WRF-Chem is the model novel regional atmospheric dynamic chemical coupling model. Benefits are coupling of chemical transmission mode and meteorological mode in time resolution and spatial resolution [40]. This model was used by authors in the presented work in [25].

5.9 Wavelet Transform

WT is based on the idea of short-time Fourier transform localization. This method surpasses a problem of constant window with frequency. Furthermore, it acts as best apparatus to analyze the system and to process. Mallat algorithm is used for wavelet decomposition which is used in experiment in papers [33, 35].

5.10 Differential Evolution (DE)

Experiments in paper [35] used DE algorithm which is a stochastic, population-based algorithm and direct search algorithm, which has the characteristics of simple structure, less control, fast convergence, parameters along with significant advantages of dealing with convergence, and strong robustness and non-differentiable. The DE-BP model is introduced to improve the function approximation capacity of BP neural network, especially on the catastrophe points. The study describes the purpose of DE algorithm to optimize the weight matrices and thresholds. Hybrid WT-VMD-DE-BP forecasting model is established for daily PM 2.5 concentration.

6 Conclusion

The PM 2.5 accumulation of a specific point can be estimated by suitable methods with exact information; nevertheless, it is difficult to get values constituting to the restrictions over a full examination area where setting up the essential tools is not attainable. Hence, an expanding number of techniques for assessing PM 2.5 focuses in zones were introduced by the scientists/researchers. The primary methodologies are partitioned into two classes in this study, to be specific are (a) statistical/mathematical methods (b) ML techniques. Be that as it may, they normally overlap as far as source data and preprocessing are considered. Nonetheless, ML techniques have risen as the most famous procedure during late years, and it is worth an extraordinary notice. Presented work audited pertinent as of late distributed papers and just as old style papers in the field of techniques and utilization of PM 2.5 prediction and assessment. The primary discoveries include the following:

1. In light of the ascent in computing power, quick improvement in ML has become a significant examination center territory for assessing and anticipating issues through persistent procurement, joining and investigation of an assortment of heterogeneous and huge information in urban areas. Consolidating physical models with ML holds evident potential for assessing spatial-transient powerful dissemination of metropolitan PM 2.5 focuses.
2. The conventional methodologies for assessment of PM 2.5 focuses are through obsolete in light of the fact that they give insights about future possible options and development in study area. Moreover, there is an expanding number of integrated ensemble models and strategies applied to different restrictive applications.
3. The most advantageous and time-effective technique is using ML techniques. Accuracy of these strategies is generally high. The more intricate of the above strategies consistently identifies with ML which can anticipate the obscure information of PM 2.5 on spatiotemporal scales. There are numerous enhancements and reconciliations of various strategies which can give more precise outcomes under explicit conditions.
4. There are different degrees of exactness/accuracy, which are controlled by an enormous components including conditions explicit to exploration territory, goal of the source information, boundaries picked by explicit models, and the details all the while, utilized by various techniques to appraise the PM 2.5.

References

1. Y.F Xing, Y.H. Xu, M.H. Shi, Y.X. Lian, "The impact of PM2.5 on the human respiratory system". https://doi.org/10.3978/j.issn.2072-1439.2016.01.19, https://doi.org/10.3978/j.issn.2072-1439.2016.01.19.
2. Health effects of particulate matter, Policy implications for countries in eastern Europe, Caucasus and central Asia, World Health organization, PublicationsWHO Regional Office for EuropeUN City, Marmorvej 51DK-2100 Copenhagen Ø, Denmark, ISBN 978 92 890 00017
3. L. Miller, X. Xu, "Ambient PM 2.5 Human Health Effects—Findings in China and Research Directions", Atmosphere 2018, **9**, 424. https://doi.org/10.3390/atmos9110424m,MDPI
4. D. Zhu, C. Cai, T. Yang, X. Zhou, "A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization", big data and cognitive computing, MDPI, 24 Feb 2018.
5. B. Pan, "Application of XGBoost algorithm in hourly PM2.5 concentration prediction", ICAESEE 2017, IOP Publishing. https://doi.org/10.1088/1755-1315/113/1/012127
6. Q. Di , H. Amini, L. Shi, I. Kloog, R. Silvern, J. Kelly, MB. Sabath, C. Choirat, P. Koutrakis, A. Lyapustin, Y. Wang, LJ. Mickley, J. Schwartz, "An ensemble-based model of PM 2.5 concentration across the contiguous United States with high spatiotemporal resolution", Environment International. **130**, 104909 (2019), ELSEVIER
7. J. Li, X. Li, K. Wang, "Atmospheric PM 2.5 Concentration Prediction Based on Time Series and Interactive Multiple Model Approach", Hindawi Advances in Meteorology Volume 2019, Article ID 1279565, 11. https://doi.org/10.1155/2019/1279565
8. M. Zamani Joharestani, C. Cao, X. Ni, B. Bashir, S. Talebiesfandarani, "PM 2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data", atmosphere, MDPI, 4 July 2019.
9. J. Boonphun, C. Kaisornsawad, P. Wongchaisuwat, "Machine learning algorithms for predicting air pollutants", E3S Web of Conferences **120**, 0 30 0 4 (2019) CGEEE 2019, https://doi.org/10.1051/e3sconf/20191 200 3004
10. J.K. Deters, R. Zalakeviciute, M. Gonzalez, Y. Rybarczyk, "Modeling PM 2.5 Urban pollution using machine learning and selected meteorological parameters", Hindawi J. Electric. Comput. Eng. **2017**, Article ID 5106045, 14 https://doi.org/10.1155/2017/5106045
11. J. Shen, "PM 2.5 concentration prediction using times series based data mining"
12. C.R. Aditya, C.R. Deshmukh, D.K. Nayana, P.G. Vidyavastu, "Detection and prediction of air pollution using machine learning models", Int. J. Eng. Trends Technol. (IJETT). **59**(4) (May 2018)
13. M.R. Delavar, A. Gholami, G.R. Shiran, Y. Rashidi, G.R. Nakhaeizadeh, K. Fedra, S.H. Afshar, "A novel method for improving air pollution prediction based on machine learning approaches: a case study applied to the capital city of Tehran", Int. J. Geo-Inf. MDPI (2019)
14. Y. Chen, "Prediction algorithm of PM2.5 mass concentration based on adaptive BP neural network", Computing **100**, 825–838 (2018). https://doi.org/10.1007/s00607-018-0628-3. Crossmark
15. X. Wang, B. Wang, "Research on prediction of environmental aerosol and PM2.5 based on artificial neural network". Neural Comput. Appl. **31**, 8217–8227 (2019). https://doi.org/10.1007/s00521-018-3861-y. Crossmark
16. H. Liu, Z. Duan, C. Chen, "A hybrid multi-resolution multi-objective ensemble model and its application for forecasting of daily PM2.5 concentrations", Inf. Sci. Elsevier (2019)
17. Y. Park, B. Kwon, J. Heo, X. Hu, Y. Liu, T. Moon, "Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks", Environmental Pollution, journal homepage: www.elsevier.com/locate/envpol, Elsevier (2019)
18. B. Zhang, H. Zhang, G. Zhao, J. Lian, "Constructing a PM 2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks", Environ. Model. Softw. J. Homepage: http://www.elsevier.com/locate/envsoft (2020)

19. R. Zhao, X. Gu, B. Xue, J. Zhang, W. Ren, "Short period PM 2.5 prediction based on multivariate linear regression model", PLOS ONE | July 26, 2018. https://doi.org/10.1371/journal.pone.0201011
20. P. He, B. Zheng, J. Zheng, "Urban PM 2.5 diffusion analysis based on the improved gaussian smoke plume model and support vector machine", Aerosol Air Qual. Res. **18**, 3177–3186 (2018), ISSN: 1680–8584 print / 2071–1409 online. https://doi.org/10.4209/aaqr.2017.06.0223
21. A. Masood, K. Ahmad, "A model for particulate matter (PM 2.5) prediction for Delhi based on machine learning approaches", Science Direct, Proc. Comput. Sci. **167**, 2101–2110 (2020), Elsevier
22. KS Harishkumar, KM Yogesh, I. Gad "Forecasting air pollution particulate matter (PM 2.5) using machine learning regression model", Procedia Comput. Sci. **00**, 000–000 (2019). ScienceDirect, Elsevier
23. J. Wang, S. Ogawa, "Effects of meteorological conditions on PM 2.5 concentrations in Nagasaki, Japan", Int. J. Environ. Res. Public Health, ISSN 1660–4601 www.mdpi.com/journal/ijerph, **12**, 9089–9101 (2015). https://doi.org/10.3390/ijerph120809089
24. J. Langner, L. Gidhagen, R. Bergström, E. Gramsch, P. Oyola, F. Reyes, D. Segersson, C. Aguilera, "Model-simulated source contributions to PM 2.5 in Santiago and the central region of Chile", Aerosol Air Qual. Res. **20**, 1111–1126 (2020), ISSN: 1680–8584 print / 2071–1409 online. https://doi.org/10.4209/aaqr.2019.08.0374
25. J. Ma, Z. Yu, Y. Qu, J. Xu, Y. Cao, "Application of the XGBoost machine learning method in PM 2.5 prediction: a case study of Shanghai", Aerosol Air Qual. Res. **20**, 128–138 (2020), ISSN: 1680–8584 print / 2071–1409 online. https://doi.org/10.4209/aaqr.2019.08.0408
26. Y. Zeng, D.A. Jaffe, X. Qiao, Y. Miao, Y. Tang, "Prediction of potentially high PM 2.5 concentrations in Chengdu, China", Aerosol Air Qual. Res. **20**, 956–965 (2020), ISSN: 1680–8584 print/2071–1409 online. https://doi.org/10.4209/aaqr.2019.11.0586
27. I. Cheng Chang, "Identifying leading nodes of PM 2.5 monitoring network in Taiwan with Big Data-oriented social network analysis", Aerosol Air Qual. Res. **19**, 2844–2864 (2019), ISSN: 1680–8584 print/2071–1409 online, https://doi.org/10.4209/aaqr.2019.11.0554
28. M. Filonchyk, V. Hurynovich, "A study of PM2.5 and PM10 concentrations in the atmosphere of large cities in Gansu Province, China, in summer period", Article in J. Earth Syst. Sci. (August 2016). https://www.researchgate.net/publication/306052138
29. M. Opera, Sanda Florentina Mihalache ad Maraian Popescu, "Computational intelligence-based PM2.5 air pollution forecasting", Int. J. Comput. Commun. Control, ISSN 1841–9836, **12**(3), 365–380, June 2017, Research Gate.
30. D.J. Lary, T. Lary, B. Sattler, "Using machine learning to estimate global PM2.5 for environmental health studies", Environ. Health Insights, **9**(s1), SAGE Publishing, https://doi.org/10.1177/EHI.S15664
31. T. Xayasouk, HwaMin Lee, G. Lee, "Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models." Sustainability **12**, 2570 (2020). https://doi.org/10.3390/su12062570. MDPI
32. P. Wang, H. Zhang, Z. Qin, G. Zhang, "A novel hybrid-Garch model based on ARIMA and SVM for PM 2.5 concentrations forecasting", http://www.journals.elsevier.com/locate/apr, Atmospheric Pollut. Res. **8**, 850e860 (2017)
33. W. Qiao, W. Tian, J. Zhang, "The forecasting of PM2.5 using a hybrid model based on wavelet transform and an improved deep learning algorithm", http://creativecommons.org/licenses/by/4.0/, **7** (2019)
34. J. Shah, B. Mishra, "Analytical equations based prediction approach for PM2.5 using artificial neural network", Springer Nature Switzerland AG 2020, SN Appl. Sci. **2**, 1516 (2020). https://doi.org/10.1007/s42452-020-03294-w
35. D. Wang, Y. Liu, H. Luo, C. Yue, S. Cheng, "Day-Ahead PM 2.5 concentration forecasting using WT-VMD based decomposition method and back propagation neural network improved by differential evolution". Int. J. Environ. Res. Public Health, **14**, 764 (2017). https://doi.org/10.3390/ijerph14070764, MDPI

36. H. Karimian, Q. Li, C. Wu, Y. Qi, Y. Mo, G. Chen, X. Zhang, S. Sachdeva, "Evaluation of different machine learning approaches to forecasting PM 2.5 mass concentrations", Aerosol Air Qual. Res. **19**, 1400–1410 (2019), ISSN: 1680–8584 print / 2071–1409 online. https://doi.org/10.4209/aaqr.2018.12.0450
37. X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, J. Wang, "Artificial neural networks forecasting of PM 2.5 pollution using air massct rajectory based geographic model and wavelet transformation". https://doi.org/10.1016/j.atmosenv.2015.02.030 1352–2310/© 2015 The Authors. Published by Elsevier Ltd
38. X. Xu, W. Ren, "Application of a hybrid model based on echo state network and improved particle swarm optimization in PM 2.5 concentration forecasting: a case study of Beijing, China", Sustainability, **11**, 3096 (2019). https://doi.org/10.3390/su11113096 www.mdpi.com/journal/sustainability.
39. W. You, Z. Zang, L. Zhang, Y. Li, X. Pan, W. Wang, "National-scale estimates of ground-level PM2.5 concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD", www.mdpi.com/journal/remotesensing, Remote Sens. **8**, 184 (2016) https://doi.org/10.3390/rs8030184
40. G. Zhang, X. Rui, Y. Fan, "Critical review of methods to estimate PM 2.5 concentrations within specified research region", Int. J. Geo-Inf., MDPI (2018)
41. N. Zhang, H. Huang, X. Duansd, J. Zhao, B. Su, "Quantitative association analysis between PM 2.5 concentration and factors on industry, energy, agriculture, and transportation", www.nature.com/scientificreports (2018)

Performance Analysis of Modified TCP New Reno for MANETs

Sharada U. Shenoy, Udaya Kumar K. Shenoy, and M. Sharmila Kumari

Abstract The transmission control protocol (TCP) controls packet loss due to congestion by additive increment (AI) and multiplicative decrement (MD) algorithm on congestion window size. This works fine in case of wired networks. In wireless ad hoc networks, the packet loss may not always be due to congestion, whereas most of the time the reason could be the problems in wireless medium. In case of path break, the routing algorithm will reestablish the route and packet loss may last for short duration. The MD phase of TCP algorithm reduces the congestion window by 50% on detection of packet loss. This reduces the transmission rate drastically. So instead, the congestion window can be reduced by a small fraction until the path gets resumed by a routing algorithm. So in this paper, a comparative study has been carried out by reducing the congestion window only by 20% instead of existing 50% during congestion in the MD phase of TCP New Reno for MANETs. This resulted in improved performance in terms PDR, throughput, and packet delivery rate compared to existing method.

Keywords MANET · TCP · Transmission control protocol · Congestion control · Congestion window · Modified TCP New Reno

1 Introduction

The transmission control protocol (TCP) congestion control algorithms work based on adjustment of congestion windows to cope with the congestion and packet losses. The reduction in congestion window size to 50% in case of packet loss decreases the data rate until the congestion situation is resolved. This is carried out irrespective of

S. U. Shenoy (✉) · U. K. K. Shenoy
NMAM Institute of Technology, Nitte, India
e-mail: sharadauday@nitte.edu.in

U. K. K. Shenoy
e-mail: ukshenoy@nitte.edu.in

M. S. Kumari
PA College of Engineering, Mangaluru, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_7

causes for packet loss. Especially in wireless networks, the packet loss is not always due to congestion [1]. The packet loss in wireless networks is usually due to the error-prone medium of transmission. Possible reasons may be path break, scattering, interference, multipath propagation, and weak signals unlike wired networks. In wired networks, packet losses are mostly due to congestion. Hence, it is possible to get fair results in terms of throughput, PDR, and packet delivery by modifying the congestion control mechanism used in traditional TCP. Especially in ad hoc networks where there is no infrastructure, reduction of congestion window by 50% causes severe degrading in data transmission. At the same time, the dynamic routing algorithm used in wireless networks will successfully find the alternative paths and regain the connection quite fast. Thus, the performance of TCP can be increased if we do not reduce the congestion window by a factor of 0.5 (50%) but reduce with a smaller fraction because packet loss in wireless networks is due to temporary problems while rerouting will be done by the routing algorithm quickly. This paper discusses the result of reducing the congestion by only 20% during packet loss and continuing with existing additive increment phase later on. This paper also provides a comparative study on the existing method of 50% window reduction.

The transmission delay and packet loss rate are determined by data link layer (DLL) parameters also, and they affect the TCP performance. Hence, the link errors and congestion need to be managed in a different way. The worse performance at the DLL and physical layer will primarily cause transmission errors like packet loss in the TCP at transport layer [2, 3, 4].

The traditional TCP which is designed for wired networks is the reliable protocol at large demand. But the situation in wireless networks is entirely different, and packet loss may happen because of reasons other than congestion, like mobility, path loss, signal fading, reflection, or weakening of the signal [1]. Traditional TCP senses any loss as congestion and shrinks the congestion window by usually 50% with an intention to decrease the network load. This continues if congestion persists or until timeouts. According to the additive increment and multiplicative decrement (AIMD) approach [1, 5], TCP at the connection setup initializes the congestion window (cwnd) to 1. Then, it starts with a slow start algorithm which does additive increment (AI) of the window size. Then, it continues in the AI phase until the slow start threshold (ssthresh) is reached. Then, increments the window linearly up to maximum limit on window size. The congestion window size gets decremented by 50% and in case of congestion till the lower limit being cwnd = 1. This technique is called as multiplicative decrement (MD) and works well with wired networks where the congestion is because of overflow in the router.

Most of the real-time traffic also uses TCP protocols in wireless. Considering the comparison of TCP New Reno and CUBIC, the study shows that New Reno recovers fast from single packet loss and suffers from slow start when multiple packets are lost. CUBIC is fair and stable, but data flow is lesser compared to New Reno [6]. The problem is with the window decrement strategy when congestion is detected and slow start during recovery. So few algorithms, receiver-assisted slow start (RAAS) and feedback-assisted recovery (FAR), proposed in [7] modify the initial window size according to the acceleration factor that is calculated based on the receiver advertised

window, and it follows slow growth of congestion window. It works based on three sizes of initial windows. This gives additional throughput of 37% and 52% against TCPNCE and New Jersey.

In another approach, a Markov chain model is defined for real-time delay distribution in TCP. It checks for real media data, voice and video streams and finds that there will be lesser TCP delays when RTT is 100 ms with error rate is less than 3% [8]. It is not only enough to change the window size by taking only the packet loss and need to consider other factors too in the wireless network. So feedback on the cause of loss can be taken from the underlying layers to enhance the performance of TCP. The queue length value is written to the advertised window (awnd) field of TCP in which the active window mechanism (AWM) in the access gateway [9]. The gateway informs the source to send controlled data with the help of this optimum value.

TCP should get a fair amount of bandwidth and also need to be stable while coexisting with other TCP protocols. Hence, the performance of TCP also needs to be examined when multiple flows coexist. The increment congestion window size in the additive increment (AI) algorithm phase of TCP can take feedback from the bottleneck node, while the multiplicative decrement (MD) is calculated separately by each source [9]. The AIMD window adjustment is carried by the window controller component at the source node. Similar work shows scaling of the window size for maintaining the traffic shape, and working of MD phase based on the RTT value is done by the traffic controller at bottleneck node [10]. Another approach takes the feedback from the bottleneck node, based on contention ratio (CR) and channel utilization (CU) [11]. This feedback regulates transmission rate of source nodes. In case there exist fluctuations in the channel due to signal fading or mobility, the cross layer (CL) approaches too can lead to enhanced results. The delay information-based fairness congestion control (DFCC) protocol that estimates congestion through the contention in the medium uses CL approach [12]. The packet delay is used by each source to determine average congestion window size. Similarly, modification is done to TCP Vegas where the transmission rate is based on the calculation of RTT [13]. The node increases the congestion window by 1, if the number of packets transmitted is greater than the limit, alpha. It decrements the congestion window by 1, if the limit is in between alpha and beta, else it remains constant. An active window that can interact with TCP source is also found to be useful [9] and implemented on a gateway to interact with TCP source. The router uses "awnd" field of TCP header to regulate the transmission rate of the source.

The protocols mentioned so far focus on window increase strategies and not on window decrease. The protocols that pay importance to window decrease strategies are mentioned below. Multiplicative increment multiplicative decrement (MIMD) [14] proposes a method where a constant time is taken to recover from error and it still depends on RTT, leading to overflow of bottleneck buffer causing severe packet loss. This affects stability and fairness. Alternatively, binary increase congestion control (BIC)-TCP [15], uses linear and logarithmic functions at MD phase. It decreases the window by β on loss. The maximum congestion window size, Wmax, will be retained to the previous size which was just before the start of packet loss. Minimum window

size, Wmin, will be set to the size that was reduced after packet loss detection. Later it sets the window size to the middle value of Wmax and Wmin, for every positive ACK. In case the distance of midpoint and Wmin is greater than the predefined maximum increment, Smax, it uses logarithmic increment. This works fine for the droptail queue in the routers. The compound TCP (CTCP)[16] uses a delay-based approach for AIMD. BIC-TCP and CTCP reduce the window by a factor of $\beta = 0.125$ because reducing it to 50% is not suitable. A quite different method is used in Westwood TCP [17], and it considers BDP (bandwidth delay product) instead of 50% reduction and is known as additive increment adaptive decrement (AIAD). İt adjusts the window to available BDP upon packet loss. In TCP Westwood + [18], which is advanced Westwood, the filtering approach is modified since the ACKs that are compressed and delayed were not allowing the Westwood to accurately measure the available BDP.

Few solutions are proposed to improve the transmission rates in case the packet loss is not due to congestion [19]. A TCP protocol introduced the explicit link notification (ELN) [20]. Here, the reason for the transmission error is conveyed by the receiver to the sender through ELN bit.

The issues of channel fading, transmission delay, and high BER cause the route rescheduling and retransmissions. Increased retransmission also causes congestion. So heaps of cross layers were introduced and are capable of identifying issues like channel attenuation, transmission delay, BER, or channel interference [3]. A forward error correction (FEC) that passes the information of error in DLL along with the ELN [21] is shown to enhance the performance by almost 100%. A revision of existing mechanisms in [21], ARQ along with FEC is proposed in [22]. Mapping based on SNR value is proposed in [22], and it conveys the errors in the physical layer / DLL to higher layers, and AMC is also used that has improved performance and QoS.

In MANETs, each node must work as a node as well as router. The existing congestion control method is not suitable since the frequent packet drop is a common scenario in wireless networks. In the existing approach, the rate of packets sent decreases and the available bandwidth is underutilized, thus resulting in reduced throughput. MANETs are popular, and it demands high data rates for data transfer like audio, video, live streaming, etc. Conventional TCP is not suitable because of low throughput due to its multiplicative decrement congestion management [15, 16]. If the window decrement mechanism of conventional TCP can be adapted for wireless scenarios, then this can provide better performance. So this paper has made an attempt to modify the MD stage of TCP New Reno and is able to get better performance for the given scenario.

2 Result Analysis for TCP New Reno

2.1 Performance Parameters for Existing TCP New Reno

This section provides analysis of TCP New Reno with existing method of congestion control.

Tables 1 and 2 show the performance parameters for the existing cwnd reduction of 50% for varying time over 50 and 100 nodes MANET. Table 3 shows the performance parameters for the existing cwnd of 50% by varying the speed over 50 nodes and 100 nodes mobile ad hoc network. In Tables 1 and 2, the average jitter is around 649 ms and ranges from 414.18 ms to 786.93 ms. The e2e delay ranges from 0.87 ms to 1.1217 ms. The average throughput varies from 545 to 555kbps. The PDR is in the range of 93.139% to 96.83%, throughput range of 465.3kbps to 596.96 kbps as seen in Tables 1 and 2.

With varying speed and constant simulation time, the packets sent is between 34,000 to 36,000 and average of 33,364 packets are sent as shown in Table 3.

Table 1 TCP New Reno performance metrics with existing approach over varying time for 50 nodes

TCP New Reno—Performance metrics for existing cwnd = 50%, 50 nodes						
Time (s)	Jitter (ms)	e2e delay (ms)	Throughput (kbps)	Number of packets sent	Number of packets received	PDR (%)
100	414.18	0.880579	510.74	5101	4598	90.1392
150	464.24	0.87843	485.93	10,597	10,086	95.1779
200	566.66	0.946554	516.89	14,159	13,518	95.4728
250	650.7	1.01921	554.01	17,930	17,296	96.464
300	729.55	1.11597	561.46	19,846	19,072	96.11
350	645.09	1.01547	564.02	23,097	22,095	95.662
400	631.36	1.04358	529.53	23,605	22,462	95.1578
450	730.33	1.12182	567	32,298	31,106	96.3094
500	710.67	1.03279	548	36,514	35,099	96.1248
750	769.74	1.11213	547.24	53,403	51,448	96.3392
1000	786.93	1.00143	570.79	68,001	65,820	96.233
1500	755.34	1.09226	590.43	111,725	107,762	96.4529

Table 2 TCP New Reno performance metrics with existing approach over varying time, 100 nodes

TCP New Reno—Performance metrics for existing cwnd = 50%, 100 nodes						
Time (s)	Jitter (ms)	e2e delay(ms)	Throughput (kbps)	Number of packets sent	Number of packets Received	PDR (%)
100	625.33	0.942108	465.3	5622	5390	95.8734
150	683.89	0.950464	519.36	9473	9173	96.8331
200	709.41	1.10076	515.03	13,068	12,026	92.0263
250	595.78	0.909859	550	17,133	16,431	95.9026
300	627.82	0.982212	547.16	20,692	19,687	95.1431
350	564.04	0.915938	554	24,782	23,785	95.9769
400	692.84	1.06943	572	28,559	27,594	96.621
450	646.53	0.980932	585.7	32,945	35,895	96.9009
500	607.51	0.952863	584.14	37,247	35,895	96.3702
750	664.23	0.942374	590.18	53,883	51,929	96.3736
1000	637.16	0.952118	596.95	73,251	71,119	96.178
1500	743.42	0.919559	584.78	100,521	96,730	96.2286

2.2 Result Analysis for Congestion Window Set to 80%—TCP New Reno Modified Window (MW

For the cwnd = 80%, the performance metrics have been plotted in Tables 4, 5 and 6. The e2e delay has been in the range of 0.637–1.081 ms as seen from the above-mentioned tables. This is less compared to existing approach and also the modified 60 and 70% approach. The jitter value starts from 285.42 ms to 779.24 ms and is less than the existing method as well as the 60% and 70% congestion window sizes.

The throughput has values ranging from 532.2kbps to 608.2kbps for varying simulation times from 100 to 1000 s and 570.87kbps to 614.75kbps for varying speed from 10 to 30 mph. This is high compared to the existing approach. The PDR is 95.31–97.4%, whereas the existing method of cwnd = 50% has 90.14–96.9% PDR. There is slight increase in PDR. If the packet count is considered, the number of sent and received packets is more than the existing approach as seen in Tables 4, 5 and 6 than seen in Tables 1, 2, and 3.

3 Improvement Over Existing Method

This section discusses an improvement seen by setting cwnd to 80% over the existing method.

Table 3 TCP New Reno—performance metrics existing approach for varying speed

TCP New Reno—Performance metrics for cwnd = 50% (existing) for varying speed												
Nodes	Throughput (kbps)		PDR (%)		e2e delay (ms)		Jitter (ms)		Packets sent		Packets received	
speed (mps)	50	100	50	100	50	100	50	100	50	100	50	100
10	541.73	567.75	94.4	95.22	1.07971	1.00382	676.99	704.4	33,941	35,945	32,040	34,226
15	562.2	579.67	95.91	96.09	1.00568	1.00783	715.84	639.02	35,031	34,620	33,596	34,006
20	569.87	575.52	95.9	96.2	0.99029	0.971939	656.53	649.92	32,927	35,353	31,578	34,017
25	575.54	573.96	95.13	95.03	0.970074	0.878402	633.68	573.96	35,100	35,248	33,391	33,497
30	571.59	572.58	95.18	91.91	0.874423	0.95093	566	644.41	34,842	36,053	33,163	33,135

Table 4 TCP New Reno-MW performance metrics with cwnd set to 80% over varying time for 50 nodes

TCP New Reno MW—Performance metrics for cwnd = 80%, 50 nodes						
Time (s)	Jitter (ms)	e2e delay (ms)	Throughput (kbps)	Number of packets sent	Number of packets received	PDR (%)
100	587.51	0.73292	529.62	5418	5182	95.64
150	555.45	0.83152	542.6	7743	7306	94.36
200	511.33	0.87593	532.2	13,143	12,664	96.36
250	703.22	0.89869	564.2	17,280	16,578	95.94
300	701.29	0.94256	577.6	21,441	20,804	97.03
350	692.36	0.95408	534.6	24,737	23,765	96.07
400	698.29	1.06347	592.2	29,515	28,568	96.80
450	733.16	1.05839	587	36,802	35,487	96.43
500	701.36	1.08121	597	37,365	36,128	96.69
750	741.97	1.00301	566.64	56,113	54,191	96.58
1000	710.23	1.00111	608.96	73,301	70,540	96.233
1500	721.26	1.00056	616.67	116,893	112,973	96.5

Table 5 TCP New Reno–MW performance metrics with cwnd set to 80% for varying time, 100 nodes

TCP New Reno MW—Performance metrics for cwnd = 80%, 100 nodes						
Time (s)	Jitter (ms)	E2e delay (ms)	Throughput (kbps)	Number of packets sent	Number of packets received	PDR (%)
100	285.42	0.67204	445.9	5323	5045	94.7774
150	497.85	0.78004	533.25	9035	8728	96.6021
200	505.06	0.79457	548.67	13,159	12,574	95.5544
250	586.2	0.97345	569	17,461	16,705	95.6704
300	541.06	0.91449	577.8	21,330	20,774	97.3933
350	615.37	0.96779	573.23	25,063	24,161	96.4011
400	633.53	1.01205	589	29,778	28,454	95.5538
450	708.43	1.08887	580.4	37,469	36,267	96.792
500	708.1	0.9516	567.9	57,006	54,874	96.26
750	636.83	0.63701	589.68	55,309	53,644	96.9896
1000	657.34	0.76981	585.49	76,133	73,477	97.0895
1500	733.12	0.78125	628.13	115,172	110,907	96.5234

Table 6 TCP New Reno MW—performance metrics with cwnd set to 80% for varying speed

TCP New Reno MW—Performance metrics for cwnd = 80%												
Nodes / speed (mps)	Throughput (kbps)		PDR (%)		e2e delay (ms)		Jitter (ms)		Packets sent		Packets received	
	50	100	50	100	50	100	50	100	50	100	50	100
10	614.75	587.68	96.64	96.48	1.04439	1.04077	751.33	635.45	38,447	36,059	37,154	34,787
15	595.75	570.87	95.91	96.21	1.06672	1.05404	727.81	721.49	37,512	36,432	35,106	33,885
20	599.9	602.94	96.46	96.33	1.00497	1.00969	680.71	681.47	37,926	37,715	36,582	36,328
25	571.32	596.79	95.62	96.1	1.02546	0.92486	746.23	644.9	37,854	37,773	36,193	36,243
30	592.45	594.22	95.31	95.98	1.01495	1.00614	779.24	755.07	37,002	37,395	35,263	35,888

Table 7 TCP New Reno MW—% improvement in packet transmission (%) for various simulation time

TCP New Reno MW—% improvement in packet transmission for varying time				
	50 nodes		100 nodes	
Time (s)	Improvement packets sent (%)	Improvement packets received (%)	Improvement packets sent (%)	Improvement packets received (%)
100	6.22	2.57	24.48	24.48
250	0.56	10.02	1.92	1.67
300	8.04	4.16	3.09	5.53
350	2.88	25.7	1.97	2.3
400	25.04	19.51	4.27	3.12
450	13.95	3.31	13.74	1.04
500	2.34	8.04	53.05	52.88
750	5. 3	9.18	6.68	6.48
1000	7.8	7.92	4.92	4.31
1500	4.63	2.56	14.58	14.66
Average	7.67	9.3	12.87	11.65

If Tables 1, 2, 4, and 1, 2, and 5 are considered, for different execution time, the throughput increase is seen in both cases of 50 and 100 node MANETs except at few case. There is an increase in average PDR of 1.214% for smaller and 0.51% for larger MANETs. Average of 5019 more packets are sent for 50 node and 4541 for 100 node MANETs. Jitter improvement 69.95 ms is seen for larger networks, and there is e2e delay improvement of 0.08 ms and 0.1 ms (Table 7).

From Tables 3 and 6 for varying speeds, there is no improvement in average jitter and e2e delay. The throughput has seen increment by 12.6%, and PDR has increased by 0.68% to 1.3%. There is 10.62% to 16% improvement which is seen in number of sent, 4.7% and 6.3% improvement in received packets when cwnd = 80% for 50 nodes and 100 node scenarios than the existing method as given in Table 8.

4 Conclusion

The work carried out shows that the performance can be improved by varying the congestion window of TCP New Reno to 80% will yield good performance in MANETs, and it is advantageous over traditional method. 80% congestion window size gives time to recover from problems like path breaks, temporary signal fading, or effect of mobility which will last for a short while. Further, it can continue to increment the window exponentially rather than again reverting back to a slow start phase.

Table 8 TCP New Reno MW—% improvement in packet transmission for various speeds

TCP New Reno MW—% Improvement in packets sent/received for varying speed				
	50 nodes		100 nodes	
Speed (mps)	Improvement in packets sent (%)	Improvement in packets received (%)	Improvement in packets sent (%)	Improvement in packets received (%)
10	13.28	15.96	0.32	1.64
15	7.09	6.54	5.24	6.37
20	15.18	15.85	6.69	6.83
25	7.85	8.4	7.17	8.2
30	6.2	6.4	3.73	8.31
Average	9.92	10.62	4.7	6.3

References

1. C. Siva Ram Murthym, B.S. Manoj, Adhoc wireless networks architectures and protocols, Pearson Education, 2004
2. S. Shakkottai, T.S. Rappaport, P.C. Karlsson, Cross-Layer Design for Wireless Networks. IEEE Commun. Mag. **41**(10), 74–80 (2003)
3. F. Foukalas, V. Gazis, N. Alonistioti, Cross-layer design proposals for wireless mobile networks: a survey and taxonomy. IEEE Commun. Surv. Tutorials **10**(1), 70–85 (2008)
4. P. Bender, CDMA/HDR: a bandwidth efficient high-speed wireless data service for nomadic users. IEEE Commun. Mag. **38**(7), 70–77 (2000)
5. Fu. Zhenghua, X. Meng, Lu. Songwu, How bad TCP can perform in mobile ad hoc networks. Proceedings ISCC Seventh International Symposium on Computers and Communications **2001**, 298–303 (2002)
6. M.C. Dias, L.A. Caldeira, A. Perkusich, Traditional TCP congestion control algorithms evaluation in wired-cum-wireless networks. International Wireless Communications and Mobile Computing Conference (IWCMC), Dubrovnik, 805–810 (2015)
7. M.J.A. Jude, S. Kuppuswami, Enhanced window increment and adaptive recovery TCP for multi-hop wireless networks. Electron. Lett. **53**(6), 438–440 (2017)
8. E. Brosh, S.A. Baset, V. Misra, D. Rubenstein, H. Schulzrinne, The delay friendliness of TCP for real-time traffic. IEEE/ACM Trans. Netw. **18**(5), 1478–1491 (October 2010)
9. M. Barbera, A. Lombardo, C. Panarello, G. Schembra, Queue stability analysis and performance evaluation of a TCP compliant window management mechanism. IEEE/ACM Trans. Netw. **18**(4), 1275–1288 (August 2010)
10. H. Nishiyama, N. Ansari, N. Kato, Wireless loss-tolerant congestion control protocol based on dynamic aimd theory. IEEE Wirel. Commun. **17**(2), 7–14 (April 2010)
11. X. Zhang, N. Li, W. Zhu, D.K. Sung, TCP transmission rate control mechanism based on channel utilization and contention ratio in AD hoc networks. IEEE Commun. Lett. **13**(4), 280–282 (April 2009)
12. H.J. Lee, J.T. Lim, Fair congestion control over wireless multihop networks. IET Commun. **6**(11), 1475–2148 (July 2012)
13. R.S. Cheng, D.J. Deng, H.C. Chao, Congestion control with dynamic threshold adaptation and cross-layer response for TCP over IEEE 80.11 wireless networks. Proceedings of the International Conference on Wireless Information Networks and Systems, Spain, 95–100 (2011)
14. T. Kelly, Scalable TCP: improving performance in highspeed wide area networks. ACM Comput. Commun. Rev. **33**, 83–88 (2003)

15. H. Bullot, R.L. Cottrell, R. Hughes-Jones, Evaluation of advanced TCP stacks on fast long-distance production networks. J. Grid Computing. **1**, 345–414 (2004)
16. K. Tan, J. Song, Q. Zhang, M. Sridharan, *A Compound TCP Approach for High-Speed and Long Distance Networks* (Microsoft Press, MSR-TR, 2005)
17. S. Mascolo, C. Casetti, M. Gerla, M. Sanadidi, R. Wang, TCP westwood: end-to-end bandwidth estimation of efficient transport over wired and wireless networks. Proceedings of the 7th annual international conference on Mobile computing and networking (MobiCom '01) ACM. USA, 287–297 (2001)
18. L.A. Grieco, S. Mascolo, Performance evaluation and comparison of Westwood+, New Reno, and Vegas tcp congestion control. ACM Comput. Commun. Rev. **34**, 25–13 (2004)
19. G. Xylomenos, G.C. Polyzos, P. Mahonen, M. Saaranen, TCP performance issues over wireless links. IEEE Commun. Mag. **39**(4), 52–58 (2001)
20. H. Balakrishnan, R. H. Katz, Explicit loss notification and wireless web performance. Proc. GLOBECOM Internet Mini-Conference. 1998
21. M. Miyoshi, M. Sugano, M. Murata, Improving TCP performance for wireless cellular networks by adaptive FEC combined with explicit loss notification. IEICE Communications Society: Transactions on Commun. Vol. E85-B(10), 2208–2214 (2002)
22. H. Zheng, H. Viswanathan, Optimizing TCP performance with hybrid ARQ and scheduler. Proc. 14th International Symposium on Personal, Indoor and Mobile Radio Commun. **2**, 1785–1789 (2003)

Smart Health and Pattern Recognition

Identification of Helmets on Motorcyclists and Seatbelt on Four-Wheeler Drivers

Divyansh Saini, Vedashree Arundekar, K. V. Priya, and Divya Jennifer D'Souza

Abstract The main objective of this research is to detect helmets and seatbelts on drivers in traffic scene. This will help people follow the rules and also make sure they drive safely. Its will also help the police monitor traffic from a safe place and away from elements such as heat and rain. Moreover, this project will help increase safety for both the commuters and the police. Using this data, police may also fine the individuals breaking the law. We intend to reprimand violators by making them aware of their violations by detecting features of vehicles driven by people who not only endanger themselves but also others on the road by their actions.

Keywords Helmet detection · Seatbelt detection · Machine learning · Image processing · Computer vision

1 Introduction

Traffic police personnel are always dependant on manual methods for intercepting traffic rule violators. Traditional methods are costly and often risky. Drivers especially Indians are obstinate toward wearing helmets and following traffic rules and the traffic police apply rudimentary methods to stop them like standing in front of the vehicle to forcefully stop them. Most of the drivers are adamant and apathetic, which leads the police to take impromptu and risky actions. Inclement weather is also one of the major challenges for manual traffic inspection by personnel. In this era of automation, an arrangement which requires less or no manual intervention would be really benevolent for governing authorities.

Traditional methods of identifying traffic violators include traffic police officers posted on various check posts or designated spots staying continuously alert and observing all incoming vehicles for presence of helmets/seatbelts on the drivers. In recent observations, traffic departments of various developed countries are automating this process by the use of CCTV or other medium to automate this

D. Saini (✉) · V. Arundekar · K. V. Priya · D. J. D'Souza
Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, Karnataka, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_8

process and directly levy fines on offenders with an attached image of the time and location of capture serving as a proof to the people who violate traffic safety norms. It is undeniable that it is physically tiring and demanding for the police officers to stay vigilant at all times. Our project aims to perish the stereotypes of being physically present for surveillance and panopticon purposes. With our suggested approach, one can easily identify people who are driving recklessly by not wearing helmets or seatbelts. In the coming times with advent in technology, automation of various processes would be benevolent for ease and avoidance of complications related to human intervene.

2 Literature Survey

Kunal Dahiya et al. proposed a framework for automatic detection of traffic rule violators in real time [1] who ride bike without using helmet. Their structure additionally helps the traffic police for distinguishing such violators in odd conditions such as heat and other such factors. Their experiment results exhibit the correctness of 98.88 and 93.80% for detecting bike riders with helmets and of violators, respectively. Likewise, their proposed structure naturally adjusts to new situations whenever required, with slight tuning. Their approach can be extended to identify and report number plates of violators.

In [2], Huiwen Guo et al. proposed a method for identifying seatbelts on drivers in traffic. To achieve the same, they have used gradient orientation method. To begin with, they detect the front window area and human face identification is used to limit the recognition area. If their face detection process fails, they use a straight-line fitting method and detect lines of 45°slope. Their accuracy is generally reliant on the consequence of human face recognition.

A system to give an idea about the number of traffic offenders was developed by Ramesh Babu et al. in [3]. They produced a database of all motorcycle riders driving helmetless with a proof image generated. They used open source software including tensorflow, OpenCV for reduction in costs. Provided adequate lighting, their system gave full proof and accurate results.

Vishnu proposed an efficient method for detecting people not wearing helmets, while riding bikes. In [4], they specified an algorithm for first segmenting and classifying vehicles, and then detection of offenders on motor bikes. For vehicle classification, they used various approaches for background calculation and object tracking, and achieved an accuracy of 0.9778. Their proposed system significantly low miss rates.

The papers [5] and [6] present novel methods for the detection of seat belt in a monitoring image which contains the full scene information of the moving car. First, the driver area is located based on the vehicle outline, then they use gradient orientation Image Processing Techniques.

3 Proposed Methodology

Below mentioned is the architecture we propose as a solution to the aforementioned problem statement.

A software solution to the incessant anguish of policeman from fatiguing manual surveillance of vehicles. The software we are developing is an application that enables one to detect traffic offenders who violate safety rules by not wearing helmets or seatbelts, while driving vehicles.

The proposed system contains two modules: one for helmet detection on motorcyclists and the other for seatbelt detection in four-wheeler drivers. There is an easily operable GUI which enables the user to choose from either of the options. The sequence diagram for the functioning of the project has been depicted in Fig. 1.

Once user has made a choice, control will be directed to the desired module and user will be asked to upload an image of the traffic scene. The image will be given as input to the selected module, which will further be processed and classification will be performed. Post classification, image will be resulted as a positive case, i.e., image with the presence of helmet/seatbelt, or as a negative class, i.e., absence of helmet/seatbelt.

Fig. 1 Sequence diagram for working of our project

Fig. 2 Helmet detected on subject

The user can then decide to notify the offenders once he has obtained results from our application.

(A) *Helmet Detection Using YOLO*

The model for detecting helmets on motorcyclists was made on YOLOv3 model [7]. YOLOv3 classifies images in one go and has short inference time. Thus, it was suitable to fit for our model. Additionally, Indian traffic contains a greater number of motorcycle riders than four-wheeler drivers. Hence, the model must be fast for identifying helmet and non-helmet wearers on motorcycle. A bounding box is made on the subject's head area if the confidence of detecting the presence of helmet is above 50%. This ensures that scarfs, masks and other objects resembling the helmet are not mistakenly classified as a helmet by the model. Figure 2 shows the helmet detected case with a bounding box depicting the accuracy of recognition.

Whereas Fig. 3 contains a subject not wearing a helmet and hence no bounding box drawn.

(B) *Seatbelt Detection Using Convolutional Neural Network*

The seatbelt detection is done using CNN (convolution neural network) initially by preparing image dataset of positive (pictures with seatbelt) and negative (pictures without seatbelt) images, which are pre-processed by subjecting it to various edge detection and thresholding algorithms; post which the image is loaded and resized to square image of 224*224 pixels to train the CNN using Tensorflow as backend. The model is created by developing four convolution layers with four max pooling layers in between. The last fully connected layer with software activation function is added to it in order to get probabilities of prediction. Lastly, the model is trained to evaluate and verify its accuracy and test result by giving the 'fit' and 'predict' command.

For detecting the seatbelt, we applied Canny edge detection algorithm first. Then, the same image was passed through Hough Transform. Now as specified by Huiwen

Fig. 3 No bounding box as helmet is not present on the subject

et al. in [2], seatbelt is usually at a specific angle. So, lines with slopes between 30 to 60 degrees are considered for being distinguished as seatbelts.

As shown in Fig. 4, the subject is wearing a seatbelt, and the same upon detection by the convolutional neural network is highlighted in red. A message indicating the presence of seatbelt is also printed in the console.

Fig. 4 Seatbelt correctly detected by our algorithm

Fig. 5 Seatbelt not present on the subject

Similarly, in Fig. 5 as there in the absence of seatbelt an appropriate message is printed in the console.

4 Results and Discussions

The purpose of our project was to automate the process of detection of traffic rule violators who jeopardize the safety of themselves as well as others by not wearing helmets or seatbelts as a mandatory precaution. Traffic police controllers risk their lives in an attempt to stop the speedy vehicles for interrogation and imposing fine on safety violators. Our project mainly aims at making the process computerized as feeds from CCTV or other sources in the form of video or photos can be processed and drivers not following rules can be identified. The UI of the said application is depicted in Fig. 6.

For this purpose, we have developed an application wherein we have added features to process images/frames and a bounding box is formed if a helmet is detected, whereas on detection of seatbelts highlighted lines are formed. For the same, we have implemented trained neural networks with a significant accuracy for detection. This project work will also help the authorities to minimize man power in order to observe the ongoing traffic.

Features from images were extracted and subjected to KNN Algorithm for comparison with our approach and the following results were obtained.

Our YOLO model shows an accuracy of 92.4% for helmet detection whereas KNN algorithm showed an accuracy of 87.7% (shown in Fig. 7). Similarly, for

Fig. 6 User interface of the application

Fig. 7 Comparison of YOLO and KNN for helmet detection

seatbelt detection our trained CNN showed 91.8% efficiency and KNN had 85.7% accuracy (Fig. 8).

The test images were manually clicked images for helmet detection whereas for seatbelt detection there was a combination of manually clicked photos and images obtained from Google. It proves that model is fit for real world application.

Fig. 8 Comparing CNN and KNN accuracy for seatbelt detection

5 Conclusion

To detect helmet on motorcyclists, we have proposed an approach which with given input image of a traffic scene, applies image processing and machine learning techniques. We have used techniques like Haar Cascade Classifier and LBP transform for face detection, YOLO Architecture for Helmet Detection [8], and Hough Transform and convolutional neural network for Seatbelt Detection. We analysed KNN classifier for comparative purposes and found that our method of operation provided better results. Wearing seatbelts and helmet, while driving car or motorcycles is an effective measure for safety and adequate attention should be given to it. The project helps in improving ways to have watch over traffic for police personnel and introduces a refined way for detection of offenders. It will also help the authorities to minimize man power in order to observe the ongoing traffic.

6 Future Work

The system that we have currently developed has multiple features that benefit the traffic police personnel in apprehending violators not following traffic safety rules. Further, we want to focus more on making the application more user-friendly and host the application on a server so that it can be accessed by authorities via a website too. The above project currently works on images and it needs to be transformed for operation on videos similar to the approach in [9] and [10]. We plan to focus on improving the overall performance of the system. License plate detection of offenders' vehicles [11] in front view is a task that we want to accomplish. The

creation of a database that links personal information of a vehicle owner with the vehicle registration number is also a feature we look forward to add and maintain. A system to notify the traffic violators along with photographic proof of location and time of their negligence in not using safety precautions such as helmets and seatbelts would also be a part of our development. We also intend on developing an android application for our model which is a better platform for interaction and each traffic personnel can have it installed in their phones for flexible use.

Acknowledgements We thank our mentor Ms. Divya Jennifer D'Souza for her valuable inputs and guidance. We also extend gratitude to Dr. Niranjan N Chiplunkar, Principal, NMAM Institute of technology and Dr. Uday Kumar Reddy, Head of department, NMAM Institute of technology for their support and giving us an opportunity to conduct this research project in college.

References

1. K. Dahiya, D. Singh, C.K. Mohan, "Automatic detection of bike-riders without helmet using surveillance videos in real-time," Visual Learning and Intelligence Group (VIGIL)
2. H. Guo, H. Lin, S. Zhang, S. Li, "Image-based seatbelt detection," Speech Commun., 978–1–4577–0577–9IEEE, pp. 161–164 (2011)
3. R. Babu, D.R. Amandeep Rathee, K. Kalita, M. Singh Deo, "Helmet detection on two-wheeler riders using machine learning," ARSSS Int. Conf. **41**(4), 603–623 (Sept 2018)
4. C. Vishnu, D. Singh, C. Krishna Mohan, S. Babu, "Detection of motorcyclists without helmet in videos using convolutional neural network", Elsevier, Visual Intelligence and Learning Group (VIGIL), Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad
5. B. Zhou, L. Chen, J. Tian, Z. Peng, "Learning-based seat belt detection in image using salient gradient". Ind. Electron. Appl. (ICIEA) 2017 12th IEEE Conference on, pp. 547–550 (2017)
6. D. Yu, H. Zheng, C. Liu, "Driver's seat belt detection in crossroad based on gradient orientation". Inf. Sci. Cloud Comput. Companion (ISCC-C) 2013 International Conference on, pp. 618–622 (2013)
7. J. Redmon, A. Farhadi, "YOLOv3: an incremental improvement", University of Washington, https://arxiv.org/abs/1804.02767
8. J. Wang, G. Zhu, S. Wu, C. Luo, Worker's helmet recognition and identity recognition based on deep learning. Open J. Model. Simul. **9**, 135–145. https://doi.org/10.4236/ojmsi.2021.92009 (2021)
9. L. Shine, C.V. J., Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN. Multimed Tools Appl. **79**, 14179–14199 (2020). https://doi.org/10.1007/s11042-020-08627-w
10. X.H. Qin, C. Cheng, G. Li, X. Zhou, "Efficient seat belt detection in a vehicle surveillance application", Ind. Electron. Appl. (ICIEA) 2014 IEEE 9th Conference on, pp. 1247–1250 (2014)
11. Y. Kulkarni, S. Bodkhe, A. Kamthe, A. Patil, "Automatic number plate recognition for motorcyclists riding without helmet", Current Trends Towards Converging Technologies (ICCTCT) 2018 International Conference on, pp. 1–6 (2018)

Prediction of Autism in Children with Down's Syndrome Using Machine Learning Algorithms

D. N. Disha, S Seema, and K. Aditya Shastry

Abstract Many human diseases that have a causative association with genetic components are defined as inherited diseases. It is a very difficult task to identify the different variants linked with inherited diseases using only lab-based technologies. One of the case studies of inherited diseases considered here is autism spectrum disorder. There are various methods are available to detect autism such as expensive screening methods that are used in well-equipped hospitals. But due to cost and also the screening timings, it is not able to efficiently detect the autism disease well in advance. Hence, it is best to utilize machine learning techniques to diagnose the disease well in advance and classify based on the severity using classification methods. One of the best ways to diagnose autism is through questionnaire method. Using machine learning methods, we concentrate on efficient detection of autism disease by using a real-time questionnaire.

Keywords Autism spectrum disorder · Machine learning · K-means clustering · Logistic regression · Support vector machine

1 Introduction

Inherited diseases are the type of disorders that need to be treated with the utmost care, and precision has to be maintained when it comes to its diagnosis. They directly affect the individual's behavior, the way they interact with the outside world, and the way they respond to situations. Autism is one among the several disorders that

D. N. Disha (✉) · K. A. Shastry
Department of ISE, NMIT, Bangalore, India
e-mail: disha.dn@nmit.ac.in

K. A. Shastry
e-mail: adityashastry.k@nmit.ac.in

S. Seema
Department of CSE, MSRIT, Bangalore, India
e-mail: seema.s@msrit.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_9

an individual can be diagnosed with and can often be confused with Asperger's syndrome and attention-deficit/hyperactivity disorder (ADHD).

There was an incident that took place in Bangalore, Karnataka, India, where an individual, who was actually autistic and was exhibiting the symptoms that fell into the spectrum. But the doctor could not diagnose him in a short span of time, and eventually, was misdiagnosed. The treatment toward the wrongly diagnosed disorder, therefore, did not help the individual in any way, and no improvement was seen in the individual. After this observation, another doctor was approached, where the individual was properly diagnosed with autism through expensive screening techniques. To make sure that this suffering should not be faced by anyone else and provide a cost-efficient, user-friendly solution to the issue, this was taken as a motive for the research.

1.1 Issues and Challenges

The major issues are that the symptoms of autism are not specific. ASD stands for autism spectrum disorder. The term 'spectrum' is used for the disorder as the symptoms vary along a very broad range; hence, prediction of autism becomes difficult. Also, as autism is an inherited disorder, the acquiring of the information needed for the dataset to train the algorithm is extremely difficult as the information is kept confidential by doctors and chooses to not disclose the information about the patient as they consider the data as highly sensitive. Selecting of symptoms in order to yield efficient diagnosis is completely analysis-based, and hence, it is time-consuming.

1.2 Objectives

- Identify behavioral patterns of the inherited diseases by collecting useful inferences from behavioral data.
- To create a budget-friendly, efficient method to detect autism and differentiate from other disorders with maximum accuracy using machine learning.
- To build a simple interface that can be used by anyone to collect data for initial screening purposes.
- To provide a budget-friendly, self-assessment method to parents to detect autism in an individual at initial stages before consulting a neurologist or a pediatrician.

1.3 Architecture of the Proposed Study

Schematic architecture is as follows. Initially, the behavioral data is collected and is tested with several machine learning algorithms. Based on the behavioral patterns detected, efficient prediction can be achieved (Fig. 1).

2 Literature Survey

Najwani Razali et al. [1] proposed that autism is identified by using motor movements of the child; the motor movements considered here are: imitating a finger tapping, clinching a hand. Dataset: Dataset ids are collected using EEG device (a device which is mainly used to record a brain signal) and multilayer perception (MLP) is used as classifier to discriminate between autistic and normal kid. Previous studies: Many researchers concentrated on finding the relationship between motor functions, cognition language development, and social communication disorder. Some concentrated on building a brief report about autism and impairment of motor function. Proposed method: Feature extraction is done by using Gaussian method and Fourier series.

Joaquin Rapela et al. [2] proposed that EOG technology is used as a therapy. Here, computer game is designed, and from the game, eye training will be given to the child. Based on the eye movement, speed and accuracy of eye movement and concentration, a disease can be detected.

Shyam Sundar Rajagopalan et al. [3], in this paper, algorithm is developed for identifying the self-stimulatory behaviors from the public dataset, which contains the video of a child which is tracked by selecting poslet bounding box prediction using the algorithm K-nearest neighbor.

Fig. 1 Schematic architecture

Histogram of dominant motion can be computed using dominant motion flow by using detected body regions. The motion model built using this descriptor can be used for detecting self-stimulatory behavior.

The proposed study uses postlets and is used to identify bounding box predictions. In related studies, two broad categories have been identified in self-stimulatory behaviors.

In the first category, they use wearable sensors either in the hands or the body which help in tracking of the sensor data for fixed time period. From the results of tracked body parts, we can analyze repetitive self-stimulatory behaviors. Wetyn et.al used wearable wristband in order to track the hand motion, and they built Markov model using for classifying the behaviors.

a. Pdoetz et.al with the help of sensors fixed to the legs, acceleration data can be obtained. Obtained data can be segmented to generate the behavior episodes which can be used for extracting the features in order to train the model which can be used for identifying the aggression, disruption, and self-injury behavior.

Second category includes tracking the child's body part motion in a video and then analyzing the track to detect the behaviors.

a. Marwa et.al identified a new descriptor to find and localize the rhythmic body movements by using color and depth images.

Anna Sudar [4] used Sulinet (an educational webpage). Web page usability and face perception patterns in the population of people with autism spectrum disorders are noted.

a. Face recognition in typical development: Usually, infants with normal condition recognize the human face and gaze at them, but this would not be in autism-affected children.

Several researches have been done by using eye-tracking movements. Latest is based on stimuli and stimulus dataset.

Some other researchers have given a movie snippet to the children. A normal child could able to recognize but autism child was only concentrating on mouth and surroundings.

b. Eye tracking and website ergonomics

Allowed to browse for free, and how children scan different faces are examined. Autistic child treats everything as an object, and it fails to detect face.

Alejandra Ornelas Banrajas et al. [5]. Here, serious game for the children is prepared with tangible user interface (TUI) and geographical user interface (GUI). Later, effects of using serious games in play therapy are noted. Common game therapy is 'building block toys.'

Arodami Choriannopoulou et al. [6]. Here, the degree of children interaction with their parents is identified. All the features that are obtained from the participants

are identified such as acoustic and dialogs. Visual investigations are also performed by observing the recorded video sessions consisting of the tasks performed by the children. From the video, patterns related to autism can be derived.

Mohana et al., a checklist called M-CHAT is used for the children of age between 16 to 30 months. By using the checklist, best classifier with reduced features can be found and which is used for autism risk level prediction. At last performance, evaluation for all the classifiers is evaluated.

Aim is to find toddler autism level by using minimum attributes.

Mythili et.al., discussed various symptoms of autism spectrum disorder and classification models are used for the detection of levels of ASD. They considered autism student database and considered the factors with respect to language, behavior, social interaction, and levels of autism. With the help of an open-source tool called WEKA, analysis, and interpretation of the classification is done, and performance analysis of the students affected with autism are calculated.

Wenbo Liu et al., here, autism is predicted by continuous observation of face scanning and eye movement data ASD prediction is done based on the face scanning eye movement data. Eye movements were recorded using Tobii T60 eye tracker. A framework for prediction is proposed.

Jianbo Yuan et al., in this paper, autism detection is done with the help of NLP based on the information obtained from medical forms of autistic patients. Here, they converted unstructured and semi-structured medical forms into a readable digital format, preprocessed it and finally done with classification.

3 Study Area and Methodology

The area of concentration is on the health care, and, in particular, the prediction of inherited diseases called autism spectrum disorder well in advance. Early prediction of the inherited disease is important as it helps the child to get the treatment and therapy.

Steps are as follows.

3.1 Collection of Data

i. **Dataset to train the model**

It is one of the important steps as it involves collection of behavioral patterns of autistic patients. Neurologists, pediatricians and Head of Autistic schools were consulted to help us in order to collect the dataset. Due to the confidentiality issue, none of the doctors and medical institutions provided the data. Hence, a dataset for autism has been taken from UCI repository. https://archive.ics.uci.edu, which

is created by Fadi Fayez Thabtah, was sufficient to train the machine learning algorithms.

ii. **Real-time data collection for testing model**

After the data was collected from the UCI repository, in order to make the algorithm usable with the data that exists in the real world, data was collected from peers and was circulated to the consulted doctors, and also to the autistic centers. This was made possible by hosting a web page online, which consisted of a questionnaire, that is, a set of questions that had exactly two answers, yes and no. The questionnaire was formed in such a way that it included all the attributes that could determine if the individual is autistic or not. The responses were recorded onto an Excel sheet, where the final score was calculated and fed into the algorithm for testing purposes.

The data which is recorded is based on the responses provided by people who have either been diagnosed for autism or show behavioral symptoms of autism. The main intention of this paper is to make sure that autism is diagnosed accurately, and not be mistaken for any other disorder. The treatment for a misdiagnosed disorder will be of no significance and will be rendered useless. In order to accurately achieve results, machine learning has been incorporated. With the usage of algorithms, the aim is to achieve an accurate diagnosis based on the answers received. The most accurate algorithm among all the algorithms used will be selected as the optimum solution to detect autism.

3.2 Data Cleaning

Data cleaning is very important as it makes it easier to feed the algorithm with data and achieve the expected results. When any form of a dataset is created or retrieved, the data is not perfect. It contains values that are not essential for the computation of results, and there are values that are most needed for the computation of results but do not exist in the dataset. The dataset had to be restructured. In the dataset, all the missing values were filled with the mean of all the values in the attribute's column and all the duplicate values were merged together. Outlier values were also eliminated from the dataset, and the attributes that did not make a difference to the computational process were removed, leaving with a limited number of attributes that had complete records, thus making it convenient for the algorithm to train with.

Among the 21 attributes, the unnecessary attributes were removed as they only took up space and did not contribute to the computation of the data. The scores of 10 questions, age, gender, ethnicity, jaundice, autism history, country and result, were retained. All the spelling errors were corrected, and the missing values were filled with the mean of all the data in that column. All the outliers were discarded. In this manner, the crude dataset was refined and transformed into a usable state. Figure shows all the missing data have been substituted and all the irrelevant features or attributes have been removed (Fig. 2).

Fig. 2 UCI Dataset after cleaning

3.3 Implementation Process to Build a Model (Module 1 Dataset Using UCI Repository)

With the help of UCI repository data, the following algorithms were executed on the training dataset.

A. **k-Means Clustering**

K-means clustering is one of the most important algorithms that is used to solve clustering problems. The way the algorithm works is by classifying a given dataset into a certain number of clusters. The number of clusters is denoted by k, which means that the entire dataset is divided into k clusters with k centers. The clusters are being placed far away from each other in order to obtain maximum accuracy based on the dataset fed into the algorithm. Each point is placed to the nearest center. New centers have to be recalculated based on the clusters in the previous iteration. The k centers change their location until no more changes are possible. For the dataset, the attributes 'age,' 'gender,' 'jaundice,' and 'result' have been used to divide the dataset into two major clusters, namely autistic and non-autistic (Fig. 3).

Observation: The algorithm was successful in classifying into two main classes, namely autistic and non-autistic. The red dots represent the autistic cluster and the green dots represent the non-autistic cluster. There is a graphical representation of the outcome of the K-means Clustering.

B. **Logistic Regression**

It is the most widely used technique for performing binary classification problems. It identifies the correlation between the dependent variable and independent variables.

Fig. 3 UCI dataset after K-means clustering (autistic and non-autistic clusters)

It computes a weighted sum of the input variables and runs the result through a nonlinear function that is the sigmoid function in order to produce the result of the computation. With the help of logistic regression, the output of the algorithm can be converted into a class variable, that is 0, which is 'no', and 1, which is 'yes,' and for this, the sigmoid function is used.

The UCI repository dataset was split into a training set and a testing set. The features that were used to train the algorithm are age, gender, jaundice, and result.

On training the algorithm and using the training the set and feeding the testing set, an accuracy of 94.52% was obtained.

C. **Support Vector Machine**

It is a very fast and a dependable classification algorithm that is efficient, even with a concise dataset. In the dataset, there are two major tags, autistic and non-autistic. From the data-set, two main features are being considered, that is 'jaundice' and 'result' attributes. An SVM takes these points and produces a hyperplane, which is a decision boundary that helps in classifying data points, thus dividing the dataset into two classifications, autistic and non-autistic. In SVM, the linear function's output is considered, and if it exceeds 1, it is identified as one class, and if it is –1, it is identified as the other class, thus giving a range of ([−1,1]) which becomes the margin. A margin is the separation of the line to the closest class points. A margin is said to be considerable when the separation is large for both of the classes. The points remain in the respective classes without merging into the other class. A bad margin is when the separation is small and where there are points, which are supposed to be a part of one class, exists in the other.

Observation: A hyperplane is produced, which separates both the classes. The features that were selected for this was 'Jaundice' and a combination of results gathered from UCI dataset parameters. Any score that ranged from 0 to 6 are classified as non-autistic individuals and any score that exists beyond 6 are classified as autistic. A function was created, that allows an individual to enter the values for the features

Fig. 4 SVM graph with hyper plane and intermediate values

Jaundice and the combined result parameters in order to check if the algorithm is classifying it accurately or not. Finally, the accuracy with UCI dataset came up to 83.3% (Figs. 4 and 5).

3.4 Implementation Process to Build a Model (Module 2 Real-Time Dataset)

A. **Module 2 Dataset**.

This phase involves the collection of data from the public, which includes parents of autistic and non-autistic children, students and faculty members of autistic schools belonging to various parts of the country, and also doctors of some hospitals. The data

Fig. 5 SVM graph classifying into autistic and non-autistic

was collected through a Google Form questionnaire that was integrated into a web page and was hosted online. All the entries were recorded onto an Excel sheet, and on cleaning the data, they were fed into the algorithms for computation. Along with the existing features, more features were added. These features feed in data related to the history of autism in an individual's family, data regarding physical disabilities of the individual, and the individual's reaction in a social environment, which is classified as 'socially uncomfortable,' 'normal,' and 'extreme' behavior (Figs. 6 and 7).

The algorithms that were used for the Module 1 data are being used for the Module 2 data as well. Fabrication of data will lead to the yielding of results that might be too good to be true; hence, a waiting period was kept to hear back from the people who have received the questionnaire. More than 100 entries were received and were used to train the algorithm. This methodology can be used by an individual, or an autistic clinic/school, or by hospitals to diagnose or detect autism accurately in a cost-efficient manner.

For this data, k-means algorithm has been used, and the features that have been selected are 'Behavior in Social Environments (BISE)' and 'result.' In the above figure, two centroids have formed with two different types of clusters (Fig. 8).

After implementing k-means algorithm, the k-nearest neighbors (KNN) algorithm was implemented in order to classify the data into two distinct classes, yielding an accuracy of 92.78.

Logistic regression was also used for the given dataset. Due to the huge number of combinations of the dataset, the accuracy that has been achieved is lower than the accuracy of the dataset retrieved from the internet. Initially, the accuracy of the data

Fig. 6 Web page with autistic questionnaire

was below the half-way mark, but after the processing of data and clearing out the unwanted information, the algorithm yields an accuracy of 96.0. The SVM was also fed with the live collected data, which yielded an accuracy of 88%.

For the Module 2 data, three more algorithms were used to examine whether the algorithms could be trained with the newly collected data and perform accurate predictions. The accuracies of the algorithms exclusively used for the Module 2 data are listed in Table 2 and also compared with the accuracies of the same algorithms used in the research paper [6] (Table 1).

The accuracies of Module 2 data when compared to the research paper are:

A comparative study of the accuracies that were achieved had to be done and to carry that out, the K-means clustering algorithm, the support vector machine

Fig. 7 Questions to assess autistic individuals

algorithm and the naive Bayes algorithm were programmed from scratch and the accuracies were obtained. The Table 3 has a comparison of the accuracies.

4 Conclusion and Future Work

After running through all the different types of data, the algorithms that have been implemented have been trained efficiently and are successfully differentiating autistic individuals from non-autistic individuals. Also, these algorithms have been yielding a higher accuracy when compared to the accuracy obtained by the authors of the research paper that was referred, thus, improving the efficiency and accuracy of the algorithms. Logistic regression and K-nearest neighbor algorithms yielded the highest accuracy among the algorithms used for Module 2 data, thus narrowing down the algorithms that can be utilized. This becomes a cost-efficient, user-friendly tool to help doctors as well as a layman to determine whether someone is autistic or on the verge of being autistic or not. The reach among the public can be easily expanded with the help of the web page to collect the data as computers are available in almost every corner of the world. Smartphones also have not been eliminated

Fig. 8 Module 2 data after K-means clustering

Table 1 Accuracy of algorithms used for module 1 and module 2 data

Algorithm	Module 1 Data (%)	Module 2 Data (%)
K-means	57.19	92.7
SVM	–	88
Logistic regression	94.52	96

Table 2 Accuracy of algorithms used for module 2 data only

Algorithm	Module 2 Data (%)	Results in paper [6] (%)
K-nearest neighbors	96.00	–
Random forest	94.25	85.1
Naive Bayes	93.33	86.5

Table 3 Comparison of built-in and coded algorithms

Algorithms	Built-in algorithm (%)	Algorithm from scratch (%)
K-means classifier	92.70	92.70
Support vector machine	88	70
Naive Bayes	93.33	84

as there are smartphones that are affordable by everyone, and hence, working on a mobile application, either on iOS and Android, will make the module portable and user friendly.

There is definitely a scope for future work on this front. For better accuracy of prediction, the implementation of neural networks and attention networks can be incorporated. With this advancement, a huge amount of data can be taken in at once, and collective analysis can be carried out to produce greater results. As a product, this can be passed on to various medical institutions to help with the diagnosis of autism. Integration of hardware components in order to obtain data is also possible. The use of retina scanners and facial recognition cameras will help us to obtain information by processing the data with the help of techniques like image processing.

References

1. H. Abbas, F. Garberson, E. Glover, D.P. Wall, Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening. In 2017 IEEE International Conference on Big Data (Big Data), p. 35583561. IEEE (2017)
2. S.R. Dutta, S. Giri, S. Datta, M. Roy, A machine learning-based method for autism diagnosis assistance in children. In 2017 International Conference on Information Technology (ICIT), p. 3641. IEEE (2017)
3. W. Liu, X. Yu, B. Raj, L. Yi, X. Zou, M. Li, Ecient autism spectrum disorder prediction with eye movement: a machine learning frame—work. In 2015 International Conference on An active Computing and Intelligent Interaction (ACII), p. 649655. IEEE (2015)
4. M.J. Maenner, K. Van Naarden Braun Marshalyn Yeargin-Allsopp, D.L. Christensen, L.A. Schieve, Development of a machine learning algorithm for the surveillance of autism spectrum disorder. In—PLoS One (2016)
5. A. Pahwa, G. Aggarwal, A. Sharma, A machine learning approach for identification diagnosing features of neurodevelopmental disorders using speech and spoken sentences. In 2016 International Conference on Computing, Communication and Automation (ICCCA), p. 377382. IEEE (2016)
6. B. van den Bekerom, Using machine learning for detection of autism spectrum disorder. In—IEEE (2017)

Speech Emotion Recognition Using K-Nearest Neighbor Classifiers

M. Venkata Subbarao, Sudheer Kumar Terlapu, Nandigam Geethika, and Kudupudi Durga Harika

Abstract The development of wide verity algorithms in artificial intelligence makes it easier in the identification of the emotion of a user. Speech emotion recognition (SER) systems may create a great impact on the working environment by timely identifying the mood of employees. Besides, it is also a key technology in next-generation security applications. This paper presents a verity of K-nearest neighbor (KNN) classifiers for automatic recognition of emotions in speech. For SER, a set of features such as mel-frequency cepstrum coefficients (MFCC) and its temporal derivatives are considered for training the proposed KNN classifiers. Performance analysis is carried on the Berlin dataset with different training rates and with different noisy conditions. Simulation results depict the superiority of the proposed classifier than the existing methods.

Keywords Emotion recognition · KNN classifiers · Distance function · Majority selection

1 Introduction

Emotion plays an important part in everyday social human relations. This is crucial to everyone's rational and smart decisions. It aids everyone to match and recognize the moods of others by passing on their feelings and giving response to others. The study has exposed the role that emotion shows in determining human social communication. Emotional presentations carry significant information about the psychological state of an individual. This makes a new research field in SER to know the desired emotions of a speaker. In the beginning, several researchers came up with a verity of approaches to recognize the speaker's emotions such as physiological signals [1], speech [2], and facial expressions [3]. The majority of investigators are showing interest in speech because of its easiness than other approaches. SER has many applications such as clinical studies, interface with robots, computer games,

M. Venkata Subbarao (✉) · S. K. Terlapu · N. Geethika · K. D. Harika
Department of Electronics and Communication Engineering, Shri Vishnu Engineering College for Women(A), Bhimavaram, AP, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_10

Table 1 Recent SER techniques [4–10]

Ref. no.	Features	Classifier	No. of emotions	Accuracy	Year
4	Pitch, format, phoneme	SVM, HMM	3	75–80	2017
5	LPCC, MFCC	GMM, HMM, SVM	5	82–93	2007
6	Mel spectrogram, harmonic percussive	DNN, KNN	9	90	2018
7	MFCC	SNN	6	83	2019
8	MFCCAcceleration, velocity	DTCNNLSTM	7	80	2017
9	Prosody,quality	SVMPSVM	4	80.7586.75	2017
10	Linguistic and non-linguistic	SVM,DT	6	84.39	2018

cardboard systems, banking, audio surveillance, commercial applications, and entertainment. For laboratory experimentation or online learning, improving the quality of teaching data about the emotional state of students plays a crucial role. SER may help an instructor for developing the new strategies to manage student emotions in the classroom. Some of the recent works on SER is tabulated in Table 1.

From the literature, it is observed that many authors considered a limited number of voice data in the dataset. So that the accuracy achieved is about 80%. There are many features in order to recognize emotion. But the existing techniques use a limited number of features to recognize emotion. For the existing technique, the classification accuracy is calculated only for fixed training rates. Consider these points as input; this paper analyzes the performance with various training rates.

The rest of the paper organization is as follows. Section 2 describes feature extraction and framework of the approach. Section 3 gives the idea about the proposed approach with necessary equations. Section 4 presents the simulation results. Finally, Sect. 5 concludes the work.

2 Framework

In recent approaches, AI techniques like machine learning and deep learning are used for SER. While considering the existing techniques, the performance is degrading due to external disturbances such as noise and environmental conditions. In order to increase the performance of a system, the evaluation is carried out under different noisy conditions with the help of MFCC features and proposed classifiers. The extraction of MFCC features from a voice signal is shown in Fig. 1

Figure 2 represents the framework of our SER systems. It consists of training and testing phases.

Fig. 1 Extraction of MFCC features

Fig. 2 Framework of SER system

3 KNN Classifiers

KNN classification is supervised classification approach. Figure 3 represents an example of KNN algorithm. It is a two-set problem, so it contains two reference classes, Class A and Class B. X1 and X2 are the two features of the classes, which are used in classification. The triangles indicate Class A, rectangles indicate class B, and the unknown class is represented by star. In KNN, firstly it measures the distance between all the reference classes to unknown class by using predefined distance function.

For the given example, initially K value is taken as 3 and it checks for the 3 nearest neighbors, the majority class among 3 is in Class B so the predicted class is Class B. For the next case, K value is changed to 6 so it checks for the 6 nearest neighbors and the majority in the group is Class A, so the predicted class is Class A. From this example, it is noted that K value and distance function affects the performance of the classifier. The best choice of K depends upon the data.

Fig. 3 Example of recognition using KNN algorithm

Table 2 KNN classifiers

KNN classifier	Distance function	K	d
Fine	Euclidean	1	$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$
Medium		10	$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$
Cosine	Cosine	10	$\frac{\acute{P}.\acute{Q}}{\lvert\bar{P}\rvert\lvert\bar{Q}\rvert}$
Cubic	Cubic		$\left(\sum_{i=1}^{m}\lvert p_i - q_i\rvert^3\right)^{\frac{1}{3}}$
Weighted	Distance weighting		$w = \frac{1}{d^2}$

Based on the distance function and K value, the KNNs are subcategorized into different types shown in Table 2 [11]. Here, d is the formula for distance function.

4 Simulation Results

Table 3 represents the simulation parameters for the simulation environment. It consists of seven emotions which are anger, anxiety, boredom, disgust, happiness, sadness, and neutral. The performance is measured with additive white Gaussian noise as disturbance ranging from 0 to 40 dB. The size of dataset is 80000 × 82 which consists of 7 emotions with 39 statistical features with one label. The dataset for training is ranging from 90 to 50%, and the dataset for testing is ranging from 10 to 50%.

Table 3 Simulation parameters

Parameters	Description
Number of emotions	7
Emotion classes	1. Anxiety/fear 2. Anger 3. Boredom 4. Disgust 5. Happiness 6. Sadness 7. Neutral
Features	MFCC, MFCC Δ, and MFCC ΔΔ
Dataset	80,000*39
Rate of training	50–90%
Rate of testing	10–50%

Figure 4 represents the original dataset for the training phase at 0-40 dB SNR for MFCC at 90% training.

Figure 5 represents the performance of fine KNN. Figure 5a shows the confusion matrix, and Fig. 5b represents the accuracy of each emotion at 90% training. The overall accuracy obtained from the fine KNN is 90.1%.

Figures 6 and 7 represent the performance of medium and cosine KNN classifiers, respectively. The overall accuracy obtained from these two are 63.4% and 65.6%. From Fig. 6, it is observed that the medium KNN has very low accuracy in identifying the disgust emotion of a speaker, and the overall performance is poor as that of fine KNN.

Fig. 4 Seven class dataset points for simulation

Fig. 5 Performance of fine KNN

Fig. 6 Performance of medium KNN

Fig. 7 Performance of cosine KNN

Table 4 Performance at different training rates

Classifier	Training rate (%)							
	50	55	60	65	70%	75	80	85
Fine KNN	75.9	77.7	79.3	81.2	83.1	84.7	86.5	88.0
Medium KNN	60.0	60.8	61.3	62.2	62.5	62.6	63.1	63.5
Cosine KNN	62.1	63.0	63.8	64.1	64.7	65.0	65.4	65.8
Cubic KNN	59.8	60.0	60.6	61.4	61.9	62.2	62.8	62.7
Weighted KNN	76.1	78.0	79.8	81.5	83.3	84.9	86.5	87.9

Fig. 8 Performance of cubic KNN

Fig. 9 Performance of weighted KNN

Figures 8 and 9 represent the performance of cubic and weighed KNN classifiers, respectively. The overall accuracy obtained from these two is 62.8 and 88.9%. Similarly, the performance of KNN classifiers with 50% to 85% training rates and 15–50% testing rates are shown in Table 4. From the simulation, it is shown that fine KNN and weighed KNN classifiers achieve better accuracy in recognition of emotion in a speech even at less training rates.

5 Conclusion

KNN classifiers are developed and investigated for SER. To train the classifiers, different MFCC features are extracted from each class of emotion. To verify the performance of classifiers, Berlin dataset is considered. Further, the performance of proposed KNN classifiers is analyzed under different training rates with various values of SNR. Finally, a performance comparison of all proposed KNN classifiers is made. Out of all the KNN classifiers, fine KNN and weighed KNN classifiers attain good accuracy in recognition of emotion in a speech even at less training rates.

References

1. H. Ali, M. Hariharan, S. Yaacob, A.H. Adom, Facial emotion recognition using empirical mode decomposition. Expert Syst. Appl. **42**(3), 1261–1277 (2015)
2. Z.T. Liu, M. Wu, W.H. Cao, J.W. Mao, J.P. Xu, G.Z. Tan, Speech emotion recognition based on feature selection and extreme learning machine decision tree. Neurocomputing **273**, 271–280 (2018)
3. M . Ragot, N. Martin, S . Em, N. Pallamin, J.M. Diverrez. Emotion recognition using physiological signals: laboratory vs. wearable sensors. In: International Conference on Applied Human Factors and Ergonomics. Springer, pp. 15–22
4. M.S. Likitha, S.R.R. Gupta, K. Hasitha, A.U. Raju, "Speech based human emotion recognition using MFCC," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, pp. 2257–2260 (2017), https://doi.org/10.1109/WiSPNET.2017.8300161.
5. H. Hu, M. Xu, W. Wu, "GMM supervector based SVM with spectral features for speech emotion recognition," 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI, pp. IV-413-IV-416 (2007). https://doi.org/10.1109/ICASSP.2007.366937
6. K. Tarunika, R.B. Pradeeba, P. Aruna, "Applying machine learning techniques for speech emotion recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore, pp. 1–5 (2018). https://doi.org/10.1109/ICCCNT.2018.8494104
7. E. Mansouri-Benssassi, J. Ye, "Speech emotion recognition with early visual cross-modal enhancement using spiking neural networks," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1–8. https://doi.org/10.1109/IJCNN.2019.8852473
8. S. Basu, J. Chakraborty, M. Aftabuddin, "Emotion recognition from speech using convolutional neural network with recurrent neural network architecture," 2017 2nd International Conference

on Communication and Electronics Systems (ICCES), Coimbatore, pp. 333–336 (2017). https://doi.org/10.1109/CESYS.2017.8321292

9. Z. Han, J. Wang, "Speech emotion recognition based on Gaussian kernel nonlinear proximal support vector machine," 2017 Chinese Automation Congress (CAC), Jinan, pp. 2513–2516 (2017). https://doi.org/10.1109/CAC.2017.8243198
10. J. Mao, Y. He, Z. Liu, "Speech emotion recognition based on linear discriminant analysis and support vector machine decision tree," 2018 37th Chinese Control Conference (CCC), Wuhan, pp. 5529–5533 (2018). https://doi.org/10.23919/ChiCC.2018.8482931
11. M. Venkata Subbarao, P. Samundiswary, "K—nearest neighbors based automatic modulation classifier for next generation adaptive radio systems", Int. J. Secur. Appl. (IJSIA), NADIA, **13**(4), 41–50 (2019)

Object Detection and Voice Guidance for the Visually Impaired Using a Smart App

Ramya Srikanteswara, M. Chandrashekar Reddy, M. Himateja, and K. Mahesh Kumar

Abstract In the world, there are around 300 million people ("World Sight Day 2017:" Available: https://www.indiatoday.in/education-today/gk-current-affairs/story/world-sight-day-2017-facts-and-figures-1063009–2017-10–12 (2017) ["World Sight Day 2017:" Available: https://www.indiatoday.in/education-today/gk-current-affairs/story/world-sight-day-2017-facts-and-figures-1063009-2017-10-12 (2017)]), who have problems with vision, among those 40 million are completely blind and around 260 million have poor vision due to some cases being partial or complete visual disability. And among them around 95 percent of these stay in developed countries, where they find it very difficult to perform basic day-to-day activities like commuting. They are unable to read traffic warning signs and regulatory signs, also they cannot read the information signs to know their exact position and they often rely on other pedestrians to guide them to their destination. This proposed model aims to provide a method to solve this issue through an application that contains an image recognition system that detects nearby objects in surroundings. It makes the lives of visually impaired better by making them independent.

Keywords Object detection · Voice guidance · Visually impaired · Text to speech

1 Introduction

Human beings have the most useful senses, the eyes that help them look around the world. But there are many people around the world who are unfortunate and lack vision. Millions of people [1] around the world are dealing with the incapacities of understanding and navigating the environment due to partial or complete blindness. Even though they can challenge to perform their day-to-day work, they might possibly have difficulty in following or directing a route as well as social incommunicado. It is very difficult for them to find particular things in unexpected, unfamiliar, and

R. Srikanteswara (✉) · M. C. Reddy · M. Himateja · K. M. Kumar
Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: ramya.srikanteswara@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_11

dangerous environments. Blind and visually impaired people find the maximum difficulty than any other is in knowing whether a person is nearby or not. Computer vision plays an important role, by providing various algorithms for detecting objects. They can be categorized into two main methods, deep learning approach [2, 3] and the conventional approaches [4–6]. Computer vision technologies with unprecedented power of the deep convolutional neural network or simply ConvNets combined with computing power opened huge possibilities in computer vision. It is suitable to use the current technologies to give a ray of hope to the visually challenged people. The proposed model explores the possibility of helping the blind, by familiarizing them, with the world around them.

2 Background

From the past few decades, as image processing algorithms became more powerful, a number of technical projects around the world have been developed which help the blind. One of the methods [7] made use of the concept of local features extraction. Pictorial representation of the surrounding objects for these people makes use of recognizing an object in the video scene where state-of-the-art techniques like key points mining required data and structures matching are used for object identification. For simulating the results, SFIT algorithm was used. The method was successful in detecting the objects around. But the drawback of this method was that the outputs depend on the quality of the image.

Researchers have brought up wide-variety solutions like navigation systems [8] which are effectively inexpensive. This guides the visually impaired by sensing the hindrances in front of the person using sensors and microcontrollers. One more technique effectively recognizes the objects with wearable solutions [9] that uses microcontrollers. These are configurable and instantaneous objects that keep track of the surrounding to and give appropriate output. In this technique, a NOIR camera is being used. The TensorFlow API helps in creating different models which can identify and classify various objects. The Raspberry Pi3 helps in with the system. The system used may be quite expensive, and out of reach of common man.

When the visually impaired people want to lead a normal life, they need to be independent. Few models have been implemented, which help them in moving around. One of the procedures [10] makes use of IoT and Raspberry Pi to deliver sound direction over crosswalks. Recognition of image was made use of in this procedure. Identifying the light of the cross-walk images of significant situations was used as input data for training. These were captured and saved during the day and in the dark. The training of the model was done by Google Object Detection API. This had a limited dataset.

In this era, communication plays an important task. Data is available in most places, in the role of written text. It is a problem for the visually impaired people as they miss out this information. A technique [11] was introduced where the text was converted to speech. The main modules included voice module for searching, image

module, and voice output processing module. Raspberry Pi and sensors were used here. This project mainly dealt with limited trained data, i.e., the bus number and bus name details. A text to read system [12] was described in one of the projects. This was helpful for the visually challenged people. This module was implemented using the OCR and scanner. This device was small scale, less expensive, and portable device. This made use of Raspberry Pi to convert text to speech. But one drawback was the use of scanner.

The systems developed are mostly travel aid for the visually impaired that guide them in travel from one place to another with the help of onboard maps or GPS. These systems are not reliable in certain situations like indoors or when the obstacles are not in the range of the sensors. There may be need of multiple sensors, microcontrollers, and devices in order to be configured. They may also require many other troubleshooting and expertise personnel, whereas a common man might not be suitable for the task. The most important function is that the system should provide the results in real time. Most of the systems are used for a specific purpose like obstacle avoidance which needs to be more generic that can be used for multipurposes like navigation, detection of the objects in and around the environment. Hence, there is a need for a system, that guides the blind people to find out objects around them in real time. The need for a speech output is the main requirement of the hour.

3 The Proposed System

The proposed system is capable of delivering on real time which must have the potential of delivering reliable object detection and also rough calculation of objects position pipeline, with the goal of informing the user with the synthesized audio about the surrounding objects, their spatial position with respect to the user along with the name of the objects.

The block diagram in Fig. 1 shows the pipeline in which the system works. The use of an input camera device is made for video capture, and these images are passed these video data frames to the framework which returns detection data which is communicated to the users.

The proposed system has the capability of being done, effected and also accomplished by using object detection, object relationship analysis to locate a blind person's environment, thus helping them along routes to destinations, providing them with reliable information about nearby points and objects of interest, and indeed with the synthesized audio guiding them.

The technologies used are as listed below.

A. YOLO/Darknet Framework:

YOLO is a network model [13] which is used for detecting objects. It is a model that is very simple to understand. The method used here is straightforward. Prior to YOLO, many classifiers were used to achieve object detection. In just one evaluation,

Fig. 1 Proposed system design

this framework predicts the object. YOLO is very fast and flexible. It achieves 51 ms on a [14] Titan X GPU. The YOLO network assimilates the various boxes extraction, extraction of features, and classification methods into a framework. Direct extraction [15] of the candidate boxes from the images is done for detecting objects.

B. OpenCV:

OpenCV is a frequently used computer vision library which is open source with wide community support. Its original cause of growth was by the Intel corporation. OpenCV currently supports most of the programming languages and also many deep learning APIs or frameworks like Keras, TensorFlow, MXNet, Caffe, Gluon, torch according to a defined list of supported layers.

C. Speech Engine:

Voice assistance is delivered using a cross-platform library [16], pyttsx3. This engine is used to synthesize text into audio users that will be able to hear through their speakers. This package works well in both Windows and Linux operating machines. This engine uses a naive speech driver which helps in good voice synthesis so it is the most flexible text-to-speech package in the open-source world of Python.

4 System Architecture and Implementation

4.1 Methodology

Once the application is started, the camera captures the frames. These frames are fed to the framework, which detect illustrations of objects in video input by inserting boundary containers around them. Then these annotated texts are converted to audio

Fig. 2 Architecture of the framework

outputs. They give the state of each thing as the user can perceive. So, the detected objects will be all made to store in a data file, and this file later referred for the text-to-speech translation of the object names to voice, to communicate to the users.

4.2 Description of the Process

A. **Training Data**:

The training of the model here is done with the Common Objects in Context (COCO) dataset which has 20 classes. The COCO dataset is an admirable dataset for object detection with 80 object categories, around 1 lakh training images.

B. **Model**:

The prototypical design described in Fig. 2 is you only look once (YOLO) object detection framework that is framed of tremendously compound ConvNets which is convolutional neural network, where this architecture is named as dark net. The architecture is simple, and it is just a convolutional neural network. YOLO makes use of the total of the square errors among the predicted values and the value of the correct value which is the expected values [17], finally to calculate the losses which will help to decide the typical accuracy. The loss function will be having both the classification loss and the object's localization loss and the corresponding errors between the predicted bounding boxes, ground truth and also the confidence. It uses these multiple layer types which are ConvNets with a (3,3) kernel and max pooling layers with a (2,2) kernel [18].

C. **Input Data**:

The input device used is a camera video stream as shown in Fig. 3. This camera feeds the images as frames every second (FPS) to the trained model, and it can be set up to

Fig. 3 Camera input

work with every alternate frame to fasten the procedure. This video capture is possible with OpenCV library's module called Video Capture. This class is for video sequence capturing from webcams, cameras, video format files, and sequences of images. It also provides multiple API calls for the video capturing from the above-mentioned input.

D. **Processing and calculations**:

In the classification and localization, the data normally that comes out of the framework is in the form (X,y), where X image data matrix and y is an array containing all the class labels that corresponds to image X, bx, by, bw, and bh [19] as shown in Fig. 4 above, where bx = Detections box x-coordinate, by = Detections box y-coordinate, bw = Detections box width, and bh = Detections box height.

The image is divided into boxes to do object localization tasks so the ConvNets [20–25] in place here. Then one more output layer predicts bounding box coordinates and do the required tweaking the loss function. The input image is passed on in the pipeline to the framework which then divides into grids in a single pass. The process of image object classification and determination of object location are applied on each of the grid. Then it predicts the bounding rectangle and their corresponding class probabilities for objects in the box.

If there is an object located in a grid, it will take the midpoint of the grid where there are objects. The corresponding detection data will then be allocated to the network which comprises of midpoint of these detected substances and the class label for the midcentral grid will be assigned. Even in some cases, if an object might be present in multiple grids, it will only be assigned to a single grid in which its midpoint is located. The x-coordinate of the detections box and y-coordinate of the detections box will be between 0 and 1 as the midpoint will always lie within the network, but width of detections box and height of detections box can be greater than one in few circumstances, when the magnitudes of the rectangle box have additional magnitudes compared to the grid.

Fig. 4 Finding the width of an object

E. **Voice Module**:

When the prediction of class is done, in each frame the objects found will be a set of object name. The coordinates of these things can be found in the picture and the location will be appended like "top," "mid," "bottom" and "left," "center," "right" to prediction of class of the object. One such example is given in Fig. 5.

The texts explanation is sent to Python Text to Speech application that uses the pyttsx3, a multiplatform library. It lets you synthesize audio from text that a user can hear through a speaker. Once the object is detected, the message "object found" is fed as text to the console and to the text-to-speech package.

The message is as shown in Fig. 6.

Figure 7 depicts the model used for text-to-speech conversion. The pyttsx3 Python module supports naive text-to-speech API packages.

Fig. 5 Finding the objects label and position

Fig. 6 Message displayed after object is detected

Fig. 7 Text-to-speech conversion

5 Results

The proposed model is proven to be more accurate than the existing models. This helps the visually impaired identify objects around them. Objects in the scene will be detected with their object class label, confidence on top of the bounding box, and this information along with the positions communicated to the users also printed in the console. As shown in Fig. 8, objects in the scene, a bag is detected with 87% confidence with blue color bounding box and a bottle in yellow box with 96% confidence.

Fig. 8 Objects in the scene a bag detected

The coordinates of the bounding box can also be obtained for every object detected in each frame passed. This will overlay the boxes on the objects detected and return the stream of frames as a video playback. This is scheduled to get a voice feedback on the first frame of each second with the object name along with its position of the camera view and also the voice feedback to the user.

Figure 9 shows predictions of objects with their relative positions, like the top/bottom, left/right. Figure 10 shows the output predicted and the statistics displayed in the console.

6 Test cases

The model was tested in several scenarios. All these are listed in Table 1, with input, expected result, and the output. There are four cases listed in the table. Out of this one case which fails is when the objects are not in the training dataset.

7 Conclusion

The proposed model will be helpful to the visually impaired people. A straightforward, flexible, configurable, easy to handle app with any requirement for technical expertise voice guidance system is the proposed system. This is capable of providing the valuable assistance and support for most blind and visually impaired people. The objects around them can now be heard. This enables them to know their position and helps them to decide where to go without help from anyone else. It not only makes

Fig. 9 Predictions of objects with their relative positions

Fig. 10 Output predicted and the statistics displayed in the console

their travel easier but also saves a lot of time and energy and enhances their safety on roads.

It makes the lives of visually impaired better by making them independent. The outcomes of the proposed system show clear signs that it is flexible, reliable, effective, and individualistic capacity. It has shown 90% accuracy in specifying the objects that may be encountered by the blind. It is able to scan any environment which the input camera is pointed to by the user. This system is accessible to anyone through a simple interface.

Table 1 Test cases for the proposed system

Input	Expected result	Output (passed/failed)
Still objects shown to camera	Objects are detected and communicated to users	Passed
Moving objects	Objects are detected and communicated to users	Passed
Objects very far away from the camera or objects in fast motion	Objects are detected and communicated to users	Passed
New object which is not present in the training dataset	Object not detected	Failed

References

1. "World Sight Day 2017:" Available: https://www.indiatoday.in/education-today/gk-current-affairs/story/world-sight-day-2017-facts-and-figures-1063009-2017-10-12(2017)
2. P. Dollar, R. Appel, S. Belongie et al., Fast feature pyramids for object detection. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1532–1545 (2014)
3. P. Dollar, Z. Tu, P . Perona, et al., Integral Channel Features. British Machine Vision Conference, BMVC, pp. 1–11 (2009)
4. Y, Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. Comput. Learn. Theory, pp. 23–27 (1995)
5. D.G. Lowe, Object recognition from local scale-invariant features: International Conference on Computer Vision, pp. 1150–1157 (19999)
6. M.A. Hearst, S.T. Dumais, E. Osman et al., Support vector machines. IEEE Intell. Syst. Their Appl. **13**(4), 18–28 (1998)
7. H. Jabnoun, F. Benzarti, H. Amiri, Object detection and identification for blind people in video scene: Université de Tunis El Manar, Ecole Nationale d'Ingénieur de Tunis, In Proc. ResearchGate (Dec 2015)
8. A. Noorithaya, K. Kumar M, Dr. A. Sreedevi, Voice Assisted Navigation System for the Blind : R.V.C.E , Bangalore, India, In Proc. International Conference on Circuits, communication, control and computing (2014)
9. A. Nishajith, J. Nivedha, S.S. Nair, Prof. J.M. shaffi, Smart Cap—Wearable Visual Guidance System For Blind, In : Proc. International Conference on Inventive Research on Computing Applications (2018)
10. H. Park, H. Won, S. Ou, J. Lee (2019) Implementation of crosswalk lights recognition system for the blind's safety: Signal, Images et Technologies de l'Information (LR-SITI-ENIT), IEEE Eurasia Conference on IOT, Communication and Engineering (2019)
11. R.K. Rakshana, C. Chitra, A smart navguide system for visually impaired. Int. J. Innovative Technol. Explor. Eng. (IJITEE) ISSN: 2278–3075, **8**(6S3) (Apr 2019)
12. Dr. B. Muthusenthil, J. Joshuva, S. Kishore, K. Narendiran: Smart Assistance for Blind People using Raspberry Pi: International Journal of Advance Research, Ideas and Innovations in Technology, ISSN: 2454–132X, Volume 4, Issue 2, (2018)
13. J. Redmon, S. Divvala, R. Girshick, A. Farhadi: You Only Look Once:Unified, Real-Time Object Detection: IEEE Conference on Computer Vision and Pattern Recognition. (2016). https://doi.org/10.1109/CVPR.2016.91.
14. Joseph Chet Redmon: YOLO: Real-Time Object Detection: Available : https://pjreddie.com/darknet/yolo/
15. C. Liu, Y. Tao, J. Liang, K. Li, Y. Chen: Object Detection Based on YOLO Network: IEEE 4th Information Technology and Mechatronics Engineering Conference. https://doi.org/10.1109/itoec.2018.8740604 (2018)

16. NanoDano: Text-to-speech in Python with pyttsx3 2018, Avaiable: https://www.devdungeon.com/content/text-speech-python-pyttsx3
17. Z.Q. Zhao, P. Zheng, S.T. Xu, X. Wu:Object Detection with Deep Learning: A Review. In: Proc. arXiv:1807.05511v2 [cs.CV] (2019)
18. J. Redmon, S. Divvala, R. Girshick: You Only Look Once:Unified, Real-Time Object Detection. In : Proc. arXiv:1506.02640v5 (2016).
19. H. Yang, H. Wu, H. Chen: Detecting 11K Classes: Large Scale Object Detection without Fine-Grained Bounding Boxes, In: Proc. arXiv:1908.05217v1 [cs.CV] (2019) .
20. M.M. Kamal, A.I. Bayazid, M.S. Sadi, M.M. Islam, N. Hasan: Towards developing walking assistants for the visually impaired people , In: Proceedings IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, (2017), pp. 238–241.
21. S. Haque, M.S. Sadi, M.E.H Rafi, M.M Islam, M.K Hasan: Real-time crowd detection to prevent stampede: Proceedings of International Joint Conference on Computational Intelligence. Algorithms for Intelligent Systems, pp. 665–678. https://doi.org/10.1007/978-981-13-7564-4_56, (2020).
22. Y. Bengio, Learning deep architectures for AI. Found. Trends Mach. Learn. **2**(1), 1–127 (2009)
23. J. Bai, S. Lian, Z. Liu, K. Wang, D. Liu, Virtual-blind-road following-based wearable navigation device for blind people: IEEE Trans. Consum. Electron. **64**(1), 136–143 (2018)
24. W. Elmannai, K. Elleithy: Sensor-based assistive devices for visually impaired people: current status, challenges, and future directions :Sensors 17(3), 565–606 (2017).
25. V. Bansal, K. Balasubramanian, & Natarajan: Obstacle avoidance using stereo vision and depth maps for visual aid devices: SN Appl. Sci. **2**, 1131 (2020). https://doi.org/10.1007/s42452-020-2815-z

Application to Aid Hearing and Speech Impaired People

Akshatha Patkar, Steve Martis, Anupriya, Rakshith, and Deepthi G. Pai

Abstract One of the most priceless gifts to a natural being is the capability of vision, hear, express and react correspondingly to the situations. Interaction between deaf–dumb and ordinary beings is an inspiring mission. The hearing-impaired and the mute society depends mainly on the hand gestures known as the sign language for communication. The sign language identification is one of the revolutions for serving the specially-abled society. The exploration of identifying sign gestures is successful but involves an exclusive charge to be commercialized. For the sign language identification system to be used widely, the data acquisition process varies largely depending on the cost of the system, the methods used, limitations, etc. The course of learning, recognizing the signs and interacting via the ISL can be simplified by the proposed system that converts speech to the sequence of sign language symbols. Speech processing embraces speech recognition, the learning of identifying the vocabularies being vocalized, irrespective of who the orator is. The proposed system practices template-based detection as the key tactic where the Voice to Sign (V2S) system initially requires to be skilled with a dialogue plan based on the predefined database of signs. It correspondingly translates speech to text via the dictation recognizer of the Unity 3D tool, processes the text and maps the phrases to animations that will assist to convey the desired message. Employing the proposed system trainers will be able to teach sign language effortlessly without explicit training.

Keywords Communication technology · Indian sign language (ISL) · Sign language · Recognition · Speech to text · Text to speech

1 Introduction

The Indian sign language (ISL) is one of the most important modes to exchange ideas for dumb and deaf individuals in India. Each country has its communicating sign language which is a finely structured code of gesture, with a unique meaning mapped

A. Patkar · S. Martis · Anupriya · Rakshith · D. G. Pai (✉)
Department of Computer Science and Engineering, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi, India
e-mail: deepthi.cs@sode-edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_12

to each one. The ISL uses manual communication or body language or signs to convey thoughts and ideas. Communicating using the ISL, on the other hand, requires sophisticated training and practice. On the other hand, locating skilled and qualified teachers for their daily affairs is an extremely tough chore and also overpriced.

The latest statistics issued by the World Health Organization (WHO) demonstrates that for individuals aged above 50 years, the audible range gradually decreases over time all over the world. It is probable that above 900 million public throughout the globe will experience loss of hearing ability by 2050 [1].

Nearly 466 million individuals universally have difficulty to perceive sound and the count of youngsters' ranges in 34 million out of these. The 'Deaf' community comprises individuals of very diminutive or no hearing ability. Communication takes place through hand gestures. All around the world, different communities use variant sign languages in distinctive fragments of the globe. Verbal lingos are few in number [2].

There has been an exponential increase in the recognition and development of the ways to interact with the specially-abled. Those with physical frailties such as deafness, loss of sight, or deaf–blindness depend on their working senses to be in contact and have access to the world. The initial lingo the deaf children absorb is the sign language as their primary language. But only 386 schools for deaf and dumb are established in India as of 2017. They have an issue to express themselves or understand the printed text, numbers and also issues while making an image of abstract thought in mind and are unable to access the information which the normal people do very easily like television content, multimedia on the Web and individual public speeches. They desire to retrieve facts in the mode of hand gestures on the Internet instead of a written subject, but incredibly limited sites propose gist in sign language in the form of cinematic clips and can be retrieved by interpreting the script. Producing the material available on the Internet by means of a filmed methodology is expensive [3]. This demand for the need for a platform that acts as the intermediate between the hearing-impaired person and normal person even if he is not acquainted with the use of the internet or browser. Equipping dumb and deaf with qualified and trained teachers can be a challenging task mainly from a financial point of view.

The existing sign language applications are constrained to text inputs and cannot very clearly handle combined phrases or SOV (subject–object–verb) patterns. These issues are addressed with the assistance of natural language processing to break down the script into slighter comprehensible fragments. All the animations of Avatar will be created. The project's UI part includes a dictation recognizer. Text pre-processing part includes removing stop words, lemmatization, handling synonyms and transforming the given sentence into SOV form since ISL sentences have to be in SOV form mandatorily.

The proposed system converts speech by the conveyer to the sequence of sign language symbols for the dumb and deaf to communicate. It also converts speech to text, processes the text and maps the phrases to animations that will help convey the desired message using Avatars. There have been several considerable research activities in developing a system for the generation and automation of sign language.

This automated system takes either text or speech input to generate sign language as output.

About 4 million deaf and about 10 million find difficulty in hearing are present in India as reported by the All Indian Federation of the Deaf (AIFD). ISL is a way of interaction for a further 1 million deaf grownups and about 0.5 million deaf. The Sign-Language-Recognition (SLR) intends to equip algorithms and approaches to appropriately recognize an arrangement of shaped signs and deliver its gist in manuscript or dialogue. The two key classes of the SLR scheme: vision-based and sensor-based. Vision-based systems exploit pictures and films of symbols developed from cameras. Likewise, sensor-based systems need the custom of wearable strategies armed with sensors to abstract the hand gestures and wave of the signs [4]. The projected scheme makes use of the vision-based system to capture the gestures acted by deaf and dumb.

The SL can be staged via hand gesture either by one hand branded as isolated sign language or two hands recognized as continuous. Sign language. The sign language comprises a solo gesture consuming only a word whereas continuous ISL is an arrangement of signs that causes an expressive verdict. The proposed system operates on a continuous ISL sign recognition procedure using ISL. Continuous ISL signs are typically crafted of two-handed; besides, it is a blend of action as well as still gestures. Hence, the issues to identify signs in the actual setting increases [5]. The continuous SL recognition focuses on studying unsegmented signs of audio–visual recordings and is aimed at dispensation continuous motion films in actual organizations. Lately, deep convolutional neural networks include an incredible influence on associated errands on films. For example, Homo sapiens motion recognition, gesticulation recognition, and recurrent neural networks (RNNs) revealed the significant consequence of studying the time-based needs in symbol noticing [6]. In other words, neural networks are limited to acquiring bit-wise depictions. The hidden Markov models (HMMs) are intended for categorized learning. Though the bit-wise classification is not noise-free to exercise the deep neural networks. Besides, HMM may well be tricky to gather the intricate active dissimilarities given the partial illustration skills.

The idea can not only be used by teachers, but also by any person for that matter, who wants to communicate a given message through sign language without having any training for the same. This idea can be extended to multiple languages also, in which case, languages other than English can also be used. The proposed system can recognize hand gestures accurately with a single normal webcam and convert them into text and voice messages. The objective of this scheme is to distinguish the signs by utilizing the peak precision in addition to the minimum amount period and render the letters of ISL hooked on to the equivalent manuscript and speech in a vision-based fashion or commonly known as Avatars.

2 Literature Survey

Ruxandra Tapu et al. [1] projected a scheme that utilizes equally deep convolutional neural networks and computer vision algorithms carefully devised to sense as well as identify the characteristics of the dynamic orator in the film. As the numerous individuals stage in the identical act and occupies in a discussion, the problem that arises is to classify the individual who is presently voicing. An additional unclear condition points to the event in which the expression of the orator is undetectable. DEEPHEAR permits the audiences at ease to stick to the film information at the same time as the accompanying descriptions by locating the captions that cause to classify the involved orator. The DEEPHEAR skeleton equally utilizes deep convolutional neural networks known as CNN and computer vision procedures to accomplish the innumerable phase's essential, together with detection of a face, recognition, and tracking, film time-based breakdown, involved orator recognition, and identification, script discovery and description locating. The scheme is unqualified of being employed in circumstances of the absent appearance of a person in the cinematography. The outspread of the construction to deal with voice-to-script library for films in which the caption writes down is inaccessible in an early stage.

Richard E. Ladner [2] reviewed modules of message passing equipment for audible range enrichment machinery. Primary automated supports for hearing only improved the hearing capabilities to a little extend along with non-linear disturbance cancellation subjected with the help of some machinery tools. The assistance tools could be high-priced, smart-aids estimated in a couple of thousand dollars also cochlear transplant operation costs more than the reach of a common man. But in both the cases of automated support tools and cochlear transplants, there is a requirement of persistent safeguarding, chiefly replacement of battery has to be done regularly. For the reason of expense, a mere fraction of individuals may perhaps afford these automated hearing tools or manage to have a cochlear implant. Emerging with minimal budget automated tools is the key practical test. The two elementary modes using which the deaf community can interact: script and sign language. Each part of the world has its own distinct ways to sign language, on the contrary, there is no prevailing written form that is accepted by all.

The paper by Jestin Joy et al. [3] examined the Sign-Quiz that customs deep neural networks (DNN). The Sign-Quiz application will be capable of effortlessly being exercised by deaf people and also a normal people. Besides, it is economical, constructed to be used on a website, hand gestures sign education presentation for ISL, applying programmed identification of sign language procedure by means of DNN. Fingerspelling is represented via alphabets of a scripting scheme and digit schemes. It performs as an intermediate linking the SL and spoken lingo. The ISL is capable of signifying English letters A to Z with finger spelling. It characterizes texts that do not have any hand gesture corresponding to it or to manipulate an expression or allotted in the schooling of the SL. The operator has the privilege to hand-pick letters. For the respective letter, pictures and films are presented. But absorbing hand gestures are extremely tricky to be used by a learner lacking the assistant of a

proficient SL expert. Sign-Quiz assessments examined volunteers who participated in it, and the results showed maximum volunteers were enthusiasts of tools-based presentations.

Suhail Muhammad Kamal et al. [4] present a methodological summary of the methods in scheming a Chinese Sign Language Recognition (CSLR) structure. The sign language recognition (SLR) schemes utilized are fingerspelling in circumstances in which different vocabulary, person's name, location and terms without recognized hand gestures are enacted by hand motions. The gesticulation identification examines an arrangement of pictures or motion made by hand activities via gloves embedded with sensors or just a webcam. The Chinese Sign Language Recognition (CSLR) employs neural network and hidden Markov models (HMM). The gloves embedded with sensors testified prominent precision but typically operated on isolated expression identification. Mining of only the applicable descriptions from statistics obtained via the sensor tools may not be generally exploited for the reason of its high price, although it is easy to collect the data from the sensors. Moreover, the dataset causes problems that are aimed at attaining an extraordinary routine of the SLR scheme. The list of datasets available is publicly available dataset and also does not contain much information that can allow to improve and compare various state-of-the-art techniques; incorporating facial expressions that convey extra meaning in a sign language sentence as the state-of-the-art German sign language dataset.

The primary language of deaf people is widely recognised by Runpeng Cui et al. [6], SL, and is typically recorded or transmitted in video form. The SL is also considered to be the gestural communication that is most grammatically organised. This makes SL recognition a perfect area of study for the development of techniques to resolve issues such as human motion analysis, human–computer interaction (HCI) and user interface design, and makes it receive great attention in multimedia and computer vision. This work establishes a continuous recognition system for sign language (SL) with deep neural networks, which transcribes videos of SL phrases directly to sequences of ordered gloss labels. Hidden Markov models (HMM) typically use previous methods dealing with continuous SL recognition with restricted capability to capture temporal information. SL recognition is all about learning long-term video streams with unsegmented gestures and is more suited for processing continuous gestural images in real-world systems. This training also does not require a costly annotation for each gesture on the temporal boundary. Recognizing SL suggests that gestural gestures and appearance attributes are simultaneously evaluated and incorporated, as well as disparate body parts, and therefore, potentially using a multimodal method. The emphasis is on the issue of continuous SL recognition on images, where it is important to learn both the spatiotemporal representations and their temporal matching for the labels.

3 Summary of Literature Review with Research Gaps

The sign language may be executed either by one hand known as independent sign language or by two hands known as continuous sign language using hand gestures. Sign language consists of a single gesture with a single word, whereas continuous ISL or continuous sign language is a set of gestures that produce a meaningful expression. We are working on continuous ISL gesture recognition techniques using ISL in the suggested framework. Continuous ISL movements consist mainly of two hands and sometimes of two hands [7].

Continuous SL recognition is all about learning long-term video streams with unsegmented gestures and is more suited for processing continuous gestural videos in real-world systems. Deep convolutionary neural networks have recently had a huge effect on similar video tasks, e.g. recognition of human behaviour, recognition of movements and sign spotting, and recurrent neural networks (RNNs) have shown substantial results in learning about temporal dependencies in sign spotting [8].

Neural networks are limited to studying frame-wise depictions in other works, and hidden Markov models (HMMs) are used for sequence learning. However, for the training of deep neural networks, the frame-wise labelling is noisy, and HMMs can be difficult to understand the complex dynamic variations, given their limited capacity for representation.

The idea can not only be used by teachers but also by any person for that matter who wants to communicate a given message through sign language without having any training for the same. This idea can be extended to multiple languages also; in which case, languages other than English can also be used.

With a single standard webcam, the proposed device will reliably interpret hand gestures and transform them into text and voice messages. The purpose of this project is to identify the gestures in a vision-based setup or generally known as Avatars with optimum accuracy and at least possible time and translate the alphabets of Indian Sign Language into corresponding text and speech.

4 Dataset Description

The dataset used is obtained from the link http://www.indiansignlanguage.org, which currently consists of all the existing words in the Indian sign language dictionary; 2785 words have been defined to date, and each of them is assigned to a particular category. There are various categories such as places, banking, arts and entertainment, shapes and educational terms. Each word is associated with its respective ISL action, and video for the same can be obtained by just clicking on that word. As we know, the ISL Dictionary is constantly being updated and new words which have their meaning in the ISL are being appended.

5 Methodology

The alphabets, numbers and hand gestures used in the Indian Sign Language varies from that of the other country's sign language. Each region inherits its own hand gestures for each alphabet and number to form words and sentences. The hand gestures for alphabets used in Indian Sign Language is given in Fig. 1.

The following steps are considered in the architectural design of the application:

a. **Speech to text conversion**: Since the basic idea of the project is to convey spoken messages to ISL animation, the input in the form of speech is recorded and converted to text [10]. The input can also be in the form of direct text.
b. **Text pre-processing**: The input obtained cannot be used directly for animations because spoken languages do not adhere to the grammar rules of the ISL. Hence, the input text has to be converted to its equivalent ISL form. Certain rules are to be followed to convey a message in ISL. Words are always used in their root form and their present tense. Undefined signs are split and shown as alphabets. The most significant rule is that every ISL sentence has to be in SOV (subject-object-verb) notation. To achieve this, the text is processed using NLP functionalities such as finding root words, lemmatization, removing stop words and identifying the parts of speech (POS). The final sentence will be in the ISL format.
c. **Text to animation mapping**: The words in the processed input have to be mapped to their respective animations. In case the animation for that word is not present, its synonym is mapped. If the synonym does not exist too, it is split according to its alphabets.

Fig. 1 The Indian sign language [9]

Fig. 2 The architecture design

d. **Animation**: The output is conveyed using an Avatar. These animations are combined to communicate the ISL sentence. Hence, a given sentence is converted into its ISL form.

The application architecture diagram is a series of step-by-step as depicted in Fig. 2.

Figure 3 shows the process of speech to text conversion using Dictation Recognizer. Unity3D's dictation cognizer is used for converting speech to text. This process starts by enabling the microphone under capabilities followed by turning on online speech recognition and linking and typing under system settings. The user speaks through the microphone, which gets converted to text and displayed on the screen.

Figure 4 shows the process of pre-processing using Python. Python and NLTK are used for text pre-processing. This process starts with Tokenization where text is tokenized into words from a sentence, followed by Lemmatization which is the process of grouping together the different inflected forms of a word so they can be analysed as a single item. Finally, stop word removal where all the stop words in the tokenized sentence are removed.

Figure 5 shows the process of sentence simplification using Python. Output of the text pre-processing is now simplified and converted to ISL sentences. The sentence is brought to subject–object–verb (SOV) notation as per ISL grammar rules using

Fig. 3 Speech-to-Text API

Fig. 4 Pre-processing sentence

Fig. 5 Sentence simplify

the spacy module of Python, as all the words in English dictionary does not have a sign associated with it. In the absence of a word, its synonym sign is used. In absence of sign for a synonym, the word is split by letters. As per the ISL grammar rules, the adjectives and adverbs are handled.

Figure 6 shows the mapping of animations using Unity3D and Flask REST API. The output of the Sentence simplifier is called the ISL sentence. The mapping starts by obtaining the ISL sentence from the API call to the Python class which include the Fig. 6 methods. Animations associated with each word in the ISL Sentence is then triggered by the Animator object to display the animations. Flask REST API is used to coordinate between Python and C#. The animations are displayed in GUI which is built in Unity 3D.

Below is the sample output.

A story is dictated using the microphone and it appears as text on the screen as follows:

Fig. 6 Mapper

```
Enter split :  birbal tell to solve complaint
After split :  b i r b a l tell to solve complaint
string without verb['b', 'i', 'r', 'b', 'a', 'l', 'tell', 'to', 'complaint']
```

Fig. 7 Console output for unknown words

One day, a clever man sold his well to a farmer. The next day, when the farmer went to the well to fetch some water, the man refused him telling that he only sold the well and not its water. The farmer, with a sad heart, went to Akbar's court. Birbal was told to solve the complaint. The following day, the man and the farmer were called to the court. The clever man told—he sold his well, not the water in it. On learning this, Birbal said, 'My friend, in that condition, you either remove your water from the well or pay tax for your water because it is the farmer's well'. The man understood his mistake and asked for forgiveness.

After text processing, the output story is in the form of ISL as follows:

One day clever man his pond to farmer sell next day farmer to pond to some water go bring when man him that he only pond reject tell sell n o t it water farmer with sad heart to a k b a r court go b i r b a l tell to complaint solve follow day man farmer call to court clever man he his pond tell sell n o t water in it on this learn b i r b a l order my friend in that condition you e i t h e r your water from pond o r tax f o r your water b e c a u s e it farmer pond remove pay man his mistake understand f o r f o r g i v e n e s s ask.

Table 1 depicts the standard Indian Sign Language rules and regulations based on which the application is built.

This checks the first rule in the table. Fingerspelling is breaking each word to alphabets in case they do not have a sign. In the above case, 'Birbal' which is a proper noun does not have a specific sign in ISL. Hence, it is split to 'b i r b a l'.

Words that have no signs are checked for replacement with synonyms that exist in the ISL dictionary. If it exists, the synonym replaces the original word. In this case, 'fetch' is replaced by 'bring'. Else, the word will be split for fingerspelling.

This checks the Wh-word rule. It says that question words must always be used at the end of the sentence. In the above output, 'what' is appended to the end.

```
when farmer go to pond to fetch water
[('when', 'WRB'), ('farmer', 'NN'), ('go', 'VBP'), ('to', 'TO'), ('pond', 'VB'), ('to', 'TO'), ('fetch', 'VB'), ('water', 'NN')]
fetch replaced by bring
After synonyms :  when farmer go to pond to bring water
```

Fig. 8 Console output for synonyms

```
Enter interrogative handler :
[['what', 'you', 'do']]
[[('what', 'WP'), ('you', 'PRP'), ('do', 'VBP')]]
[('what', 'WP'), ('you', 'PRP'), ('do', 'VBP')]
Final sentence(if interrogative): you do what
Final Sentence: you do what
```

Fig. 9 Console output for wh-questions

```
Input string :  a clever man sold his pond to a farmer
After stop words removal: clever man sold his pond to farmer
```

Fig. 10 Console output for linking verbs and articles

```
Input : clever man sold his pond to farmer
Enter lemmatization with pos :
[['clever', 'man', 'sell', 'his', 'pond', 'to', 'farmer']]
After lemmatization : clever man sell his pond to farmer
```

Fig. 11 Console output for lemmatization

Fig. 12 UI of the application

ISL does not have signs or use linking verbs and articles. In the above case, these two cases are handled. Stop words are unnecessary words that are removed. The article ‘a’ is removed.

The existing live SLR approaches suffer from a few downsides of complexity in identifying composite hand gesticulations, minimal distinguishing precision for

Fig. 13 Working of the application

Fig. 14 Animations according to resultant output recording process

many active SLR, also the probable complications in the bulkier cinematic chain of information guiding. The SL put together a sequence of movements that take in swift gestures in addition to alike features. Thus, the dynamic SLR is problematic to transact in addition to the density and huge dissimilarities of terminology databases in hand gestures. Also, it possibly will misinterpret a variety of alternatives with one another. The dynamic SLR encounters difficulties in production besides the intricacy of gesture behaviours of finger movement of the wide-ranging frame environment. The mining of the utmost judicious characteristics from pictures and films is another added problem. Further, picking a suitable classifier is a desperate feature for influencing exact perception outcomes. Two of the numerous modes to resolve the setback with dynamic SLR are—the first one is the process grounded on hand

Fig. 15 Recording process

Table 1 The ISL rules and regulations

Rule	Input	Output
Finger spelling for unknown words	Birbal said	Birbal state (replaced by synonym)
Never use linking verbs (is, was etc.)	Birbal was told to solve the complaint	Birbal tell to complaint solve
Never use words with suffixes (s,'s etc.)	because it is farmer's pond	Because it farmer pond
Always sign in present tense	Clever man sold his well to a farmer	Clever man his pond to farmer sell (SOV format)
Articles are not used	The next day	Next day
Wh- questions are at the end	When are you leaving	YOU leave when
Never use gerunds	On learning this	On this leam

signs and movement curve of hand-gesticulations; the next one is the technique built on a cinematic arrangement of respective SL [3].

6 Results and Discussion

It aims to eliminate the explicit training which teachers have to undergo to teach in a Deaf and Dumb School. Using this project any person can teach the students in the deaf and dumb school. This project will ensure that deaf and dumb school students

can get access to the majority of the learning materials and resources which were otherwise unavailable to them due to language barriers.

The end product is an open-source desktop application that is platform-independent. The UI is user-friendly which consists of an avatar, an input field, a button to record the voice in case of audio input, and a submit button. The ISL equivalent of the text along with the animation being played is displayed simultaneously after the input is processed. There is no considerable delay to display the animation through an avatar on submitting the text in the input field. The animations are clear with smooth transitions.

This application will simplify the job of teaching the mute community. It is quite hard to learn and master the Sign Language, as it is time-consuming. This application eliminates the requirement of knowing the ISL beforehand. This product can not only be used by teachers, but also by any person, who wants to convey a message through sign language without having any training for the same. The figures below illustrate an idea of the implementation of the application to aid hearing and speech impaired people.

For ages, very little importance has been given to the deaf and dumb. This can be noticed in the epics of this country. We cannot find many characters being deaf and dumb in any of the greatest epics of this country. Although they have the most perceiving sensory organ of the human body, which is the eye, using it to their fullest advantage has not helped them in achieving great feats [9]. The main reason is due to lack of resources and awareness among the people, that's why it's difficult to find the deaf and dumb achievers in this country. So, this project will help the people with such disabilities to learn on level ground along with the rest of the world.

There is no professional technical solution to teach the deaf and dumb until today. Few of the mobile applications which exist now merely convert each word into its respective sign without following ISL grammar. Also, there are hardly any applications that take in voice input also. Some of the existing applications are not open source. This application follows all the ISL grammar rules and regulations. Thus, this project can be of great use to the deaf and dumb and is a humble attempt to give something back to society.

7 Conclusion

The prevailing sign language applications are constrained to text inputs and cannot handle combined phrases or SOV (subject–object–verb) patterns. Many of them are in the form of mobile applications that do not consider audio input. These matters are addressed by means of the semantics of Natural Language Processing to break down the script into slighter comprehensible fragments, and through dictation recognizer which takes in speech as input through a microphone.

Linguistics is entertained as a crucial basis of interaction within human society. Signs can stem from the movement of a body part, usually initiated as the action of expressions made by face and hands. An extent of computer vision is Body Language

Identification, it takes into account understanding the signs via a diverse set of rules. In attendance is a shared delusion regarding the SL that it is global furthermore the society of unresponsive fitting to whichever chunk of the globe uses the identical lingo. However, the aforementioned is known that numerous gesture lingos use the resemblances of one another. The precise count of the prevailing sign languages is not known, nevertheless, the 2013 publication of Ethnologies' rolls them 137.

References

1. Ruxandra Tapu, Bogdan Mocanu and Titus Zaharia, "DEEP-HEAR: A Multimodal Subtitle Positioning System Dedicated to Deaf and Hearing-Impaired People" *IEEEAccess,* vol. 7, 2019.
2. Richard E. Ladner, "Communication Technologies for People with Sensory Disabilities", *IEEE*, vol. 100, 2012.
3. Jestin Joy, Kannan Balakrishnani and Sreeraj "SignQuiz: A Quiz Based Tool for Learning Fingerspelled Signs in Indian Sign Language Using ASLR", *IEEEAccess*, vol. 7, 2019.
4. Suhail Muhammad Kama, Yidong Chen, Shaozi Li and Xiaodong Shi and Jiangbin Zheng, "Technical Approaches to Chinese Sign Language Processing: A Review" *IEEEAccess*, vol. 7, 2019.
5. Formation Kumud Tripathi, Neha Baranwal and G. C. Nandi, "Continuous Indian Sign-Language Gesture Recognition and Sentence", Eleventh International Multi-Conference on Information Processing (IMCIP), *Elsevier*, 2015.
6. Runpeng Cui, Hu Liu, and Changshui Zhang, "A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training", *IEEE*, 2018.
7. Formation Kumud Tripathi, Neha Baranwal and G. C. Nandi, "Continuous Indian Sign Language Gesture Recognition and Sentence", Eleventh International MultiConference on Information Processing (IMCIP), Elsevier, 2015
8. Mu-Chun Su, Chia-Yl Chen, Shi-Yong Su, Chien-Hsing Chow, Hsiang-Feng Hsiu and Yu-Chine Wang, "Portable communication aid for deaf-blind people", Feb. 2001
9. B. Lakshmi, "Assistive SIGN LANGUAGE Converter for DEAF AND DUMB",International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2019.
10. Suharjito, Ricky Andersonb, Fanny Wiryanab, Meita Chandra Ariestab, Gede Kusumaa, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output", 2nd International Conference on Computer Science and Computational Intelligence 2017, Bali, Indonesia, *Elsevier*, October 2017.
11. Anish Kumar, Rakesh Raushan, Saurabh Aditya, Vishal Kumar Jaiswal, Mrs. Divyashree Y.V, "An Innovative Communication System for Deaf, Dumb and Blind People", International Journal for Research in Applied Science & Engineering Technology (*IJRASET*), Volume 5, Issue VI, June 2017.

Variants of Fuzzy C-Means on MRI Modality for Cancer Image Archives

C. K. Roopa, B. S. Harish, and R. Kasturi Rangan

Abstract The segmentation, identification and mining of contaminated tumor region from MRI images is a primary concern in medical image analysis. However, it is a monotonous and time-consuming process done by radiologists or clinical experts, and its precision is subjected to their expertise. To overcome these constraints, the use of supporting technology turns out to be very important. In this study, performance improvement and reduction of the complexity involved in the segmentation of medical images is been focused. We have investigated cancer cells using various fuzzy c-means methods and its different variants for Breast, Brain, Liver and Prostate dataset. The empirical outcomes of the various methods have been tested and corroborated on MRI for efficiency and quality analysis based on four well-known cluster validity indices. The results are very encouraging and robust in nature.

Keywords MRI · Cancer images · Fuzzy C-means · Cluster validity indices

1 Introduction

Computer aided diagnosis (CAD) has become essential in diagnosis of illness. It also plays major role in treatment planning. The other vital reason for the use of computers to assist is the development of medical imaging modalities like CT, MRI, PET and radiography. The main reason of using computer aided diagnosis is that clinical decision support system, where the output of CAD is taken as the second opinion and makes the final decision. CAD assists radiologist in interpreting the different types of abnormalities in medical images. It helps medical professionals in faster and automated diagnosis with higher rate of accuracy. The images obtained from the different medical imaging modalities are copied to CAD in a DICOM format.

C. K. Roopa · B. S. Harish (✉) · R. Kasturi Rangan
JSS Science and Technology University, JSS TI Campus, Mysore, India
e-mail: bsharish@jssstuniv.in

C. K. Roopa
e-mail: ckr@jssstuniv.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_13

This has to be prepared and analyzed using couple of measure like preprocessing, segmentation and evaluation.

Medical image is acquired using various modalities; one of its kinds is MRI. MR images give the structural details of Brain, Breast, Liver and Prostate which helps radiologist in monitoring and analysis for treatment.

The advantage of MRI is that it gives more details in the images when compared to other modalities. All though many research works are attempted on different application of MRI and MR image analysis, and the diagnosis is still a challenging and complex problem. In this paper, we attempted to explore the impact of variants of fuzzy C-means on MRI modality using four different cancer image archives [18] (i.e., Breast, Brain, Liver and Prostate). To mention some of the very well-known research works attempted are: Breast [3, 4, 6, 12, 13, 15, 21]; Brain [7, 9, 10, 25, 28]; Liver [2, 16, 17, 22, 23] and Prostate Cancer [5, 14, 20, 24, 26, 27].

2 Methodology

Accurate segmentation is a crucial step in most of the medical analytical procedures. It is also necessary to isolate the region of interest. If we can identify the boundaries of the ROI (image segment), we can also simplify decisions. Segmentation is a critical step that decides the final result of the proposal as a whole because the majority of the analysis depends solely on it. Since the image is segmented into coherent regions by segmentation, clustering procedures can be used for segmentation by extracting the global image characteristics in order to properly distinguish the region of interest from the context. Clustering can be done in numerous ways such as Density-based, Model-based and Systematic partitioning. In addition, due to the fact that most medical images are ambiguous and disordered data, Fuzzy approach becomes essential to manage uncertainty/vagueness in medical images during the segmentation process.

2.1 Fuzzy Approaches for MRI Segmentation

Fuzzy c-means (FCM) is a well-accepted clustering algorithm proposed by Bezdek [1], and it is based on traditional fuzzy set. FCM cluster the data element based on the membership value.

Though FCM showcased outstanding results on all the 4 dataset, unfortunately, it is sensitive to noise and ignored the neighborhood information. Further, FCM uses Euclidean metric that is sensitive to noise and in turn degrades the clustering results.

To handle the noise, researchers proposed a kernel version of FCM named as Kernel FCM (KFCM) [11]. The problem with KFCM-based methods is that they do not utilize the neighborhood information. To eliminate these limitations, various researchers proposed a variant of FCM named as Spatial Kernel FCM (SKFCM) [8,

19]. However, SKFCM still suffers the following limitations: work for single-feature inputs, high computation time and incorporate neighborhood information only to the objective function.

To alleviate these drawbacks, we found the way out to use two variants of FCM methods, namely Robust Spatial Kernel FCM (RSKFCM) and Generalized Spatial Kernel FCM (GSKFCM). The detailed derivation and explanation of RSKFCM and GSKFCM can be found in [10, 19], respectively.

2.1.1 Modified Intuitionistic Fuzzy C-Means (MIFCM)

The above RSKFCM and GSKFCM are based on the traditional fuzzy set where the decision for belongingness of the data point is founded using the membership value. Traditional fuzzy set assumes that the non-membership value is always the complement of the membership value. But, this assumption is not always true because of hesitation. This hesitation is caused by the uncertainty in defining the membership function. Intuitionistic fuzzy set (IFS) handles this uncertainty by considering the non-membership value and the hesitation degree. In the literature, based on the IFS, many clustering methods are proposed. The existing intuitionistic fuzzy c-means (IFCM) method employs the Sugenos and Yagers intuitionistic fuzzy complement generator function to compute the non-membership value. Unfortunately, for certain values, the aforementioned intuitionistic fuzzy complement generator fails to fulfill the fundamental requirement of intuitionism. To alleviate this limitation, the authors in [11] had proposed MIFCM by adopting a new intuitionistic fuzzy complement generator function. Further, employed the modified Hausdorff distance metric instead of the Euclidean metric. The same was used in this paper, and detailed tests were performed on all 4 datasets. MIFCM method comprises of two steps: Representation and Clustering. Next subsection explains these two steps in detail.

Representation

In IFS, each data point is associated with three parameters and it can be represented as:

$$\mathrm{IFS}(X) = \{< x_i, \mu_X(x_i), \eta_X(x_i), \pi_X(x_i) > | x_i \in X\} \tag{1}$$

where $\mu_X(x_i)$ represents the membership value, $\eta_X(x_i)$ signifies the non-membership value, and $\pi_X(x_i)$ represents the value of hesitation degree associated with data point x_i.

As it is known that existing IFS complement generator fails to fulfill the fundamental requirement of the intuitionism. Thus, to alleviate this limitation, we employed IFS complement generator [11]. The non-membership value $\eta_X(x_i)$ is computed using the Bustince IFS complement generator:

$$\eta_X(x_i) = 1 - \phi(\mu_X(x_i)) \tag{2}$$

where ϕ refers to the Sugeno's or Yager's IFS complement generator. When we use Sugeno's function, Bustince IFS complement generator function becomes

$$\eta_X(x_i)_{\text{Bustince_ Sugeno's}} = 1 - \left(\frac{1 - \mu_X(x_i)}{1 + \alpha\mu_X(x_i)}\right) \tag{3}$$

Similarly, if we use the Yager's function, Bustince IFS complement generator function becomes

$$\eta_X(x_i)_{\text{Bustince_Yager's}} = 1 - (1 - (\mu_X(x_i))^{\alpha})^{\frac{1}{\alpha}} \tag{4}$$

In this paper, we used both the Bustince IFS complement generator (i.e., Bustince–Sugeno's and Bustince–Yager's) function to compute the non-membership value. After applying the Bustince IFS complement generator function, IFS representation is given by:

$$\text{IFS}(X)_{\text{Bustince_ Sugeno's}} = \left\{\left\langle x_i, \mu_X(x_i), 1 - \left(\frac{1 - \mu_X(x_i)}{1 + \alpha\mu_X(x_i)}\right), \pi_X(x_i)\right\rangle | x_i \in X\right\} \tag{5}$$

$$IFS(X)_{\text{Bustince_ Yager's}} = \left\{\left\langle x_i, \mu_X(x_i), 1 - (1 - (\mu_X(x_i))^{\alpha})^{\frac{1}{\alpha}}, \pi_X(x_i)\right\rangle | x_i \in X\right\} \tag{6}$$

The value of the hesitation degree is computed by using the below equations:

$$\begin{aligned}\pi_X(x_i)_{\text{Bustince_ Sugeno's}} &= 1 - \mu_A(x) - \eta_A(x) \\ &= 1 - \mu_X(x_i) - \left(1 - \left(\frac{1 - \mu_X(x_i)}{1 + \alpha\mu_X(x_i)}\right)\right)\end{aligned} \tag{7}$$

$$\begin{aligned}\pi_X(x_i)_{\text{Bustince_ Yager's}} &= 1 - \mu_A(x) - \eta_A(x) \\ &= 1 - \mu_X(x_i) - \left(1 - (1 - (\mu_X(x_i))^{\alpha})^{\frac{1}{\alpha}}\right)\end{aligned} \tag{8}$$

In this method, we convert the crisp input data into intuitionistic fuzzy set (IFS) representation by employing the Bustince IFS complement generator. Let $X = \{x_1, x_2, x_3, \ldots, x_i, \ldots, x_n\}$ be the input data with samples. Firstly, we convert the input data into fuzzy form using the min–max normalization technique. The membership value for each data point is calculated using the below equation:

$$\mu_X(x_i) = \frac{x_i - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)} \tag{9}$$

Further, using these membership values, non-membership is computed using Eqs. (2) or (3) and the hesitation degree is computed using Eqs. (7) or (8). Finally, the

input data is represented in the IFS form using all the three parameters (i.e., membership, non-membership and hesitation degree). Further, in the next step, these IFS converted input is clustered using the enhanced intuitionistic fuzzy c-means method.

Clustering

Improved intuitionistic fuzzy c-means (IIFCM) is the modified conventional IFCM algorithm which groups the input data by optimizing the subsequent objective function presented below:

$$J = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^{m} \, d(X'_j, v_i) \tag{10}$$

with the constraints:

$$\sum_{i=1}^{c} \mu_{ij} = 1; \quad 1 \leq j \leq n \tag{11}$$

$$\mu_{ij} \geq 0; \quad 1 \leq i \leq c; \quad 1 \leq j \leq n \tag{12}$$

$$\sum_{j=1}^{n} \mu_{ij} > 0; \quad 1 \leq i \leq c \tag{13}$$

where $X'_j = \{\mathrm{IFS}(x_j)\}$ denotes the IFS of data point x_j, $V = \{v_1, v_2, \ldots, v_i, \ldots, v_c\}$ denotes to c number of cluster centers and each v_i is related with the membership, non-membership and hesitation degree value, m denotes the fuzzification value, μ_{ij} signifies the membership value of jth data point to ith cluster, $d(X'_j, V_i)$ is the distance between jth data point to ith cluster center.

In IIFCM, the decision of belongingness of a data point to a cluster is made based on the membership value and it is computed using the distance function. Thus, the performance of the IIFCM depends on the selection of the distance metric. The conventional and existing variants of the IFCM method employ Euclidean as the distance metric which is very sensitive to noise. To alleviate this problem, the modified Hausdorff distance metric is used instead of the Euclidean distance metric. Modified Hausdorff distance metric is computed using the following equation:

$$d_{\mathrm{H}}\left(X'_j, V_i\right) = \max\left\{|\mu_{X'_j} - \mu_{V_i}|, |\eta_{X'_j} - \eta_{V_i}|, |\pi_{X'_j} - \pi_{V_i}|\right\} \tag{14}$$

Similar to fuzzy c-means, IIFCM optimizes the objective function iteratively by updating the membership and cluster centers. The membership value is updated subsequently as follows:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d\left(X'_j, V_i\right)}{d\left(X'_j, V_k\right)}\right)^{\frac{2}{m-1}}} \tag{15}$$

Further, based on the membership value, non-membership and hesitation degree are updated. In IIFCM, comparable to data point, cluster center V_i is also related with 3 IFS variables. Consequently, the center values are essentially to be updated for all the 3 variables. In the MIFCM, cluster centers are upgraded using (16)–(18) equations.

$$\mu_{V_i} = \frac{\sum_{j=1}^{n} \mu_{ij}^{m} \mu_{X_j}}{\sum_{j=1}^{n} \mu_{ij}^{m}},\ 1 \le i \le c,\ 1 \le k \le n \tag{16}$$

$$\eta_{V_i} = \frac{\sum_{j=1}^{n} \eta_{ij}^{m} \eta_{X_j}}{\sum_{j=1}^{n} \eta_{ij}^{m}},\ 1 \le i \le c,\ 1 \le k \le n \tag{17}$$

$$\pi_{V_i} = \frac{\sum_{j=1}^{n} \pi_{ij}^{m} \pi_{X_j}}{\sum_{j=1}^{n} \pi_{ij}^{m}},\ 1 \le i \le c,\ 1 \le k \le n \tag{18}$$

IIFCM is an iterative process and it halts when the convergence condition is satisfied. Once the IIFCM converges, each data point is related with the membership values. Using the membership values, data points are grouped such that the data point is assigned to a cluster for which it has maximum membership value.

3 Empirical Evaluation

This section presents the empirical results of the proposed variants of FCM on all the four dataset. To find the optimal values of the parameters, we used cluster validity indices as a metric. We set the fuzzifier m value to 2 and stopping criteria G to 0.00001 empirically for all the four datasets. We varied the values of p and q from 0 to 3 with an increment of 1. If we set the values $p = 1$ and $q = 0$, RSKFCM becomes traditional FCM. On the other hand, if we set $p = 0$ and $q = 1$, then RSKFCM becomes Spatial FCM method. The experiments indicated that, if we set p and q to larger values, the results are deteriorating because of the loss of important information in the input data. Thus, we varied the value of p and q till 3.

We varied the window size from 3 to 9. When the window size is large, we had similar neighborhood data and the impact of neighborhood information was more. However, due to the increase in the computation time and bias results, the window size could not be too large. We varied the Kernel width σ value from 50 to 300. When we further increased the σ value, there was no improvement in the results. Thus, we set 300 as the maximum limit for σ.

In MIFCM, we set the fuzzifier m value to 2 and stopping criteria G to 0.00001. We varied α value from 0.1 to 30 with a uniform increment of 0.1. Since the results are not varied in between values, we presented the results only for those values which give different results. In experiments, we observed that there is no variation in the results when $\alpha > 30$. So, we set 30 as maximum limit for α value. In the proposed method, we adopted both the Sugeno's and Yager's function for the complement generator and presented results for the same. In the results, MIFCM_S represents Bustince with Sugeno's IFS complement generator and MIFCM_Y represents Bustince with Yager's IFS complement generator. To validate the proposed methods, we used Partition Coefficient (V_{pc}), Partition Entropy (V_{pe}), Fukuyama-Sugeno function (V_{fs}) and Xie-Beni function (V_{xb}) as a performance evaluation metrics [11]. Tables 1, 2, 3 and 4 present the empirically evaluated results on the four datasets.

Table 1 Cluster validity indices for Breast dataset

Breast dataset				
Methods	Metrics			
	V_{pc}	V_{pe}	$V_{xb}[1 \times 10^{-3}]$	$V_{fs}[-1 \times 10^{6}]$
FCM	0.816	0.286	83.93	185.77
RSKFCM	0.852	0.169	78.60	192.58
GSKFCM	0.897	0.932	71.29	205.69
MIFCM	0.916	0.108	63.35	235.85

Table 2 Cluster validity indices for Brain dataset

Brain dataset				
Method	Metrics			
	V_{pc}	V_{pe}	$V_{xb}[1 \times 10^{-3}]$	$V_{fs}[-1 \times 10^{6}]$
FCM	0.812	0.221	81.26	296.20
RSKFCM	0.826	0.125	76.45	310.67
GSKFCM	0.846	0.105	69.36	335.85
MIFCM	0.874	0.098	58.65	342.69

Table 3 Cluster validity indices for Liver dataset

Liver dataset				
Method	Metrics			
	V_{pc}	V_{pe}	$V_{xb}[1 \times 10^{-3}]$	$V_{fs}[-1 \times 10^{6}]$
FCM	0.624	0.170	79.45	170.85
RSKFCM	0.695	0.126	74.23	186.93
GSKFCM	0.762	0.108	70.60	206.58
MIFCM	0.819	0.099	66.35	211.60

Table 4 Cluster validity indices for Prostate dataset

Prostate dataset				
Method	Metrics			
	V_{pc}	V_{pe}	$V_{xb}[1 \times 10^{-3}]$	$V_{fs}[-1 \times 10^{6}]$
FCM	0.713	0.155	80.46	190.50
RSKFCM	0.765	0.119	76.53	198.65
GSKFCM	0.810	0.101	72.15	210.96
MIFCM	0.845	0.089	69.35	215.15

4 Summary

In this article, we conducted a comparative analysis on four datasets on variants of the FCM methods. From the results of the experimentation, we found that the RSKFCM and GSKFCM are less noise sensitive. Furthermore, MIFCM methods handled the uncertainty better compared to RSKFCM and GSKFCM and achieved better results. The study revealed that the FCM approaches outperform several other approaches that exist. We also presented the evaluation metrics used to assess the performance of the methods employed. Since the methods in this article are clustering-based, we used cluster validity indices as common performance evaluation metrics for all four datasets to validate the clustering results.

References

1. J.C. Bezdek, *Pattern Recognition with fuzzy Objective Function Algorithms* (Springer, 2013)
2. A. Das, S.K. Sabut, Kernelized fuzzy C-means clustering with adaptive thresholding for segmenting liver tumors. Procedia Comput. Sci. **92**, 389–395 (2016)
3. I. El-Naqa, Y. Yang, M.N. Wernick, N.P. Galatsanos, R.M. Nishikawa, A support vector machine approach for detection of microcalcifications. IEEE Trans. Med. Imaging **21**(12), 1552–1563 (2002)
4. S. Ellmann, E. Wenkel, M. Dietzel, C. Bielowski, S. Vesal, A. Maier, M. Hammon, R. Janka, P.A. Fasching, M.W. Beckmann et al., Implementation of machine learning into clinical breast mri: Potential for objective and accurate decision-making in suspicious breast masses. Plos one **15**(1), e0228446 (2020)
5. M.D. Greer, N. Lay, J.H. Shih, T. Barrett, L.K. Bittencourt, S. Borofsky, I. Kabakus, Y.M. Law, J. Marko, H. Shebel et al., Computer-aided diagnosis prior to conventional interpretation of prostate mpmri: an international multi-reader study. Eur. Radiol. **28**(10), 4407–4417 (2018)
6. A.E. Hassanien, Th. Kim, Breast cancer mri diagnosis approach using support vector machine and pulse coupled neural networks. J. Appl. Log. **10**(4), 277–284 (2012)
7. P. Hebli P, S. Gupta, Brain tumor detection using image processing: a survey 1 amruta (2017)
8. G. Hu, Z. Du, Adaptive kernel-based fuzzy C-means clustering with spatial constraints for image segmentation. Int. J. Pattern Recogn. Artif. Intell. **33**(01), 1954003 (2019)
9. M.S.S. Hunnur, A. Raut, S. Kulkarni, Implementation of image processing for detection of brain tumors, in *2017 International Conference on Computing Methodologies and Communication (ICCMC)* (IEEE, 2017), pp. 717–722

10. S.A. Kumar, B.S. Harish, Segmenting MRI brain images using novel robust spatial kernel fcm (rskfcm), in *Eighth International Conference on Image and Signal Processing* (2014), pp. 38–44
11. S.V.A. Kumar, B.S. Harish, V.N.M. Aradhya, A picture fuzzy clustering approach for brain tumor segmentation, in *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)* (2016), pp. 1–6
12. H.M. Moftah, A.T. Azar, E.T. Al-Shammari, N.I. Ghali, A.E. Hassanien, M. Shoman, Adaptive K-Means clustering algorithm for mr breast image segmenta- tion. Neural Comput. Appl. **24**(7–8), 1917–1928 (2014)
13. A.A. Nahid, Y. Kong, Involvement of machine learning for breast cancer image classification: a survey. Comput. Math. Methods Med. (2017)
14. Y. Peng, Y. Jiang, C. Yang, J.B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M.L. Giger, S.E. Eggener, A. Oto, Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with gleason scorea computer-aided diagnosis development study. Radiology **267**(3), 787–796 (2013)
15. S. Radhakrishna, S. Agarwal, P.M. Parikh, K. Kaur, S. Panwar, S. Sharma, A. Dey, K. Saxena, M. Chandra, S. Sud, Role of magnetic resonance imaging in breast cancer management. South Asian J. Cancer **7**(2), 69 (2018)
16. B.V. Ramana, M.S.P. Babu, N. Venkateswarlu et al., A critical study of selected classification algorithms for liver disease diagnosis. Int. J. Database Manage. Syst. **3**(2), 101–114 (2011)
17. M. Ramasamy, S. Selvaraj, M. Mayilvaganan, An empirical analysis of decision tree algorithms: Modeling hepatitis data, in *2015 IEEE International Conference on Engineering and Technology (ICETECH)* (IEEE, 2015), pp. 1–4
18. Repositories OAMI: http://www.aylward.org/notes/open-access-medical-image-repositories. Accessed 12 Jan 2020 (2020)
19. C.K. Roopa, B.S. Harish, S.A. Kumar, A novel method of clustering ECG arrhythmia data using robust spatial kernel fuzzy C-means. Procedia Comput. Sci. **143**, 133–140 (2018)
20. V. Shah, B. Turkbey, H. Mani, Y. Pang, T. Pohida, M.J. Merino, P.A. Pinto, P.L. Choyke, M. Bernardo, Decision support system for localizing prostate can- cer based on multiparametric magnetic resonance imaging. Med. Phys. **39**(7Part1), 4093–4103 (2012)
21. H. Shahid, J.F. Wiedenhoefer, C. Dornbluth, P. Otto, K.A. Kist, An overview of breast MRI. Appl. Radiol. **45**(19), 7–13 (2016)
22. S. Vijayarani, S. Dhayanand, Liver disease prediction using SVM and Nave Bayes algorithms. Int. J. Sci. Eng. Technol. Res. (IJSETR) **4**(4), 816–820 (2015)
23. H. Wang, Y. Liu, W. Huang, Random forest and bayesian prediction for hepatitis b virus reactivation, in *2017 13th International Conference on Nat- ural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (IEEE, 2017), pp. 2060–2064
24. J. Wang, C.J. Wu, M.L. Bao, J. Zhang, X.N. Wang, Y.D. Zhang, Machine learning-based analysis of mr radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. Eur. Radiol **27**(10), 4082–4090 (2017)
25. T. Xia, A. Kumar, D. Feng, J. Kim, Patch-level tumor classification in digital histopathology images with domain adapted deep learning, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2018), pp. 644–647
26. S. Yoo, I. Gujrathi, M.A. Haider, F. Khalvati, Prostate cancer detection using deep convolutional neural networks. Sci. Rep. **9** (2019)
27. Y. Yuan, W. Qin, M. Buyyounouski, B. Ibragimov, S. Hancock, B. Han, L. Xing, Prostate cancer classification with multiparametric MRI transfer learning model. Med. Phys. **46**(2), 756–765 (2019)
28. A. Zotin, K. Simonov, M. Kurako, Y. Hamad, S. Kirillova, Edge detection in mri brain tumor images based on fuzzy C-means clustering. Procedia Comput. Sci. **126**, 1261–1270 (2018)

A Review on Effectiveness of AI and ML Techniques for Classification of COVID-19 Medical Images

M. J. Dileep Kumar, G. Santhosh, Prabha Niranjajn, and G. R. Manasa

Abstract During the global urgency, the scientific world is searching for new innovation and technology to tackle the pandemic. It is the necessity to develop a tool that will monitor and reduce human intervention in treating COVID-19. Major techniques/tools are available to control the current situation by diagnosing the patients with the help of machine learning (ML) and artificial intelligence (AI). These techniques have helped researchers in identifying/controlling the epidemic in past decades. The use of ML and AI has helped medical fraternity to significantly improve screening, testing, predicting and treating and helped in developing vaccines for COVID-19 patients. This article presents a view of how techniques like AI and ML are used for detecting, classifying and clustering of novel coronavirus disease 2019 by using the various types of medical images.

Keywords Pandemic · COVID-19 · Machine learning · Artificial intelligence · Clustering

1 Introduction

Due to the ability of adapting to various forms of data which are not restricted by various assumptions like forms of decision-making functions, probability of distribution of variables, etc., the machine learning (ML) is emerging as a popular paradigm for many recent scientific researches. This versatile approach has enabled the use of ML in various fields ranging from forecasting financial variables [1] to analysing the

M. J. D. Kumar (✉) · P. Niranjajn
Department of ECE, NMAM Institute of Technology (VTU, Belagavi), Nitte, Karnataka 574110, India
e-mail: dileepmj@nitte.edu.in

G. Santhosh
Department of Mechanical Engineering, NMAM Institute of Technology (VTU, Belagavi), Nitte, Karnataka 574110, India

G. R. Manasa
Department of CSE, NMAM Institute of Technology (VTU, Belagavi), Nitte, Karnataka 574110, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_14

emotions of texts better known as sentiment analysis [2] and medical applications [3, 4]. Most recently, various medical professionals and medical councils across the globe are looking into a way of using this approach to detect new infectious samples by making use of X-ray images [5] and also to identify how the virus is spreading in the timeline by making use of confirmed cases [6].

With their capabilities to condensate high-level complex pattern to medium-to-low level complex pattern by making use of nonlinear interaction on the data set, the models developed exhibit a high degree of sensitivity to variations in their hyperparameters. These parameters are defined by the user and are specific for one model; generally, these parameters are used for training the process or as the sample data set.

As discussed by Claesen and De Moor [7], identifying the ideal combination of these parameters or predicting which of these trivial parameters to be used for a said task is not always required. With the emergence of new global pandemic COVID-19, it was observed that there is an urgent requirement for understanding and classifying the real-time information. While doing so, one must emphasize on origin of data, relevance of the information and about the characteristics of the virus [8]. The work of analysing, classifying and identifying the credibility of the data is to be done at a faster pace as observed, and the effective identification and early detection of how the virus spreads and how it affects can if not help in eradicating the virus will definitely help in reducing the speed of spreading of the virus. As identified by Leon et al. [9], it is important to have a large data set in order to assist decision-making on COVID-19's infections, and this will help in identifying various strategies to mitigate and/or suppress the spread of virus across various locations.

During the early stages of this pandemic outbreak, various scientific communities have published their findings [10, 11]. In those results, majority of the symptoms are classified as not so severe, but the clinical findings of these patients have varied in great proportions [12–14]. This causes a threat of normalizing the effect of virus as same across all group of people if we were to make the classification based on the simple features like age and gender alone. Apart from identifying the presence of virus, it is also required to detect/assess out of these identified people with the virus who is most likely to get severe illness and will face a greater risk including death itself. These are the important variables; if detected early, it can help in effective utilization of available resources (hospital beds, medical mask, respirator, capacity of the hospital, etc.) to save those who are severely affected, and as these are restricted or less in number, medical practitioners are made to take decision about a patient without having any knowledge to guide them. Due to all these constraints, an artificial intelligence (AI)-aided system can help them to make a decision based on available information limited not just to one place but across the globe.

Nowadays, AI forms a critical factor/component in healthcare systems, wherein it actively provides support in making several critical clinical decisions [15–17]. Effectiveness of the ML to identify the medical conditions like epilepsy [18, 19], nerve and muscle diseases [20, 21] and heart rhythms [22, 23] was all well documented. Also several major researches have happened to identify new data sets for developing new classifiers. Another effective predicting method is deep learning techniques which

can be used in various clinical findings such as early detection of cancers, virus diseases and biomedical studies. Such techniques are proven to be efficient, and they can be used to early predict COVID-19 infection. With this as primary objective, this article presents several studies conducted by various medical professional bodies across the world and their results and accuracy in detecting the pandemic disease. This article is structured as follows: Sect. 2 discusses various case studies conducted by the researchers and their findings; Sect. 3 discusses the limitations and future developments.

2 Literature Review

COVID-19 has affected people in various ways. It has observed that majority of the people who are infected with the virus show the symptoms like fever, fatigue and dry cough [24], while others experience additional symptoms like aches and pains, nasal congestion, runny nose, sore throat and diarrhoea [25]. This pandemic has brought the healthcare system across the globe to its knee and also showed the inability of healthcare systems to manage the influx of patients in record number which resulted in the anxiety among the world population. The major cause for the rapid increase in the spread of COVID-19 is due to inability of early detection techniques or early diagnosis techniques [26]. Clinical approaches like real-time reverse transcription–polymerase chain reaction (rRT-PCR) [27] and other methods such as serologic tests [28] and viral throat swab testing [29] are necessary and widely utilized for the detection of COVID-19. But several research studies have suggested that chest radiographs (X-rays) [5] and chest computed tomography (CT) scans [30] can help in identifying the various types of lung-related anomalies including COVID-19. These medical imaging techniques can be used as an early detection tool to evaluate the degree of infection caused by COVID-19. Using these findings, one can monitor severely infected person and can predict the progression of disease [31]. But, the major constraint in such emergencies is the availability of time, which does not allow the medical examiner to run all these tests using the available traditional diagnosis [32]. Generally, this diagnosis technique requires an expert physician, but usually, these findings are prone to minor human error while taking the values or while reading or interpreting these values, but such errors even though are minor are not acceptable because such might cost a person's life. Due to the speed at which this COVID-19 has spread across the globe, all the hospitals are flooded with the patients; the conditions of these patients are either recovering from the infection or becoming severe (dying) [33]. In this case, CT scan and X-ray tests should be performed with maximum speed and efficiency to save as many lives as possible [31]. In such cases, a technology-enabled intelligent system which assists in diagnosing and classifying the patients would definitely help in reducing the burden caused on doctors and technicians. Also such intelligent systems would avoid unnecessary loading of hospitals which otherwise would have been occupied by those who does not require the critical care units [5]. The application of AI has been found in various fields, mainly in medical

field wherein it is used as a tool to detect various diseases [34]. AI is used in order to get a better result or accurate detection of the disease and hence reduce strain on healthcare system [35]. By deploying AI in the field of healthcare system, it helps in reducing the time required to detect the disease if any when compared to traditional techniques [36]. As suggested in [37] in order to prevent future global health crises, we must develop an AI-enabled system which can be used to early recognize or detect the risk of epidemic diseases. Several studies have been carried out by researchers using the dataset of COVID-19 factoring in various case studies and targets [31]. Even though this software-enabled detection technique can be proven beneficial to detect and classify the COVID-19, choosing an suitable AI technique is what the major criterion which poses challenge for successful implementation of these techniques [38, 39]. The availability of diverse AI techniques is what making the task of selecting one among them to diagnose and classify the COVID-19 data a difficult procedure, as there is no profound AI method which is comparatively good from another. Along with this, most of these methods are less accurate but have least computational efficiency [40]. Another major hurdle in adopting these is associated with the method of evaluation and comparison due to availability of multiple evaluation criteria, and these criteria often have conflicts among them which further increase the complexity in selecting a suitable evaluation criterion [41]. The main goal of reviewing the articles on the topic of ML/AI techniques is to understand how the identification and classification of medical images targeted for COVID-19 can be implemented using AI/ML. Also while doing so, one can understand of what the need for research in this area and which topics are over-studied or under-evaluated.

In [31], authors reviewed the response of AI-empowered medical imaging techniques applied for COVID-19. The method of acquiring the image using the AI technique will help greatly in automating the procedure thereby reducing the contact of the medical personal with the patients and thus helping in preventing the spread of virus for these laboratory technicians. This study is mainly focused on all the process involved in diagnosing COVID-19 like medical imaging, segmentation, diagnostic report generation and finally suggesting for follow-ups to the clinic by making integrating AI with X-ray and CT images. In the study [42], it was shown that several deep learning methods can be used for diagnosing COVID-19 with the help of medical images obtained through various imaging techniques. In the decision-making confusion matrix, the authors relied on the class labels to identify the presence of infection. The identification of the disease is done on the rate of false-negative (FN) results, and this assumption affects the decision-making ability like whether to keep monitoring the patients or to discharge them. In this study, authors relied on the data set obtained by the available information from ten patients. But the results showed that out of ten negatively identified cases, two were identified as positive for COVID-19 by making use of rRT-PCR method of testing, thus resulting in 20% FN rate. Hence while designing an automated method or to create dataset for ML techniques, one must consider more number of patient information thereby reducing the FN rate of detection.

In the work [43], a scoring tool was designed targeted to understand the severity of COVID-19. The need for such tool is to provide necessary aide for the healthcare

professionals in identifying and determining whether patients are suspected with or confirmed with COVID-19 infection and whether such patients are in need of respiratory interventions. In the study, the researchers followed the guidelines set by WHO in classifying the severity as mild/moderate and severe. It was identified that for a patient who was diagnosed at critical are the ones who require the ventilation, and the category of people who are diagnosed under severe category required the administration of oxygen, whereas the patients diagnosed as moderate level need not have to be administered with oxygen even though they show the symptoms of pneumonia. In this study, the researchers have utilized the information of 13,500 COVID-19 patients to form necessary data sets and to develop required decision-making matrices. The study showed the classification at 93.6%, also the underestimated value at 0.8% of patients and 5.7% of patients as overestimated. Another study involving the deep learning method for identification of COVID-19 using MobileNetV2 and Squeeze Net was carried out. Along with feature sets gathered by the deep learning method, the data was processed by applying social mimic optimization technique. This technique also made use of preprocessing technique to improve the quality of detection.

The preprocessing was done using fuzzy colour technique, thereby restructuring data classes, and structured images were stacked with the original images. After that, the required features were grouped together and then classified with the help of support vector machines (SVMs). This technique shows to be getting the rate of classification at 99.27%. A system including convolutional neural network along with patch-based technique was studied in [44]. In this, parameters are designed by making use of small numbers of training parameters to diagnose the COVID-19. This work emphasizes on the statistical study of biomarkers on the chest X-ray images, which along with identifying COVID-19 can potentially be helpful in detecting any other medical conditions also. The proposed method involves preprocessing of the images from cross-database, thus improving the accuracy of segmentation. From the result, it can be observed that preprocessing of database will result in increasing the performance of effective image segmentation. COVIDiagnosis-Net [45], an AI-enabled detection technique for COVID-19, was developed based on Bayes optimization technique. It makes use of deep learning technique to detect and classify the patients with the symptoms of pneumonia. This study showed the result with an accuracy of 98.3% in classifying the cases as normal, pneumonia and COVID-19. Using various lung images and considering multi-class and hierarchical approach, a classification method was designed in [46], to identify COVID-19. This technique used re-sampling and re-balancing for class distribution. This re-sampling method has improved macro-average F1 score to 0.89 from the earlier score of 0.65 with the use of multi-class method for the detection of COVID-19 in hierarchical classification scenario.

Classification models for AI techniques were developed using 3D imaging technique in [47], wherein one model considered entire lung region and another model considered mean score obtained by considering multiple regions inside the lung image with a fixed image resolution. The images were trained as full 3D: an original lung image with fixed resolution and hybrid 3D and a segmentation of original image with the fixed length. The training set considering full 3D showed the accuracy of

91.7%, while the hybrid model showed an accuracy of 92.4%. The other classification is detecting the other disease vs. COVID-19; the most accuracy was achieved when 3D model for classification was used; this model results in 90.8% of accuracy. While this model predicts the classification at the highest rate, there are several limitations to the proposed model, the major limitation is the dataset used for training, and all the data used for training are of those which have returned positive results from RT-PCR tests. But it was observed that the CT results were often found to be negative even when a person is found positive by RT-PCR test. So one cannot decide based on CT images only whether a person is infected with COVID-19 or not. In conclusion, using multi-national data set for training, one can develop an AI system capable of delivering a classification model with acceptable performance metrics for classifying CT images for COVID-19 disease. Also it is observed that CT image alone cannot be used to predict the COVID-19, and it can help in the assessment of the findings thereby helping in developing the variable to design a tool for AI approach. As proven by many studies, deep learning techniques have achieved far greater success in the field of radiology. In [48], a 3D deep learning framework was suggested to detect COVID-19. Using the proposed method, extracting both local and global features from the image was achieved. Various studies have applied and obtained the desirable results in detecting pneumonia in paediatric chest radiographs by making use of deep learning technique and further to differentiate viral and bacterial pneumonia in 2D paediatric chest radiographs. The study has been carried out by collecting a large number of CT samples from different locations, covering more than 1296 COVID-19 CT images. Along with those many positive samples, the data set also included 1735 CAP and 1325 non-pneumonia CT images as the control groups in the study, so that the model will be able to efficiently detect, differentiate and identify COVID-19 and other type of lung diseases. In this study, a deep learning model was designed and evaluated for detecting COVID-19 from chest CT images. On an independent testing data set, the model achieved high sensitivity (90% [95% CI: 83, 94%] and high specificity of 96% [95% CI: 93, 98%] in detecting COVID-19. The AUC values for COVID-19 and community acquired pneumonia (CAP) were 0.96 [95% CI: 0.94, 0.99] and 0.95 [95% CI: 0.93, 0.97], respectively.

3 Conclusion

The effect of COVID-19 has been felt at every corner of the world. All governments across the globe and various scientific communities are looking into a way to mitigate this new-found pandemic. Already various forms of medical examinations like X-rays and CT scans are conducted to detect the virus. The medical practitioners across the globe are limited with the knowledge about COVID-19; a computerized program can help in making the critical decisions. As observed in the study, an attempt is already being done to develop suitable model which can be used to monitor and predict the progression of the disease in patients.

References

1. B.M. Henrique, V.A. Sobreiro, H. Kimura, Literature review: machine learning techniques applied to financial market prediction. Expert Syst. Appl. **124**, 226–251 (2019)
2. J. Singh, G. Singh, R. Singh, Optimization of sentiment analysis using machine learning classifiers. HCIS **7**(1), 32 (2017)
3. M. Motwani, D. Dey, D.S. Berman, G. Germano, S. Achenbach, M.H. Al-Mallah, D. Andreini, M.J. Budoff, F. Cademartiri, T.Q. Callister et al., Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur. Heart J. **38**(7), 500–507 (2017)
4. J.A. Sidey-Gibbons, C.J. Sidey-Gibbons, Machine learning in medicine: a practical introduction. BMC Med. Res. Methodol. **19**(1), 64 (2019)
5. T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of covid-19 cases using deep neural networks with x-ray images. Comput. Biol. Med. 103792 (2020)
6. V.K.R. Chimmula, L. Zhang, Time series forecasting of covid-19 transmission in canada using lstm networks. Chaos Solit. Fract. 109864 (2020)
7. M. Claesen, B. De Moor, Hyperparameter search in machine learning. arXiv:1502.02127 (2015)
8. P. Song, T. Karako, Covid-19: real-time dissemination of scientific information to fight a public health emergency of international concern. Biosc. Trends (2020)
9. D.A. Leon, V.M. Shkolnikov, L. Smeeth, P. Magnus, M. Pechholdová, C.I. Jarvis, Covid-19: a need for real-time monitoring of weekly excess deaths. Lancet **395**(10234), e81 (2020)
10. K. Liu, Y.-Y. Fang, Y. Deng, W. Liu, M.-F. Wang, J.-P. Ma, W. Xiao, Y.-N. Wang, M.-H. Zhong, C.-H. Li et al., Clinical characteristics of novel coronavirus cases in tertiary hospitals in Hubei province. Chin. Med. J. (2020)
11. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet **395**(10223), 497–506 (2020)
12. J. Cai, J. Xu, D. Lin, L. Xu, Z. Qu, Y. Zhang, H. Zhang, R. Jia, X. Wang, Y. Ge et al., A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features. Clin. Infect. Diseases (2020)
13. K. Kam, C. Yung, L. Cui et al., A well infant with coronavirus disease 2019 (covid-19) with high viral load. Clin Infect Dis. **10**
14. Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, M. Wang, Presumed asymptomatic carrier transmission of covid-19. JAMA **323**(14), 1406–1407 (2020)
15. F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, Artificial intelligence in healthcare: past, present and future. Stroke Vasc. Neurol. **2**(4), 230–243 (2017)
16. T. Davenport, R. Kalakota, The potential for artificial intelligence in healthcare. Future Healthcare J. **6**(2), 94 (2019)
17. S. Reddy, J. Fox, M.P. Purohit, Artificial intelligence-enabled healthcare delivery. J. R. Soc. Med. **112**(1), 22–28 (2019)
18. T.B. Alakus, I. Turkoglu, Detection of pre-epileptic seizure by using wavelet packet decomposition and artifical neural networks, in *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)* (IEEE, 2017), pp. 511–515
19. N. Memarian, S. Kim, S. Dewar, J. Engel Jr., R.J. Staba, Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. Comput. Biol. Med. **64**, 67–78 (2015)
20. J. Yousefi, A. Hamilton-Wright, Characterizing emg data using machine-learning tools. Comput. Biol. Med. **51**, 1–13 (2014)
21. P. Karthick, D.M. Ghosh, S. Ramakrishnan, Surface electromyography based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms. Comput. Methods Programs Biomed. **154**, 45–56 (2018)
22. M. Alfaras, M.C. Soriano, S. Ortín, A fast machine learning model for ecg-based heartbeat classification and arrhythmia detection. Front. Phys. **7**, 103 (2019)

23. C.A. Ledezma, X. Zhou, B. Rodriguez, P. Tan, V. Diaz-Zuccarini, A modeling and machine learning approach to ECG feature engineering for the detection of ischemia using pseudo-ECG. PloS One **14**(8), e0220294 (2019)
24. S. Kooraki, M. Hosseiny, L. Myers, A. Gholamrezanezhad, Coronavirus (covid-19) outbreak: what the department of radiology should know. J. Am. Coll. Radiol. (2020)
25. Y. Wang, Y. Wang, Y. Chen, Q. Qin, Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (covid-19) implicate special control measures. J. Med. Virol. **92**(6), 568–576 (2020)
26. T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. Radiology 200642 (2020)
27. Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest CT for covid-19: comparison to RT-PCR. Radiology 200432 (2020)
28. H. Zeng, C. Xu, J. Fan, Y. Tang, Q. Deng, W. Zhang, X. Long, Antibodies in infants born to mothers with covid-19 pneumonia. JAMA **323**(18), 1848–1849 (2020)
29. D.A. Schwartz, An analysis of 38 pregnant women with covid-19, their newborn infants, and maternal-fetal transmission of sars-cov-2: maternal coronavirus infections and pregnancy outcomes. Arch. Pathol. Lab. Med. **144**(7), 799–805 (2020)
30. M. Li, P. Lei, B. Zeng, Z. Li, P. Yu, B. Fan, C. Wang, Z. Li, J. Zhou, S. Hu et al., Coronavirus disease (covid-19): spectrum of CT findings and temporal progression of the disease. Acad. Radiol. (2020)
31. F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Rev. Biomed. Eng. (2020)
32. H.S. Maghdid, A.T. Asaad, K.Z. Ghafoor, A.S. Sadiq, M.K. Khan, Diagnosing covid-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. arXiv: 2004.00038 (2020)
33. World Health Organization et al., Coronavirus disease 2019 (covid-19): situation report. 72 (2020)
34. D.D. Miller, E.W. Brown, Artificial intelligence in medical practice: the question to the answer? Am. J. Med. **131**(2), 129–133 (2018)
35. K.-H. Yu, A.L. Beam, I.S. Kohane, Artificial intelligence in healthcare. Nat. Biomed. Eng. **2**(10), 719–731 (2018)
36. A. Albahri, R.A. Hamid et al., Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (covid-19): A systematic review. J. Med. Syst. **44**(7) (2020)
37. Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai et al., Modified SEIR and AI prediction of the epidemics trend of covid-19 in China under public health interventions. J. Thorac. Dis. **12**(3), 165 (2020)
38. M. Alsalem, A. Zaidan, B. Zaidan, M. Hashim, O. Albahri, A. Albahri, A. Hadi, K. Mohammed, Systematic review of an automated multiclass detection and classification system for acute leukaemia in terms of evaluation and benchmarking, open challenges, issues and methodological aspects. J. Med. Syst. **42**(11), 204 (2018)
39. M. Alsalem, A. Zaidan, B. Zaidan, O. Albahri, A. Alamoodi, A. Albahri, A. Mohsin, K. Mohammed, Multiclass benchmarking framework for automated acute leukaemia detection and classification based on BWM and group-VIKOR. J. Med. Syst. **43**(7), 212 (2019)
40. A. Zaidan, B. Zaidan, M. Alsalem, O. Albahri, A. Albahri, M. Qahtan, Multi-agent learning neural network and Bayesian model for real-time IoT skin detectors: a new evaluation and benchmarking methodology. Neural Comput. Appl. **32**(12), 8315–8366 (2020)
41. A. Zaidan, B. Zaidan, O. Albahri, M. Alsalem, A. Albahri, Q.M. Yas, M. Hashim, A review on smartphone skin cancer diagnosis apps in evaluation and benchmarking: coherent taxonomy, open issues and recommendation pathway solution. Health Technol. **8**(4), 223–238 (2018)
42. D. Li, D. Wang, J. Dong, N. Wang, H. Huang, H. Xu, C. Xia, False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome

coronavirus 2: role of deep-learning-based CT diagnosis and insights from two cases. Korean J. Radiol. **21**(4), 505–508 (2020)
43. L.A. Wallis, Covid-19 severity scoring tool for low resourced settings. Afr. J. Emergency Med. (2020)
44. Y. Oh, S. Park, J.C. Ye, Deep learning covid-19 features on CXR using limited training data sets. IEEE Trans. Med. Imaging (2020)
45. F. Ucar, D. Korkmaz, Covidiagnosis-net: deep Bayes-squeezenet based diagnostic of the coronavirus disease 2019 (covid-19) from X-ray images. Med. Hypotheses 109761 (2020)
46. R.M. Pereira, D. Bertolini, L.O. Teixeira, C.N. Silla Jr, Y.M. Costa, Covid-19 identification in chest X-ray images on flat and hierarchical classification scenarios. Comput. Methods Programs Biomed. 105532 (2020)
47. S.A. Harmon, T.H. Sanford, S. Xu, E.B. Turkbey, H. Roth, Z. Xu, D. Yang, A. Myronenko, V. Anderson, A. Amalou et al., Artificial intelligence for the detection of covid-19 pneumonia on chest CT using multinational datasets. Nat. Commun. **11**(1), 1–7 (2020)
48. L. Li, L. Qin, Z. Xu, Y. Yin, X.Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song et al., Using artificial intelligence to detect covid-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology **296**(2) (2020)

Medical Image Encryption Using SCAN Technique and Chaotic Tent Map System

Kiran, B. D. Parameshachari, and H. T. Panduranga

Abstract In today's digital world, security during medical image transmission is very important. The proposed method uses the concept of SCAN method and chaotic tent map for medical image encryption. It consists of bit plane decomposition, pixel rearranging stage and diffusion operation. Amount of information in the image is distributed into eight binary planes with a different proportional. Higher four bit planes are shuffled using spiral SCAN pattern, and lower bit planes are shuffled using diagonal SCAN pattern. Then perform pixel wise xor operation between shuffled image and the random image generated from the chaotic tent map. Experimental results show that proposed method achieves high level of security with lesser computation.

Keywords Medical image · SCAN · Tent map · Bit planes · Security

1 Introduction

Transmission of images through a communication channel requires high security which is attained by advanced encryption techniques. As the major communication is done through digital media, the importance for the security of digital images has been increased. In many applications such as military image database, medical imaging systems, confidential video conferencing, Pay-TV, etc., high end security is required for transmission as well as storing the digital images. Many consumer electronic devices such as mobile phones, tabs, laptops have an option of saving and transmitting the images with high end security on both ends via wireless networks by using service such as multimedia messaging.

Kiran (✉)
Department of ECE, Vidyavardhaka College of Engineering, Mysuru, India

B. D. Parameshachari
Department of Telecommunication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, India

H. T. Panduranga
Department of ECE, Govt. Polytechnic, Turvekere, Tumkur, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_15

Akkasaligar et al. [1] explained DNA-based medical image encryption. Medical image pixels are encoded into DNA sequence then performed by permutation and diffusion operation. Zhang [2] proposed a fast image encryption scheme with the help of lifting scheme and chaos. Janani et al. [3] ROI-based secure image encryption has been proposed and which uses the concept of ROI encryption. Low security level due to only ROI part encryption. Akkasaligar et al. [4] DNA-based approach has been proposed for providing security to images. Zhang [5] proposed "an exchange and random access strategy" is employed to replace the traditional confusion operations. Ali and Ali [6] proposed conventional approach to encrypt the image along with Boolean operation. Anwar and Meghana [7] chaotic map-based image encryption scheme has been proposed. Ramasamy et al. [8] proposed block Scrambling and Modified Zigzag SCAN method for image encryption. Liu et al. [9] proposed a scheme for medical image encryption based on this new chaotic system. Khan and Masood [10] explained about multiple chaotic system-based image encryption scheme. For the confusion and diffusion operation, different kinds of chaotic system are employed. Nezhad et al. [11] proposed an hybrid method for encrypting an image based on DNA and tent mapping techniques. Rashmi and Supriya [12] describe encryption algorithm make use the technique of permutation and diffusion process. Dr. Parameshachari et al. [13] explained about controlling the amount of encryption using LSIC. In the exiting work, many algorithms use the concept of bit level permutation and ROI-based approach. So, due to this existing method suffers from low security level and to overcome proposed an high security for medical images.

Organization of the paper is as follows. Overview of SCAN basics and tent map are explained in Sect. 2. Working of proposed system as shown in Sect. 3. Performance parameters and their analysis for proposed work described in Sect. 4. Section 5 concludes the proposed work.

2 SCAN Method

SCAN method transforms 2D image to 1D form and those transformed results will be applied by SCAN language. There are many SCAN letters in this language. There are different SCAN orders representations for each SCAN letter. We can generate different secret images using different SCAN letters combinations. The SCAN string has to be generated once the SCAN letters combination is determined. The original image SCAN order is defined by these strings. Further, in the determined order, the original image is scanned by this method and also, the SCAN string is encrypted by commercial cryptosystems. The original image will be secured as the correct SCAN string cannot be obtained by illegal users. The image size is very large as there is no image compression technique used here. Therefore, it is very difficult to directly decrypt or encrypt the image.

The SCAN is a formal languages family which depends on 2D spatial accessing methodologies that can generate and represent a huge number of paths of scanning. Generalized SCAN, Extended SCAN and Simple SCAN are the formal languages

SCAN family that generate and represent specific scanning paths. The grammar rules which define the SCAN language are used to define simple patterns of SCAN, and the same is utilized to generate the complex SCAN patterns with the help of these simple SCAN patterns.

2.1 Chaotic Map

The chaos, one of the latest and basic modern science concepts, is the new concept which makes us understand better of our universe. The random and unorganized behavior of the real world events gives the meaning for the word chaos. For the information encryption, the new strategies are designed for chaotic dynamic systems which are widely used since 1990s. To be precise, the control parameters and the initial condition dependence with the growth of organic nature give a very good chance to use chaos for the cryptographic systems as a basis. Equation for chaotic is given by

$$X(n+1) = r * X(n) * (1 - X(n)) \tag{1}$$

where X is initial condition and its range is [0, 1], r is control parameter. If the system is chaotic system, then r value is in the range [3.5, 4].

2.2 Tent Map

The following equation gives the expression for Tent map:

$$X_{n+1} = \begin{cases} aX_n, 0 \leq X_n < \dfrac{1}{2} \\ a(1 - X_n), \dfrac{1}{2} \leq x < 1 \end{cases} \tag{2}$$

where the initial value is given by $\boldsymbol{X}0$ and $\boldsymbol{X}0\ \varepsilon$ [0, 1]. A control parameter is between (0 and 1) and it is $a/ = 0.5$ when the system is chaotic. The tent map function is as shown in Fig. 2.

3 Proposed Method

Block diagram for proposed medical image encryption as shown in Fig. 3.

Original image having an m number of rows and n number columns undergo bit plane decomposition process. Where original image is decomposed into eight binary

Fig. 1 **a** Basic SCAN pattern, **b** partition patterns and transformation

planes represented as B0, B1, B2, B3, B4, B5, B6, B7. B0 represents least significant bit plane and b7 represents most significant bit plane. Amount of information distributed in each plane as shown in Table 1. From Table 1, around 94% information presents in higher four bit planes and approximately 6% information presents in lower bit planes. Rearranging of the higher and lower bit planes as shown in Fig. 4. Then, pixels in the expanded higher four bit planes and lower four bit planes are shuffled

Fig. 2 Tent map signal with $a \in [0, 1]$

Fig. 3 Block diagram of proposed medical image encryption system

Table 1 Percentage of pixel information contributed by different bits

Bit position (i) in the pixel	Percentage $p(i)$ of the pixel information
1	0.3922
2	0.7843
3	1.5686
4	3.137
5	6.275
6	12.55
7	25.10
8	50.20

Fig. 4 Bit plane rearranging process

according to spiral SCAN pattern and diagonal bit pattern, respectively. Generate the chaotic sequence from chaotic tent map with initial condition. Later, these sequences are converted into integer value with the range 0–255 using following Eq. 3. Integer sequence converted into matrix of size $m * n$ that image matrix called as random image. Finally, encrypted image is obtained by performing pixel wise xor operation between bit plane permuted image and random image generated from tent map.

$$\text{Int_ sequence} = \text{mod}((\text{tent_ map sequence} * (10^9)), 256) \tag{3}$$

4 Performance Analysis for Proposed Work

4.1 Histogram Analysis

Image histogram graphically represents the distribution of pixels based on their occurrence. For highly secured encryption technique, histogram of cipher image should be uniformly distributed.

4.2 Entropy

Entropy [14] gives the measure of information randomness in the image. The equation for entropy is given by

$$H(S) = \sum_{i=0}^{2^M-1} P(si) \log_2 \frac{1}{P(si)} \tag{4}$$

Ideal value of entropy is 8 for a random image. If it is less, the chance of predictability is more.

4.3 Mean Square Error

Mean squared error (MSE) [14] is defined as an average of the square of the difference between plain image and encrypted image. The MSE is given by the equation

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [X(i, j) - Y(i, j)]^2 \tag{5}$$

4.4 Number of Pixel Change Rate (NPCR)

Plain image and encrypted image represented by C_1 and C_2, respectively. $C_1(i, j)$ and $C_2(i, j)$ are original image pixel and encrypted image pixel, respectively. The NPCR [15] is then defined as,

$$\text{NPCR} = \frac{\sum_{i,j} D(i, j)}{M \times N} \times 100\% \tag{6}$$

where D is bipolar array.

$$D(i, j) = \begin{cases} 1, & C_1(i, j) \neq C_2(i, j) \\ 0, & \text{otherwise} \end{cases}$$

4.5 Peak Signal to Noise Ratio (PSNR)

The peak signal to noise ratio is evaluated in decibels and is inversely proportional to MSE. It is given by the equation

$$\text{PSNR} = 10 \log_{10} \frac{255}{\text{MSE}} \tag{7}$$

4.6 Unified Average Changed Intensity (UACI)

It is used measure the intensity rate difference between the plain image and cipher image [15].

$$\text{UACI} = \frac{1}{N} \left[\sum_{i,j} \frac{|C_1(i, j) - C_2(i, j)|}{255} \right] \tag{8}$$

4.7 Experimental Results

Here, the proposed work has been evaluated and validated with its performance. We used MATLAB 2017b for the purpose of simulation. The simulation is carried out using one set of medical images as test images. We took these medical images from the database of Open-i image. The medical images size we used is 256 × 256. The medical images used for the simulation purpose are in the Table 2.

In our proposed scheme, Table 3 illustrates the histograms of the plain images, cipher images and decrypted encrypted image are evenly distributed, similar to white noise.

Table 4 illustrates the original image, encrypted image and decrypted image. In the encrypted images clearly shows that entire image is encrypted so that will not give any information about original image.

Table 2 Medical images used in the proposed system

Medical Image 1	**Medical Image 2**	**Medical Image 3**
Medical Image 4	**Medical Image 5**	

From Tables 5, 6 and 7 provides the performance parameter analysis of proposed work. Entropy values of encrypted images are very close to ideal value. MSE, PSNR, NPCR and UACI values are calculated between original image and encrypted image whose values also meet the security requirements, which shows that our encryption scheme meets the robustness requirements against differential attacks.

Table 8 provides the comparison of proposed work with existing method in terms of performance analysis like entropy, NPCR and UACI. From the parameter values, it is observed that proposed method achieves better values as compared to existing work.

5 Conclusion

In this paper, we proposed a medical image encryption based on SCAN pattern and tent map. Proposed method mainly consists of bit plane decomposition, SCAN-based shuffling and diffusion operation. SCAN method applied to higher and lower four bit planes separately to increase the randomness of the image. Then, xor operation is involved for the diffusion operation. Experimental results show that proposed method achieves higher security in terms of entropy, statistical analysis and differential attacks.

Table 3 Histogram analysis of proposed system

Medical images	input image histogram	Encrypted image histogram	Decrypted image histogram

Table 4 Original images, encrypted image and decrypted image of proposed system

Medical images	Encrypted image	Decrypted image

Table 5 Entropy analysis of proposed method

Medical images	Entropy of original image	Entropy of encrypted image
Medical image1	3.5963	7.9965
Medical image2	5.0030	7.9972
Medical image3	4.2654	7.9970
Medical image4	6.8129	7.9974
Medical image5	5.0923	7.9969

Table 6 NPCR and UACI of proposed method

Medical images	NPCR (%)	UACI (%)
Medical image1	99.6658	43.9856
Medical image2	99.6445	41.3774
Medical image3	99.6292	42.0747
Medical image4	99.6307	40.4066
Medical image5	99.6445	43.0771

Table 7 MSE and PSNR analysis of proposed method

Medical images	MSE	PSNR (db)
Medical image1	33.8168	22.8395
Medical image2	30.7966	23.2458
Medical image3	57.8667	20.5065
Medical image4	64.2554	30.0517
Medical image5	26.9480	24.8255

Table 8 Comparative analysis

Image	Proposed method			Existing method [16]		
	Entropy	NPCR	UACI	Entropy	NPCR	UACI
Medical image 2	7.9972	99.644	41.37	7.0425	99.56	33.46

References

1. Akkasaligar, T. Prema, S. Biradar, Medical image encryption with integrity using DNA and chaotic map, in *International Conference on Recent Trends in Image Processing and Pattern Recognition* (Springer, Singapore, 2018)
2. Y. Zhang, The fast image encryption algorithm based on lifting scheme and chaos. Inf. Sci. **520**, 177–194 (2020)
3. T. Janani, Y. Darak, M. Brindha, Secure similar image search and copyright protection over encrypted medical image databases. IRBM (2020)
4. P.T. Akkasaligar, S. Biradar, Selective medical image encryption using DNA cryptography. Inf. Secur. J.: Glob. Perspect. **29**(2), 91–101 (2020)

5. W. Zhang, Yu. Hai, Z.-L. Zhu, Color image encryption based on paired interpermuting planes. Optics Commun. **338**, 199–208 (2015)
6. T.S. Ali, R. Ali, A new chaos based color image encryption algorithm using permutation substitution and Boolean operation, in *Multimedia Tools and Applications* (Springer, Berlin/Heidelberg, Germany, 2020), pp. 1–21
7. S. Anwar, S. Meghana, A pixel permutation based image encryption technique using chaotic map. Multimedia Tools Appl. **78**(19), 27569–27590 (2019)
8. P. Ramasamy et al., An image encryption scheme based on block scrambling, modified zigzag transformation and key generation using enhanced logistic—Tent map. Entropy **21**(7), 656 (2019)
9. J. Liu et al., A new simple chaotic system and its application in medical image encryption. Multimedia Tools Appl. **77**(17), 22787–22808 (2018)
10. M. Khan, F. Masood, A novel chaotic image encryption technique based on multiple discrete dynamical maps. Multimedia Tools Appl. **78**(18), 26203–26222 (2019)
11. S.Y.D. Nezhad, N. Safdarian, S.A.H. Zadeh, New method for fingerprint images encryption using DNA sequence and chaotic tent map. Optik 165661 (2020)
12. P. Rashmi, Dr. M.C. Supriya, Encryption of color image to enhance security using permutation and diffusion techniques. Int. J. Adv. Sci. Technol. **28**(12), 375–384 (2019)
13. B.D. Parameshachari, R.P. Kiran, M.C. Supriya, Rajashekarappa, H.T. Panduranga, Controlled partial image encryption based on LSIC and chaotic map, in *ICCSP* (2019), pp. 60–63
14. Y. Wu, J.P. Noonan, S. Agaian, NPCR and UACI randomness tests for image encryption. Cyber J.: Multidisc. J. Sci. Technol. J. Sel. Areas Telecommun. 31–38 (2011)
15. J. Ahmad, F. Ahmed, Efficiency analysis and security evaluation of image encryption schemes. Int. J. Video Image Process. Netw. Secur. **12**, 18–31 (2012)
16. K.N. Madhusudhan, P. Sakthivel, A secure medical image transmission algorithm based on binary bits and Arnold map. J. Ambient Intell. Human Comput. (2020). https://doi.org/10.1007/s12652-020-02028-5

Utilization of Dark Data from Electronic Health Records for the Early Detection of Alzheimer's Disease

Sonam V. Maju and O. S. Gnana Prakasi

Abstract Information management and the presence of dark data is the most alarming and vulnerable topic that needs to be addressed amidst the rapid growth of technology. The unused, untapped, or unstructured data in our archives can be called as dark data. Under the HIPPA regulations, the clinical data needs to be stored and secured for years, so the amount of dark data in healthcare databases consumes a large amount of storage space. The paper shows the comparative performance of the random forest algorithm with and without dark data from the patient's health record for the early detection of Alzheimer's disease. The model executes in two ways, (1) considering only the Alzheimer's disease parameters and (2) including diabetes disease parameters which are considered as dark data for the Alzheimer's disease. The results of the research work clearly say that utilizing the health dark data has increased the accuracy by 16.3% and hence helps in making better decisions.

Keywords Dark data · Electronic health record · Machine learning · Random forest classifier

1 Introduction

We have discovered, designed, and invented supercomputers and expert systems to solve the complexity of every problem a human has to face in his/her daily life. The amount of data thus collected, processed, analyzed, and stored in our databases is beyond calculation and consumes more storage resources. Dark data is the data thus stored in our databases in a loose format, and we as users fail miserably to make use of it for other purposes. Dark data does not have any one true definition [1]; however, in reviewing the literature, common themes are found. Gartner originally

S. V. Maju (✉) · O. S. Gnana Prakasi
Computer Science and Engineering, School of Engineering and Technology, CHRIST (Deemed to be University), Bengaluru, India
e-mail: sonam.maju@res.christuniversity.in

O. S. Gnana Prakasi
e-mail: gnana.prakasi@christuniversity.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_16

coined the term 'Dark Data' and outlined it as, any data that is collected for a specific purpose but not used for other suitable purposes as well. Gartner describes dark data as every information asset that organizations/institutions collect, process, and store throughout the regular business activities, but usually fails to use for other purposes.

Majority of the dark data can be further labeled as unstructured content. Factually, 90% of all the information used in an organization on an everyday basis is in the unstructured format, and its escalating rate is much faster than structured data. So, dark data can be any data available and stored in our databases but not being used for multiple purposes.

Presently, almost every medical institution mostly depends on electronic health records (EHR) or the electronic medical record (EMR) to monitor a patient's present condition, to record diagnostic evidence, procedures performed on the patient, and past treatment results. EHR is one of the valuable resources for large-scale medical data analysis [2]. The development of technology and hospital information system is giving high importance and also been widely promoting EMR/EHR thus considered as a treasure for large-scale medical analysis of clinical/health data. This EMR can be classified into three kinds: structured data, semistructured data, and unstructured data. Structured data is the data stored that is easily available in databases in a particular framework which includes basic information like birth details, medications, vital signs, etc. Semistructured data will have a flow chart format, like resource description files, name, value, and time-stamp. Unstructured data is considered as one of the loosely structured data with no framework, and at the same time, it carries highly valuable information including handwritten clinical notes, surgical images, discharge reports, radiology, and pathology reports. By the conversion of these unstructured data to structured one, a huge amount of data can be made available for research analysis and learning from the last 20 years by the amalgamation of technology and health care. The results literally saved and are saving millions of lives as the diseases can be diagnosed early through data analysis on machine learning models and treatment can be done forehand.

Alzheimer's is one such disease that can be predicted in a very early stage considering different parameters. Along with these parameters, this research work makes it possible that early detection of diseases is possible by identifying the relationship between diseases and the presence of one disease can trigger the occurrence of another disease.

In this paper, dark data is extracted from the health records and is used to derive efficient and useful conclusions like identifying the relationship between Alzheimer's and its prediction at a very early stage using dark data.

2 Review of Literature

Dark data is an emerging term; the existence, identification, and utilization of dark data are one of the prominent research domains in the future.

Every electronically stored information (ESI) will come under focus to legal discovery if a threat of litigation arises, even in incomplete or trivial data [3]. The untagged/untapped or forgotten data called dark data can bring huge security risks if we did not manage it properly. The study [3] reveals the hidden risk of selling used electronics without improper deletion and management of data. The pre-owned users of the electronic devices can actually retrieve confidential information like the login passwords and misuse them, if it is not managed wisely.

IBM is doing an enormous amount of research on dark data and [4] IBM claims that 80% of the data around us is dark data and not all the collected IoT sensor data are being used for the analysis.

In the field of medical industry, the health informatics emerged back in 1960s, and the first system, MED-SYNDIKATE, on medical language processing which automatically retrieves clinically relevant information was released in 2000s [5]. Four different supervised multi-label classifiers were experimented [5], which are decision trees, multinomial naive Bayes, conditional random fields (CRFs), and support vector machines (SVM). From the observations, the performance of conditional random fields (CRFs) and support vector machines (SVM) was significantly better than other classifiers like naive Bayes and decision trees. Also, comparing CRFs and SVMs, CRFs performed better because it can identify entities in longer spans.

Identifying the relationship between diseases is an important parameter in the early detection of diseases. Patients with seizures, epilepsy, and Parkinson's are at increased risk of developing Alzheimer's disease. Friedman et al. [6] suggested the positive association between two neurological disorders, epilepsy and Alzheimer's, which are seen commonly in elderly people. Asadollahi et al. [7] discussed some additional parameters such as medications, down syndrome, dementia, and cognitive impairment as the potential risk factors that leads seizures in Alzheimer's disease.

Ott et al. [8, 9] related the association of diabetics and dementia based on the renowned large population study conducted in Rotterdam and the observations from the study shows that diabetic patients treated with insulin are having high risk of Alzheimer's disease with 4.3 fold. In the same time frame, another longitudinal cohort study was conducted with T2DM patients [10] and shows a bigger risk of all dementias and Alzheimer's (2.27 fold for men and 1.37 fold for women), supporting the Rotterdam study by proving it again that Alzheimer's and diabetics are related. The studies [11, 12] concluded that T2DM can increase the risk of Alzheimer's amid 1.3 fold and 1.8 fold. Schilling [13] analyzes and integrates multiple approaches used in different genre of studies, making it possible to unravel the research opportunities on Alzheimer's disease.

3 Proposed Model

Any data which is collected and stored but fails to use for improving the performance of other purposes is called dark data. Electronic health records are the systematic collection of patient details in electronic format which contains the personal information, medical reports, scanning images, consultation reports, etc., of a patient.

Diabetes parameters stored in EHR are considered as dark data in this research as it is not generally used in Alzheimer's prediction. As defined, this research focuses on the early prediction of Alzheimer's disease by considering some diabetics parameters in addition to the parameters of Alzheimer's disease.

Generally, for early prediction of Alzheimer's, features like Mini-Mental State Examination (MMSE), normalized whole-brain volume (nWBV), estimated total intracranial volume (eTIV), and Atlas scaling factor (ASF) are used from the EHR records.

However, the EHR data of any health record includes many other features; for example, the diabetes-related parameters may be recorded but not utilized for calculating the results, and thereby becoming dark data. In this research, along with the Alzheimer's features, the relation of Alzheimer's with other diseases is identified, and those parameters are also included in training the prediction model to increase the accuracy of the model and for better predictions.

The correlation of diabetes and Alzheimer's is discussed [14, 15], and the parameters of diabetes like glycated hemoglobin (Glyhb), also known as A1C test and stabilized glucose (stab.glu) from the EHR, are also included along with the Alzheimer parameters.

The above-mentioned parameters are extracted from electronic health records, preprocessed, and divided into training and testing data, and then, the data will undergo random forest machine learning classifier, and Alzheimer's is predicted more accurately. Later, area under the curve is used along with accuracy and recall as the performance metrics. The workflow of the proposed model is represented in Fig. 1.

3.1 Data Collection

As the research work requires identification of dark data and usage of the same for better decision making, manipulation of the dataset is required, and it is mandatory to make a dataset of EHR which contains all data. Thus, a new dataset is created in order to identify the relation between two or more diseases, and so, the merging of independent disease datasets is important.

The initial dataset used in the research work is longitudinal MRI data in non-demented and demented older adults by the Open Access Series of Imaging Studies (OASIS) [16]. This dataset includes a longitudinal collection of patients who are aged from 60 and above. All the patients are right-handed, and both men and women are taken into consideration. From the total number of patients, 72 patients were categorized as non-demented, 64 patients were diagnosed as demented, and remaining 51 individuals had mild to moderate Alzheimer's disease. The rest of the 14 patients were categorized as non-demented initially, and they do have the chances of showing signs of dementia in the later period.

The variables of the dataset are Patient.ID, MRI.IDs, Group (non-demented/demented), Number of Visits and MR.delay. The dataset also includes

Fig. 1 Workflow of proposed model

Gender, Hand, Age, Years of Education (EDUC), Socioeconomic status (SES) ranging from 1 to 5 (highest status and lowest status). Other variables that are considered important in the model are Mini-Mental State Examination (MMSE from 0 to 30) and Clinical Dementia Rating (CDR, 0 = no dementia, 0.5 = very mild dementia, 1 = mild dementia, 2 = moderate Dementia). The derived anatomic volumes or the estimated total intracranial volume (eTIV), normalized whole-brain volume (nWBV), which is the percent of all voxels labeled as the gray or white matter in the atlas-masked image taken from the automated tissue segmentation process and finally ASF—Atlas scaling factor—which is a computed scaling factor that changes the native-space brain and skull to the atlas target.

The original data came from the biostatistics program at Vanderbilt [17], and the dataset used here is a manipulated dataset with diabetes parameter. Patients without a hemoglobin A1c test is excluded. If the patient's hemoglobin A1c is 7 or greater, then they are labeled with diabetic. Glycated hemoglobin (Glyhb) or the A1C test is a blood test that specifies the average blood sugar level of a patient for the past few months. Standard levels are in the range of 4–5.6%, and a result between 5.7 and 6.4% is measured as pre-diabetic, and an A1C level of 6.5% or higher indicates that the patient is diabetic. A1C test is a commonly used blood test to diagnose type-1 and type-2 diabetes. A1C test is used to measure the percentage of hemoglobin. As the A1C level increases, the blood sugar control is very poor, and thereby the risk of diabetes complications is high. The new variables of the EHR datasets other than OASIS are cholesterol (chol.), stab.glu, high-density lipoprotein (hdl) also known

as good cholesterol, Glyhb, location, height, weight, waist, hip, bp.1s, bp.2s, bp.2d, time.ppn, results and condition.

3.2 Data Preprocessing

The database of electronic health record is a composition of non-homogeneous data sources which is obtained from the datasets of clinical records, and the data contained is redundant, diverse, and incomplete, which directly affects the performance of the model. Hence, the electronic clinical data must be preprocessed to make sure the accuracy, completeness, and consistency of the EHR dataset [6]. As our research work is a comparison, the data preprocessing steps are the same for both datasets. In the first dataset, SES column is having missing values in eight rows, and this issue can be dealt with two methodologies; the first one is to drop the rows having the missing values, and the second is to substitute the missing values from the eight rows with different matching values; the second methodology is called 'imputation.' While implementation, before imputation, all the missing values were replaced as NaN values and was triggering errors, and thus, we decided to impute the detected missing values by inferring them from the known parts of the available data, and we are using median for the imputation. As the dataset contains only less than 500 data, imputation is the right option to accelerate the performance of our proposed model. In the second dataset, as the diabetes features are also added, other than 8 rows of SES missing values, there are missing values in other features of EHR dataset. In the selected features, only SES and stab.glu had missing values. So, both the datasets will undergo same preprocessing steps.

3.3 Machine Learning Classifiers

Machine learning approaches transform clinical IE tasks into classification problems. In this, random forest machine learning classifier is used for the early detection, and this classifier can perform both regression and classification. Since the main objective is to focus the utility of dark data in EHR, we compare the performance of random forest classifier in Alzheimer disease dataset with and without dark data.

4 Results

In this, the prediction of the Alzheimer's disease with and without dark data is evaluated using random forest classifier. Both the datasets are first read and will undergo all the preprocessing steps, and then, the data will be split into training and testing sets, finally the model is trained and tested with random forest classifier. As

Table 1 Performance metrics

Performance metrics	Alzheimer dataset	EHR dataset
Accuracy of the validation set	0.7785	0.9927
Accuracy with the best parameters	0.8421	0.9893
Recall with the best parameters	0.7	0.9787
AUC score	0.85	0.9893

the performance metric, the accuracy and recall of best parameters, accuracy of the validation set and the AUC score is calculated. The observations of implementation are shown in Table 1, and the accuracy of the model increased by 17% and AUC score by 16.3%

The importance of each feature used in training and testing the model is also calculated. The best four features of Alzheimer datasets are MMSE, nWBV, ASF, and eTIV respectively. After identifying and utilizing the dark data of EHR from this application, which is diabetics parameters, the best performing features of second dataset for the early prediction Alzheimer's are, glyhb, MMSE, nWBV, and ASF. The ROC curves of both the datasets are shown in Figs. 2 and 3.

The importance of each feature used in training and testing the model is also calculated. The best four features of Alzheimer datasets are MMSE (0.385), nWBV (0.159), ASF (0.125), and eTIV (0.088), respectively. After identifying and utilizing the dark data of EHR from this application, which is diabetics parameters, the best performing features of the second dataset for the early prediction of Alzheimer's are Glyhb (0.9), MMSE (0.02), nWBV (0.01) and ASF(0.01). The ROC curve of both the datasets is shown in Figs. 1 and 2.

Fig. 2 ROC of Alzheimer's dataset

Fig. 3 ROC of EHR dataset

5 Conclusions

In this paper, the utilization of dark data helps to improve the prediction of Alzheimer's in a very early stage. Thus, the additional parameters of diabetes from the electronic health records increase the AUC score to 0.98 from 0.85 which clearly shows the better performance of the proposed model. Along with the AUC score, other performance metrics were accuracy and recall. Accuracy of the validation set got increased from 0.77 to 0.99, the accuracy of best parameters increased from 0.84 to 0.98, and the recall score increased from 0.7 to 0.9. Overall, the performance of the model increased when the diabetes feature got added to the Alzheimer's features for the prediction of early Alzheimer's.

6 Future Scope

The future scope of this work is to identify more associativity of Alzheimer's and other diseases to enhance the early prediction of Alzheimer's in a broad manner. Seizure, epilepsy, and Parkinson are a few of the factors that can be considered, which leads to Alzheimer's disease. The comparison of different machine learning classifiers and hybrid machine learning classifiers can be used on the same datasets to find a better performing classifier. Also, not only Alzheimer's but similar architecture can be modeled, trained, and tested for the early prediction of other diseases.

References

1. V. Seehusen, E. Maldonado, Using a roadmap in the back alleys of dark data. J. Technol. Res. **9** (2020)
2. W. Su, Z. Cai, Y. Li, F. Liu, S. Fang, G. Wang, Data processing and text mining technologies on electronic medical records: a review. J. Health Care Eng. (2018)
3. W. Dimitrov, C. Сярова, L. Petkova, Types of dark data and hidden cyber-security risks. Technical Report (2018). https://doi.org/10.13140/RG.2.2.31695.43681
4. https://developer.ibm.com/technologies/analytics/articles/ba-data-becomes-knowledge-3/
5. M. Tao, Clinical information extraction from unstructured free-texts. PhD thesis (Department of Information Science, University at Albany, 2018)
6. D. Friedman, L.S. Honig, N. Scarmeas, Seizures and epilepsy in Alzheimer's disease. CNS NeuroSci. Ther. **18**(4) (2012). https://doi.org/10.1111/j.1755-5949.2011.00251
7. M. Asadollahi, M. Atazadeh, M. Noroozian, Other Demen, seizure in Alzheimer's disease: an underestimated phenomenon. Am J Alzheimers Dis. **34**(2), 81–88 (2019). https://doi.org/10.1177/1533317518813551
8. A. Ott, M.M.B. Breteler, F. Van Harskamp, D.E. Grobbee, A. Hofman, Incidence of Alzheimer's disease and vascular dementia in the Rotterdam study. 970–973 (1996). https://doi.org/10.1136/bmj.310.6985.970
9. A. Ott, R. Stolk, F. Van Harskamp, H. Pols, A. Hofman, M. Breteler, Diabetes mellitus and the risk of dementia: The Rotterdam study. Neurology **10**, 1937–1942 (1999)
10. C.L. Leibson, W.A. Rocca, V.A. Hanson, R. Cha, E. Kokmen, P.C. O'Brien, P.J. Palumbo, The risk of dementia among persons with diabetes mellitus: a population-based cohort study. Ann N Y Acad Sci **826**, 422–427 (1997)
11. R. Peila, B.L. Rodriguez, L.J. Launer, Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies—The Honolulu-Asia aging study. Diabetes **51**, 1256–1262 (2002)
12. W.L. Xu, C.X. Qiu, A. Wahlin, B. Winblad, L. Fratiglioni, Diabetes mellitus and risk of dementia in the Kungsholmen project—A 6-year follow-up study. Neurology **63**, 1181–1186 (2004)
13. M.A. Schilling, Unraveling Alzheimer's: making sense of the relationship between diabetes and Alzheimer's disease. J Alzheimer's Dis. **51**(4), 961–977 (2016). https://doi.org/10.3233/JAD-150980.PMID:26967215;PMCID:PMC4927856
14. N.R. Rajeshkanna, S. Valli, P. Thuvaragah, Relation between diabetes mellitus type 2 and cognitive impairment: a predictor of Alzheimer's disease. Int. J. Med. Res. Health Sci. **3**(4), 903 (2014)
15. A.M. Ortiz Zuñiga, R. Simó, O. Rodriguez-Gómez, C. Hernández, A. Rodrigo, L. Jamilis, L. Campo, M. Alegret, M. Boada, A. Ciudin, Clinical applicability of the specific risk score of dementia in Type 2 diabetes in the identification of patients with early cognitive impairment: results of the MOPEAD study in Spain. J Clin Med. **9**(9), 2726 (2020). https://doi.org/10.3390/jcm9092726.PMID:32847012;PMCID:PMC7565958
16. https://www.kaggle.com/jboysen/mri-and-alzheimers
17. https://data.world/informatics-edu/diabetes-prediction

Brain Tumor Segmentation Using Capsule Neural Network

Jyothi Shetty, Shravya Shetty, Vijaya Shetty, and Chirag Rai

Abstract In medical science, one of the deadly diseases demonstrated by specialists is a brain tumor. Early identification of brain tumors is responsible for improving treatment possibilities, and chances of survival of the patients can be increased. Deep learning techniques can evaluate the huge amounts of MRI image data. In this paper, capsule neural network (CapsNet)-based model is used for the image segmentation to extract tumor regions. The proposed method was assessed using BraTS18 data consisting of two sets, HGG and LGG. The model is using dice coefficient, sensitivity and specificity metrics and also is compared with the U-Net model.

Keywords Brain tumor · Deep learning · Segmentation · Capsule neural network

1 Introduction

In the last decades, brain tumor segmentation has been the most trending topic in the research field. In today's world, brain tumor has become one of the reasons in rise of mortality among the people. The neoplasm called gliomas can affect the functionality of the brain depending on its location and size of growth [1]. Medical imaging technique termed as magnetic resonance imaging (MRI) is used to examine the brain tumor in the nervous system. Brain tumors cannot be identified at the beginning stage, and some tumors cannot be recognized ever after the appearance of the symptoms. Deep learning method helps in the early detection of tumor. In this work, BraTS 2018 [2] datasets consist of four different MRI modalities as shown in Fig. 1. These modalities help to give information about the irregular shaped tumors. The four modalities of an MRI image differ in their structure based on their shape, tumor border or color [3]. It is possible to find several methods in brain tumor

J. Shetty (✉) · S. Shetty · C. Rai
NMAM Institute of Technology, Nitte, India
e-mail: jyothi_shetty@nitte.edu.in

V. Shetty
NMIT, Bangalore, India
e-mail: vijayashetty.s@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_17

Fig. 1 Four different MRI modalities [3]

segmentation that are gaining importance in medical science and are complicated because of its unstable tumor shape [4].

The brain tumor is detected by reading the MRI image of brain and applying segmentation method to identify the tumor region [5]. Various machine learning-based segmentation methods were implemented based on their similarity or dissimilarity measurements. In this paper, a segmentation model is proposed using a new architecture called capsule neural network (CapsNet) with dynamic routing has been introduced. Capsule neural network model stores the information about the input by replacing max pooling layers with convolutional strides and dynamic routing [6].

This model is evaluated using dice coefficient, sensitivity and specificity metrics for HGG and LGG cases for these different models. This model is also compared with U-Net model, and results show that CapsNet model performs better than the U-Net model [7]. The U-Net and CapsNet designs are assessed using leaderboard BraTS 2018 datasets consisting of LGG and HGG patients. The predicted result is differentiated with manual described as ground truth; CapsNet model acquired good results.

2 Literature Review

Segmentation is the method where objects having same kind of properties are put in a group together and dissimilarity ones are separated. Segmentation of brain tumor has a great consequence in medical image segmentation. The researchers have proposed several papers on brain tumor segmentation. Several machine learning algorithms are used to analyze and understand the concept of brain tumor segmentation. K-means clustering gave best result when comparing to other algorithms [8].

Subbanna and Arbel [9] used Markov random fields (MRFs) for segmenting the brain tumor. Wu et al. [10] introduced superpixel features. To segment brain tumor images, superpixel features were employed using conditional random fields framework, but the results were not up to the mark as it did not perform well with LGG cases. Randomized forest was introduced for classifying features that achieved dice score of 83% [11]. Yang et al. [12] merged classification technique randomized trees with another algorithm named superpixel-based segmentation techniques. Single FLAIR sequence-based MRI is used for the performance of the model. Pereira et al. [13] implemented CNN algorithm for segmentation. The fully connected layers were used as classifiers in their model.

3 Dataset

The experiments were conducted on multimodal brain tumor segmentation challenge (BraTS) 2018 datasets to assess the performance of the model. It consists of two sets of data containing 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG). It includes four MRI modalities such as T1-weighted (T1), T2-weighted (T2), FLAIR and T1-weighted imaging with gadolinium enhancing contrast (T1c) images with corresponding ground truth images. These four modalities of MRI are included in both HGG and LGG data. The three modalities of MRI are T1, T2 and FLAIR images of each patient cataloged into the MRI modality of T1c image. These four modalities differ from each other based on their features and also contain instructions to predict the area and tumor type. Voxels of each image are classified into five categories: enhancing core, necrosis, non-enhancing core, edema and background noise. BraTS18 data consists of 3D image, each with size of $240 \times 240 \times 155$.

4 Preprocessing

Preprocessing of data is used to transform unprocessed data to clean data for better performance. It helps to train the network very smoothly and also to enhance the overall execution of the model to obtain good results. The code has been executed using Colab notebooks which support free GPU. The preprocessing techniques used

are bias correction, 2D slicing, cropping and data augmentation. As smooth and less frequency noise signal is attached to the image, N4ITK method is used to remove this effect. 2D slicing BraTS 2018 datasets consist of 3D images. These 3D images are sliced as 2D NumPy array. Each image generated 155 NumPy arrays. All 155 slices do not show tumor region, and hence, only the mid-portion of slices needs to be considered. We took from 30th slice to 120th slice for creating the final data. All are resized and cropped to center with final dimension of (N1,192,192,4). Data augmentation technique is used to solve data imbalancing problem that results in unequal distribution of classes. Augmentation helps in improving balancing classes in the training datasets. Augmentation approach like rotation, flipping and enhancement of image is being used by the experts as a preprocessing method. In this work, flipping method is used. Finally, the data was randomly split into training, validation and test data with 60%:20%:20% of ratio, respectively.

5 Evaluation Metrics

Dice similarity co-efficient:

It measures the comparability between the ground truth label and the predicted label.

$$\mathrm{DSC} = 2 * \frac{\mathrm{TP}}{\mathrm{FP} + 2 * \mathrm{TP} + \mathrm{FN}}$$

Sensitivity:

Sensitivity calculates the amount of true positive rate or recall rate (risk class).

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Specificity:

Specificity computes the amount of true negatives (normal class).

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

6 Results

In capsule neural network architecture, the model is assessed using training and validation sets for HGG and LGG cases. Firstly using HGG data, the model is trained

on 3780 samples, validated on 1260 samples. The results obtained for CapsNet model are shown below.

Figure 2a, b shows the result of dice coefficient score and dice coefficient loss for training and validation of HGG set, respectively.

In Fig. 3a, b, results of sensitivity and specificity for validation and training HGG set are plotted.

Next, using LGG set of data that consists of 75 patients scan images. We have trained CapsNet model on 4050 samples, validated on 1350 samples.

The result of dice coefficient score and loss for training and validation set is plotted, respectively, in Fig. 4a, b.

In Fig. 5a, b, sensitivity and specificity results are plotted. Figures 6 and 7 show the prediction of HGG and LGG data from BraTS18 datasets using CapsNet model.

Table 1 shows the dice similarity coefficient, sensitivity and specificity for HGG and LGG test sets using CapsNet model. We have compared these results with the dice similarity coefficient (DSC), sensitivity and specificity results for HGG and LGG test sets using U-Net model. Results of U-Net model are as shown in Table 2. This study shows that the CapsNet model has given best result on testing data.

Fig. 2 **a** Model score for HGG set. **b** Model loss for HGG set

Fig. 3 **a**. Sensitivity for HGG set. **b** Specificity for HGG set

Fig. 4 **a**. Model score for LGG set. **b** Model loss for LGG set

Fig. 5 **a**. Sensitivity for LGG set. **b** Specificity for LGG set

7 Conclusion

In biomedical engineering, brain glioma segmentation is the task of segmenting tumors from normal brain tissues. Fully convolutional network (U-Net) algorithm and capsule neural network (CapsNet) algorithm are constructed to extract the tumor regions. Validation is done using BraTS2018 leaderboard dataset holding two sets: 70 high-grade gliomas (HGG) and 75 low-grade gliomas (LGG) patient scans each set split into training, validation and testing data. U-Net and CapsNet networks are trained and assessed using dice coefficient, sensitivity and specificity of HGG and LGG data. Comparing the U-Net results with the CapsNet results, capsule neural network model gave a promising result.

Fig. 6 Prediction of a HGG subject using CapsNet model

Fig. 7 Prediction of a LGG subject using CapsNet model

Table 1 Dice similarity coefficient (DSC), sensitivity and specificity results for HGG and LGG test sets using CapsNet model

Dataset	Model name	Metrics	HGG data	LGG data
BraTs 2018 Leaderboard	CapsNet model	DSC	0.9841	0.9953
		Sensitivity	0.9826	0.9947
		Specificity	0.9944	0.9953

Table 2 Dice similarity coefficient (DSC), sensitivity and specificity results for HGG and LGG test sets using U-Net model

Dataset	Model name	Metrics	HGG data	LGG data
BraTS 2018 Leaderboard	U-Net model	DSC	0.9796	0.9792
		Sensitivity	0.9780	0.9797
		Specificity	0.9928	0.9927

References

1. M. Lyksborg, O. Puonti, M. Agn, R. Larsen, An ensemble of 2d convolutional neural networks for tumor segmentation, in *Scandinavian Conference on Image Analysis* (Springer, 2015), pp. 201–211
2. B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby et al., The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)
3. B. Menze et al., The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)
4. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networksfor semantic segmentation in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), pp. 3431–3440
5. F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: fully convolutional neural networks for volumetric medical ımage segmentation (2016), pp. 1–11
6. S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)* (2017) pp. 3859–3869
7. O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical ımage segmentation. arXiv, cs.CV (2015)
8. K.M. Nimeesha, R.M. Gowda, Brain tumour segmentation using K-means and fuzzy c-means clustering algorithm. Int. J. Comput. Sci. Inf. Technol. Res. Excell. **3**, 60–65 (2013)
9. N.K. Subbanna, T. Arbel, Probabilistic Gabor and Markov random fields, segmentation of brain tumours in MRI volumes, in *Proceedings of the MICCAI-BRATS* (2012)
10. W. Wu, A.Y.C. Chen, L. Zhao, J.J. Corso, Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features. Int. J. Comput. Assist. Radiol. Surg. **9**(2), 241–253 (2013)
11. L. Szilágyi, L. Lefkovits, B. Benyó, Automatic brain tumor segmentation in multispectral MRI volumes using a fuzzy c-means cascade algorithm, in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (2015), pp. 285–291
12. M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T.L. Jones, T.R. Barrick, F.A. Howe, X. Ye, Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. Int. J. Comput. Assist.
13. S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging **35**, 1240–1251 (2016)

Forgery Detection and Image Recommendation Systems

Using Machine Learning for Image Recommendation in News Articles

Rohit Jere, Anant Pandey, Hasib Shaikh, Sulochana Nadgeri, and Pragati Chandankhede

Abstract The recent developments in the machine learning domain have changed the journalism workflow significantly. A growing interest is being observed in adapting to technologies that can scan, process, filter, and even suggest photos for news articles. This research paper focuses on image selection by harnessing keyword extraction. The terms present in the article are first converted to a machine-readable format and then ranked based on three characteristics, namely first occurrence, term frequency, and entity category. Four variations of machine learning approach and simple statistical calculations are applied for the image selection process followed by that. Then, the highest-ranking terms are used for search query generation that involves a heuristic method predicting threshold probability values for the terms.

Keywords Machine learning · Open Calais entity detection · Statistical calculation

1 Introduction

Artificial Intelligence is streamlining the workflow for media outlets globally for curating as well as writing news articles. Microsoft has been using AI to scan for content and then process and filter it and even suggest photos for human editors to pair it with. Microsoft had been using human editors to curate top stories from a variety of sources to display on Microsoft News, MSN, and Microsoft Edge [1].

However, selecting suitable images for these articles is not an area that is gaining traction in this domain. Images considerably impact the reader's mind and studies show that they help in recalling the article better. Selecting a large number of images for the articles can be tedious and automating this process can be immensely beneficial for the newsroom. This research leverages keyword extraction mechanisms, computationally efficient statistical calculations, and machine learning to produce search queries that return suitable images for news articles. Keyword extraction is an automated process that collects a set of terms, illustrating an overview of the

R. Jere (✉) · A. Pandey · H. Shaikh · S. Nadgeri · P. Chandankhede
Excelsior Education Society's, K. C. College of Engineering and Management Studies and Research, Kopri, Thane (East), Maharashtra, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_18

document. The term is defined as how the keyword identifies the core information of a particular document [2]. The images for this research are collected from Getty Images and the news articles from BBC News Website.

The plain article text is first converted into a machine-readable format through tokenization, n-gram selection, stopword removal, stemming, and feature assignment. These terms are then ranked based on first occurrence, term frequency, and entity categorization. The image's metadata is preprocessed to find better keywords for searching the right images. Metadata is "data that provides information about other data" [3]. Also, in this research paper, "Keywords" are described as terms that best illustrate a news article.

The training data is generated with a matcher component that assigns Boolean values to the terms upon comparing article terms with image tags. These terms are transformed into a vector that forms the input layer of the neural network. A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes [4]. A training data generator component is used to convert all the vector values that help in identifying fully activated and inactivated neurons.

Pre-trained feedforward neural networks were used for term classification in this research and they used different types of term feature combinations to distinguish keywords from non-keywords. Both machine learning and computationally efficient statistical approach were used to predict the probability values of the terms acquired through ranking mechanisms to assess their suitability as a keyword.

Furthermore, a heuristic method is applied to enhance the image query generation. A heuristic function, also called simply a heuristic, is a function that ranks alternatives in search algorithms at each branching step based on available information to decide which branch to follow. For example, it may approximate the exact solution [5]. In the end, the results are evaluated to identify the best performing approach.

2 Methodology

2.1 Data Collection

Data is sourced from the records of the popular visual media company Getty Images and the BBC News website. A web harvesting application is designed based on this data that stores articles gathered from BBC News website locally. Also, only the articles that consisted of images from Getty were used as training data. These are those images that convey the whole idea of the article and the application checks if its Getty ID is exposed. The application then requests image metadata from the Getty API also. Thomson Reuters Open Calais is also used to identify people, events, organizations, and other entities in the corpora. In the end, a dataset of 500,000 terms was accumulated through this process. Figure 1 describes the methodology followed through a flow chart.

Fig. 1 Flow diagram for image selection

2.2 Structuring the Data

2.2.1 Article Conversion Process

The article is converted into a machine-readable format through the following process.

Tokenization The strings of plain text in the article are first transformed into a sequence of words with the help of a tokenizer.

N-gram Selection N-grams are sequences of up to four consecutive words that are taken into consideration by the pre-processor. N-grams with punctuations are not considered though.

Stopword Removal Stopwords are the words that do not confer meaning individually such as pronouns. All such words are removed during preprocessing.

Stemming Stemming is applied for the purpose of inflection associated with a particular word or a group of words. Porter's stemming algorithm was used for performing stemming.

Feature Assignment Finally, every term is classified with a set of features. It distinguishes the terms that can be instrumental in finding an image that conveys the main idea of the article.

2.2.2 Term Labels

After the article is converted to the desired format, the extracted terms are ranked based on the following.

Term Frequency The number of times a particular term repeats in the article is recorded to measure its importance. Repeating words are combined to one list entry by the preprocessor. A term frequency of value 1 is assigned to each term and every time the term repeats in the text, the count is incremented.

First Occurrence The relative position at which a term is introduced for the first time in the article is also tracked. The term occurring earlier is identified as more important in comparison. This is inspired from the concept of inverted pyramid which involves placing the most important information at the top in order to captivate reader's attention. It is implemented by searching for the term first and then counting

Table 1 Detections

Entity name	Terms	Entity type	Entity category
Reddit	Reddit	Organization	OrganizationProtagonist
	Tech-Firm		
	Company		
Alexis Ohanian	Alexis Ohanian	Person	HumanProtagonist
	Co-founder him		

the number of characters that occur before it. The value obtained is divided by the total number of characters in the text that transforms it into a relative measure. Consequently, 0 savors the first character in the article and 1 the last.

Entity Category Automated semantic tagging was used to improve the term classification to consolidate terms that refer to the same entity. Entities can include individuals, organizations, products, and people, among others. This was achieved by using the web service—Thomson Reuters Open Calais, which assigns intelligent metadata tags to the entities. These entities are attached to the entity category, which helps in describing a term better.

This further helps in creating a better search query as it segregates names, locations, events, and other entities, and helps finding better search terms for finding suitable images. Seven entity categories are developed for this purpose that includes: Event, HumanProtagonist, OrganizationProtagonist, Position, Location, Product, and Other.

The input text and Table 1 are a representation of Open Calais entity detection. Input Text: *Alexis Ohanian—co-founder of Reddit, has resigned from the tech-firm's board and urged the company to replace him with a black candidate.*

2.3 Training the Model

Figure 2 depicts the training procedure using the flow diagram. An ideal training data for the system pertinent to this research should be the search queries that human editors use to find a particular image for their article. However, due to limitations of time and resources, it is assumed that combinations of the terms used in the image description can be utilized to fetch images similar to that image. Although this is a weak approach, it can help generate search queries that might even be identical to what human editors search. Figure 3 is image used for image preprocessing and Table 2 is a representation of the image metadata.

The image metadata is preprocessed to make the terms in the article consistent with the image description. This process is fairly identical to the Open Calais entity detection explained in Sect. 2.2.2 under term labels. However, a new list is introduced here named "image tags" comparable to the article terms. Image metadata

Fig. 2 Flow diagram depicting the training procedure

preprocessing is composed of three fields: title, caption, and image keywords. The data is converted to machine-readable format using the process described in Table 2. However, only stemming is applied to the image keywords since most of them represent fixed terms describing topics, and features such as color and orientation. Both these lists are merged to eliminate repeating terms in the output.

Fig. 3 Image used for image preprocessing

Table 2 Image metadata preprocessing

Representation	Image metadata
Title	Einstein and Wife
Caption	circa 1921: German–Swiss–American mathematical physicist Professor Albert Einstein (1879–1955) with his wife Elsa in Egypt. (Photo by Topical Press Agency/Getty Images)
Keywords	1990–1929, Adult, Africa, Black And White, Couple- Relationship, Data, Egypt, German Culture, Germany, Mathematical Symbol, Mathematics, People, Physicist, Science and Technology, Swiss Culture, Wife, Women, Albert Einstein, Elsa Einstein

2.4 Generating Training Data

The resultant list of image tags obtained from image metadata preprocessing is further processed by the matcher component. The matcher component assigns the Boolean feature is keyword to every term upon comparing the image tags with the article terms. The Boolean value is true if the article term matches with the image tag and false otherwise.

Each of these terms is transformed into a vector that forms the input layer of the neural network. A training data generator component is used to convert all the values to numbers ranging from 0 to 1. 0 indicates inactivated neuron and 1 represents fully activated neuron. Figure 4 represents an example vector for one term.

Term Frequency The sigmoid function is used to normalize the values in the input layer of the neural network. Sigmoid was chosen since its value increments swiftly for natural numbers greater than 1.

$$\mathrm{sig}(x) = \frac{1 - \mathrm{e}^{-x}}{1 + \mathrm{e}^{-x}} \tag{1}$$

Fig. 4 One term as vector in the training data

Since the lowest possible value for term frequency is 1, the value is decreased by 1 before applying the sigmoid function. This implies that a term frequency of 1 will result in an inactivated neuron ($\text{sig}(1 - 1) = \text{sig}(0) = 0$).

First Occurrence Values for first occurrence range from 0 to 1 by design and do not require normalization. They are applied to the term vector without alterations.

Entity Category This feature is represented as an eight-dimensional vector, wherein, the first value, is Entity, indicates whether the terms is an entity or not and the other seven values represent the entity categories. If the term is not identified as an entity all the values are assigned 0, else, they are assigned 1. An entity is better suited as a search term as compared to a term that is not an entity. If a term qualifies as a keyword, it acquires the value 1, else it is 0.

2.5 Training Neural Networks

Pre-trained feedforward neural networks were used for term classification in this research because they provide a simple and effective mechanism for performing classifications. Four networks were trained that used different types of term feature combinations to distinguish keywords from non-keywords which include:

- term frequency + first occurrence
- term frequency + entity category
- first occurrence + entity category
- term frequency + first occurrence + entity category.

2.6 Image Selection Process

The terms identified by the preprocessor that qualify as possible keywords are further classified regarding their suitability as search queries. Probability values were predicted for these keywords. 1 indicates that the system the considered term is a 100% search term and 0 indicates otherwise.

Five different methods are applied for predicting the probability values which include the four variations of machine learning approach described in Sect. 2.5 and another method based on statistical approach that is explained below.

Statistical Approach The calculation applied here is based on term frequency and first occurrence. Therefore, the term that earliest occurring keyword that is repeated the most in the articled can be classified as the ideal keyword. This can be described as follows:

$$p(\mathrm{tf}, \mathrm{fo}) = \frac{\mathrm{tf}}{\max(\mathrm{TF})} * (1 - \mathrm{fo}) \tag{2}$$

Here, tf denotes the frequency of a term, fo its first occurrence value, max(TF) the highest term frequency value of all terms in an article, and p the probability of this specific term being a keyword. $\frac{\mathrm{tf}}{\max(\mathrm{TF})}$ equals 1 when the term under consideration is the most frequent one in the article, in all other cases, it is less than 1. The difference $(1 - \mathrm{fo})$ converges toward 1 if fo converges toward 0. The ideal keyword in this model would therefore have a probability value of $\frac{\max(\mathrm{TF})}{\max(\mathrm{TF})} * (1 - 0) = 1 * 1 = 1$.

2.7 *Query Generator Component*

After keyword classification, the list generated is passed on to the query generator component which converts the ranked terms into a query string. To eliminate the problem of polysemy, two highest-ranking terms are selected by default and linked together to form a query. It is observed that queries consisting of more than three words become too specific, so the shorter term is disregarded if the resultant query is longer than three words. It was also conspicuous that abbreviations and short forms were higher ranked and resulted to be more suitable for a query. However, the short forms are followed by spelled-out terms in the ranking list soon. Thus, a simple heuristic is applied to replace short forms with the spelled-out terms, if they are ranked reasonably high. This method is illustrated below in Fig. 5.

Fig. 5 Term replacement at the time of search query generation

This heuristic method defines a threshold probability value first and the terms that fall above this value are considered suitable search queries. The threshold was set to 0.5 for the statistical approach and 0.1 for the machine learning approaches, respectively, after conducting manual benchmarks. Two highest-ranked terms were selected followed by that. The query generated also looks for super terms after selecting the search terms which indicates the terms that have the same meaning as the selected terms. Then, the search query is generated by connecting the resulting terms and transforming them into lowercase letters. Finally, this query is sent to Getty API, which returns a broad range of images sorted by their popularity. The most popular image is selected for the article.

3 Results

Figure 6 depicts the results of image selection process. It is observed from Fig. 6 that the machine learning approach renders appreciable results when term frequency and first occurrence are both used as features. Term frequency + First Occurrence returned 40% correct images and Term frequency + First Occurrence + Calais returned 42% correct images. Therefore, it is recognized as the best performing approach in this research. A comparison between statistical and machine learning approaches is shown in Fig. 7. From Fig. 7, it is observed that the statistical approach returns the highest percentage of incorrect images (59%).

Figure 8 shows the power of features in sculpting the performance of neural networks. Also, adding the First Occurrence feature to the networks improved the results from 30% correct images in the Term frequency + Calais network to an average of 39.66% in all other networks that use first occurrence. This shows how impactful the first occurrence of a word is in an article. Apart from that, term

Fig. 6 Results of image selection process

Fig. 7 Statistical approach versus machine learning approach

Fig. 8 Influence of features on the performance of neural network

frequency has a powerful effect when combined with first occurrence and adding entity category to this network further increases the successes rate.

4 Conclusion

Images play a vital role in influencing the reader and dictates how they may interpret the article. Therefore, this research proposes a model for extracting keywords from news articles and use them to generate search queries to acquire suitable images for the article. First, terms are extracted from the articles and then ranked based on first occurrence, term frequency, and entity category. One statistical approach and four variations of machine learning approach were applied to distinguish the keywords from the non-keywords in these terms. Image query was then generated based on

the highest-ranked terms. Finally, it was found that the machine learning approach performed best with Term frequency + First Occurrence + Entity Category network, returning an accuracy of 42%. This system can be useful for suggesting images to the human editors, which can improve the overall productivity of the newsroom. The future work of this research involves comparing the performance of these approaches on different article topics and varying the type and design of the networks.

References

1. T. Warren, Microsoft lays off journalists to replace them with AI (2020). Retrieved 19 Aug 2020 from https://www.theverge.com/2020/5/30/21275524/microsoft-news-msn-layoffs-artificial-intelligence-ai-replacements
2. H.M. Hasan, F. Sanyal, D. Chaki, Md. Ali, An empirical study of important keyword extraction techniques from documents (2017). https://doi.org/10.1109/ICISIM.2017.8122154
3. Merriam Webster. Archived from the original on 27 Feb 2015. Retrieved 17 Oct 2019
4. J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. U.S.A. **79**(8), 2554–2558 (1982). Bibcode: 1982 PNAS 79.2554H. PMC: 346238, PMID: 6953413. https://doi.org/10.1073/pnas.79.8.2554
5. J. Pearl, *Heuristics: Intelligent Search Strategies for Computer Problem Solving* (Addison-Wesley Pub. Co., Inc., Reading, MA, US, 1984), p. 3. OSTI 5127296

An Approach to Noisy Synthetic Color Image Segmentation Using Unsupervised Competitive Self-Organizing Map

P. Ganesan, B. S. Sathish, L. M. I. Leo Joseph, B. Girirajan, P. Anuradha, and R. Murugesan

Abstract The complex data is transformed as the simple but meaningful smaller data groups in the segmentation process. It is the utmost phase of the data exploration process. It is the method of allocating a tag to each pixel to make them as groups (clusters) and the pixels using the same tag have the common characteristics such as color, texture, or intensity. It is challenging to decide on the optimal segmentation method. For noisy images, segmentation becomes more difficult one. This is due to both the image and noisy pixels are considered as the same category. In this work, an artificial neural network based unsupervised self-organizing maps utilized to analyze and cluster the noisy synthetic images. The projected technique employed three levels (competition, cooperation, and adaptation) of competitive learning to segment the data into meaningful regions. The investigational end result undoubtedly revealed the proficiency of the suggested methodology to cluster the noisy images.

Keywords Synthetic image · Self-organizing map · Competitive learning · Clustering · Noise

P. Ganesan (✉)
Department of Electronics and Communication Engineering, Vidya Jyothi Institute of Technology, Aziz Nagar, C.B. Post, Hyderabad, India
e-mail: gganeshnathan@gmail.com

B. S. Sathish
Department of Electronics and Communication Engineering, Ramachandra College of Engineering, Eluru, Andhra Pradesh, India

L. M. I. Leo Joseph · P. Anuradha
Department of Electronics and Communication Engineering, SR University, Warangal, Telangana, India
e-mail: leojoseph.lmi@sru.edu.in

B. Girirajan
Centre for Social Innovation, SR University, Warangal, Telangana, India
e-mail: dir.csi@sru.edu.in

R. Murugesan
Department of Electronics and Communication Engineering, Narsimha Reddy Engineering College, Secunderabad, Telangana, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_19

1 Introduction

Segmentation, the clustering technique, is the utmost phase of the data exploration process. The complex image is simply transformed into number of small but meaningful images [1]. It is the method of allocating a tag to each pixel to make them as groups (clusters), and the pixels using the same tag have the common characteristics such as color, texture, or intensity [2]. Segmentation finds many applications in image processing. In clustering, the whole data is huddled as an amount of small data groups (clusters) in a significant comportment [3]. All the image pixels in the same cluster should share the common and unique characteristics among them. These characteristics may be color, texture, or intensity [4]. The accomplishment of the higher level image processing techniques solely depends on the outcome of the segmentation process. It is challenging to decide on the optimal segmentation method [5]. Especially for the noisy images, segmentation becomes more difficult one. This is due to both the image, and noisy pixels are considered as the same category. This is the major reason that the segmentation process is failed to achieve its goal, i.e., leads to either under or over segmentation. For image segmentation, number of methods are developed, but most of them are application oriented. The noise in image processing is considered as the superfluous, unnecessary information which corrupts the original information. This useless information is the result of the instability of image characteristics such as color or intensity [1]. The proposed method analyzed the impact of three additive noises (salt and pepper, Gaussian, and Poisson) and one multiplicative noise (Speckle). The salt and pepper (impulse) noise is mainly due to the arbitrarily dispersed black (pepper) and white (salt) pixels over the image. In this work, the impact of salt and pepper on synthetic images is analyzed before segmentation. The noise simply destroys the original content of the image. The optimal selection of the filter, to remove the unnecessary components (pixels) vehemently added to the image, mainly depends on its behavior, the nature of the assignment carried out, and the kind of the data. The proposed work utilized the median filter, powerful nonlinear filter, to remove the unnecessary impulse noisy pixels. The principle behind this filter is that all the pixel values are arranged into ascending order, and then, the median value is computed.

2 Proposed Methodology for Noisy Synthetic Image Clustering

The unsupervised self-organizing maps (SOM) changes its neuron weights using the competitive learning [6, 7]. The structural design of SOM is represented in Fig. 1. SOM consists of one input layer and feature map [8, 9].

The segmentation of color synthetic image clustering based on SOM instigate with the weight vectors initialization. Then, SOM arbitrarily selected a test vector [9]. All the weight vectors are compared with test vector to opt the best (similar) one [10].

Fig. 1 Structural design of SOM

We know that every weight vector is surrounded by similar neighboring weights. The opted weight and its neighbors are recognized as finest alike units [11]. From this step, the amount of neighbors and the learning rate decreases over time. The entire procedure is reiterated a huge amount of times with respect to the complexity of the data. This unsupervised ANN using three levels (competition, cooperation, and adaptation) of competitive learning to revise its weights [12]. In competition level, the SOM estimate the space between every neuron (Output layer) and the input data using Euclidean distance [13]. The neuron with the minimum distance is considered as the winner of this process. In the second level (cooperation), the opted weight and its neighbors are updated. The neighbor's selection is exclusively based on the kernel function and its factors such as time and the space between the winner and other neurons [14]. In the adaptation process, the winner neuron and its neighborhood simply updated using (1).

$$W_k = W_k + \eta(t)h_{ik}(t)\left(X^{(n)} - W_k\right) \tag{1}$$

The learning rate (η) changes the weights after every iteration. The learning rate gradually decreases and converges to zero after time 't'.

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{T_1}\right) \tag{2}$$

The neighborhood kernel function, ($h_{ik}(t)$) and decay rule ($\sigma(t)$) is computed using (3) and (4), respectively

Fig. 2 Suggested methodology for synthetic color image segmentation

$$h_{ik}(t) = \exp\left(-\frac{d_{ik}^2}{2\sigma^2(t)}\right) \tag{3}$$

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{T_1}\right) \tag{4}$$

d_{ik} = the distance between W_i and W_k.

The suggested methodology for synthetic color image segmentation illustrated in Fig. 2.

3 Experimental Outcomes and Discussions

The test image is acquired from the database, a collection of fifty synthetic images. The test image and its corresponding noisy and filtered version illustrates in Fig. 3.

Fig. 3 Test image 1 and its corresponding noisy and filtered version

The end result of this process is depicted in Fig. 4. The input image (Fig. 4a) is in RGB space and noise free. It is resized to smaller size to reduce the execution time. The test image is clustered into 3, 4, and 5 clusters according to its color attribute as illustrated in Fig. 4b, c, and d, respectively.

Fig. 4 Test image 1 and its clustered versions

The SOM neighbor distance and its hits (clusters) are demonstrated in Fig. 5. The total number of pixels in the test image 1 is 10,000. They are fragmented into three clusters as illustrated in Fig. 5. The cluster one comprised of 3509 pixels. The cluster two and three included 2642 and 3859 pixels, respectively.

The weight positions of the proposed methodology for this cluster process are clearly represented in Fig. 6.

The image 1 is divided into four clusters by 0.2631 s. This process completed within 12 iteration and PSNR is 14.1282 dB. The complete image clustered into four segments as shown in Fig. 7. Out of 10,000 pixels, the cluster one has 3035 pixels. The cluster two, three, and four comprised of 2169, 2391 and 2405 pixels, respectively.

The weight positions of the proposed methodology for four clusters are clearly represented in Fig. 8.

Fig. 5 Distribution of pixels for three clusters

Fig. 6 Weight positions for three clusters (image 1)

Fig. 7 Distribution of pixels for four clusters

Fig. 8 Weight positions for four clusters (image 1)

Figure 9 demonstrated that the image 1 is divided into five clusters by 0.2769 s. This process completed within 12 iteration and produced PSNR is 14.0587 dB. The complete image clustered into five segments as shown in Fig. 9. Out of 10,000 pixels, 1892 belonged to the cluster one. The cluster two, three, four, and five comprised of 1377, 1823, 2163, and 2745 pixels, respectively (Table 1).

The weight positions of the proposed methodology for four clusters are clearly represented in Fig. 10.

The test image 2 and its corresponding noisy and filtered version are illustrated in Fig. 11.

The test image 2 (Fig. 12a) is resized to smaller size to reduce the execution time. The test image 2 clustered into 3, 4, and 5 clusters according to its color attribute as illustrated in Fig. 12b, c, and d, respectively.

Fig. 9 Distribution of pixels for four clusters

Table 1 Synthetic image clustering for different clusters (image 1)

No. of iterations	No. of clusters	PSNR (dB)	Execution time (s)	Centroid		
12	3	14.3025	0.2186	192.1001 152.5753 106.7617	150.5383 96.5127 48.8623	92.3365 61.1608 36.7069
12	4	14.1282	0.2633	90.9490 136.8896 188.3347 194.836	40.1693 69.0841 124.6539 166.1575	34.1045 42.9025 70.5705 106.5883
12	5	14.0587	0.2769	183.7212 200.3471 189.4729 128.0811 84.9170	182.3987 149.9198 99.1911 62.3581 39.9321	132.1818 82.3908 44.0536 41.8389 35.4926

Fig. 10 Weight positions for four clusters (image 1)

Fig. 11 Test image 2 and its corresponding noisy and filtered version

Fig. 12 Test image 1 and its clustered versions

The SOM neighbor distance and its hits (clusters) are demonstrated in Fig. 13. The total number of pixels (10,000) in the test image is segmented into three clusters as illustrated in Fig. 13.

The cluster one is comprised of 3715 pixels. The cluster two and three included 2373 and 3912 pixels, respectively. The weights position for this process is illustrated in Fig. 14.

Fig. 13 Distribution of pixels for three clusters

Fig. 14 Weight positions for three clusters (image 2)

The image 2 is segmented into four clusters by 0.2786 s, and PSNR is 14.7792 dB. The complete image clustered into four segments as shown in Fig. 15. Out of 10,000 pixels, the cluster one has 2686 pixels. The cluster two, three, and four comprised of 2320, 1593, and 3401 pixels, respectively.

Figure 16 demonstrated that the image 2 is segmented into five clusters by 0.2681 s. and PSNR is 14.7927 dB. The complete image is clustered into five segments as shown in Fig. 5. Out of 10,000 pixels, 2316 belonged to the cluster one. The cluster

Fig. 15 Distribution of pixels for four clusters

Fig. 16 Distribution of pixels for four clusters

two, three, four, and five comprised of 1271, 2052, 2060, and 2301 pixels, respectively (Table 2).

The illustrative representation of the impact of the proposed method for synthetic color image 1 and 2 displayed in Figs. 17 and 18, respectively.

Table 2 Synthetic image clustering for different clusters (image 2)

No. of iteration	No. of clusters	PSNR (dB)	Execution time (s)	Centroid		
12	3	14.9976	0.2611	186.8795 122.0374 62.4593	110.2171 100.3241 82.9755	94.5572 96.9722 93.1347
12	4	14.77972	0.2786	63.5368 71.1992 141.2413 196.4517	60.1775 101.8814 127.3624 110.0399	66.9544 111.801 122.236 93.8887
12	5	14.7927	0.2681	66.0415 60.8333 97.9488 165.0670 202.6202	133.4127 88.3334 45.3848 74.9134 120.4674	153.2588 100.0013 45.0791 62.9312 95.5374

Fig. 17 Pictorial representation (PSNR) of the suggested approach

Fig. 18 Pictorial representation (execution time) of the suggested approach

4 Conclusion

The proposed methodology presented an artificial neural network-based unsupervised self-organizing maps to cluster the noisy synthetic images. The test image is acquired from the database, a collection of fifty synthetic images. The projected technique employed three levels (competition, cooperation, and adaptation) of competitive learning to segment the data into meaningful regions. The investigational end result tabulated and undoubtedly revealed the proficiency of the suggested methodology.

References

1. P. Ganesan, V. Rajini, Segmentation and denoising of noisy satellite images based on modified Fuzzy C means clustering and discrete wavelet transform for information retrieval. Int. J. Eng. Technol. **5**(5), 3858–3869 (2013)
2. B.S. Sathish, P. Ganesan, S. Khamar Basha, Color image segmentation based on genetic algorithm and histogram threshold. Int. J. Appl. Eng. Res. **10**(6), 5205–5209 (2015)
3. V. Kalist, P. Ganesan, B.S. Sathish, J.M.M. Jenitha, Possiblistic-fuzzy C-means clustering approach for the segmentation of satellite images in HSL color space. Procedia Comput. Sci. **57**, 49–56 (2015)
4. P. Ganesan, B.S. Sathish, G. Sajiv, A comparative approach of identification and segmentation of forest fire region in high resolution satellite images. in *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)* (2016), pp. 1–6
5. P. Ganesan, K. Palanivel, B.S. Sathish, V. Kalist, K.B. Shaik, Performance of fuzzy based clustering algorithms for the segmentation of satellite images-a comparative study. in *IEEE Seventh National Conference on Computing, Communication and Information Systems (NCCCIS)* (2015), pp. 23–27
6. T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, Self-organization of a massive document collection. IEEE Trans. Neural Netw. **11**, 574–585 (2000)
7. J. Sun, Y. Yang, Y. Wang, L. Wang, X. Song, X. Zhao, Survival risk prediction of esophageal cancer based on self-organizing maps clustering and support vector machine ensembles. IEEE Access **8**, 131449–131460 (2020)
8. S. Wang, X. Zhang, Analysis of self-organizing maps (SOM) methods for cell clustering with high-dimensional OAM collected data. in *IEEE 5th International Conference on Cloud Computing and Big Data Analytics* (Chengdu, China, 2020), pp. 229–233
9. P. Yang, D. Wang, Z. Wei, X. Du, T. Li, An outlier detection approach based on improved self-organizing feature map clustering algorithm. IEEE Access **7**, 115914–115925 (2019)
10. V. Chaudhary, R.S. Bhatia, Anil K. Ahlawat, The self-organizing map learning algorithm with inactive and relative winning frequency of active neurons. HKIE Trans. **21**(1), 62–67 (2014)
11. M. Hagenbuchner, A. Sperduti, A.C. Tsoi, A self-organizing map for adaptive processing of structured data. IEEE Trans. Neural Netw. **14**, 491–505 (2003)
12. J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map. IEEE Trans. Neural Netw. **11**, 586–600 (2000)
13. Kohonen, T.: Self-Organizing Maps. Springer Series in Information Sciences, vol. 30, 3rd edn. (Berlin, Germany, 2001)
14. X. Xiao, E.R. Dow, R. Eberhart, Z.B. Miled, R.J. Oppelt, A hybrid self-organizing maps and particle swarm optimization approach. Concurrency Comput. Pract. Experience **16**, 895–915 (2004)

Building Dataset and Deep Learning-Based Inception Model for the Character Classification of Tigalari Script

Sachin S. Bhat, Alaka Ananth, Rajashree Nambiar, and Nagaraj Bhat

Abstract Image classification and optical character recognition are important research areas in computer vision. With advancement in machine learning and deep learning techniques, these fields are attracting lot of researchers to develop models with near human perfection. Many character recognition models are available for modern languages. But, it is still a challenging task to analyze the handwritten text in Indian scripts. It is further complex for the scripts with large alpha syllabary and complex nature. This paper proposes a technique for the recognition and classification of ancient Tigalari characters from the handwritten text. Tigalari is widely used in coastal Karnataka and Kerala for documenting Sanskrit, Tulu, and Malayalam languages. Method involves the creation of database, design of deep convolution neural network (DCNN)-based architecture to classify the text, training the model with the data and recognizing text using test set. Being an inception model for this script, proposed method classifies 46 basic Tigalari characters with an impressing accuracy of 98.55%.

Keywords Tigalari · Tulu · Kannada · Character recognition · Convolutional neural network · VGG16

1 Introduction

Image classification is a process of assorting the images into classes in accordant with their resemblance. This is an extensive application of image recovery that simplifies the search for those images with specific visual content across an image dataset. Classification techniques assume that certain features represent the image in the problem domain, and that individual features belong to one of the many discrete and sole classes.

S. S. Bhat (✉) · R. Nambiar · N. Bhat
Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Karnataka 574115, India

A. Ananth
NMAM Institute of Technology, Nitte, Karkala, Karnataka 574110, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_20

Classification model uses two methods: testing and training. Prominent attributes of the image features are extracted in the training level and a unique representation of individual classification family called training class. In the next level, these feature-space partitions are used to classify image features. From the last few years to solve the classification problem, the research community is enthusiastic on machine learning (ML) and deep learning (DL) problem. Since 2015, deep learning algorithms are becoming widely popular and getting an increasing attention.

ML and DL enable the machines to mimic the human intelligence. It is all about making machines to do things at which humans are better at present. Alan Turing in his famous work starts with a question, 'Can machines think?' Which question is still unanswered. If the behavior of a machine is indistinguishable from that of a human being, then we must accept that the machine can indeed think. This is probably an important step, in the ongoing journey to further improve the abilities of computing machines. The main objective of these techniques is to make the machines more smart and effective in dealing the issues important to humans. Some of the major ML algorithms used for text classification so far are support vector machines (SVM), Ridge classifier, Naive Bayes, K-nearest neighbor, etc. Apart from these, convolutional neural network-based deep learning approaches are swiftly overtaking the traditional ML models in the recent past.

2 Tigalari Script

Having evolved from Tamil-Brahmi, Tigalari finds its origin in Grantaezhuttu. Tigalari was widely used to write Tulu, Kannada, and Malayalam languages in olden days. It was also used by the Shivalli (Tulu speaking), Havyaka (Kannada/Malayalam speaking), and Kota (Kannada speaking) communities to write Sanskrit religious texts. Also known as Arya-Ezhuttu in Kerala, this is a familiar script seen since 1200 CE scripted on almost all old manuscripts found along Karavali (coastal), Malenadu (Western Ghats) of Karnataka and northern Kerala [1]. The orthography of Tigalari is influenced by other dominant languages/scripts like Nandinagari, Kannada, Tamil Grantha, and Malayalam. Cladwell [2] in his magnum opus called this as one of the most evolved Indian scripts. In 1841, the first Karnataka printing press was started by German missionaries in Mangaluru. They started to print the literature in the Kannada script instead of Tigalari. Tigalari was hardly used in printed format which led to the disuse of the script turning to it extermination in a gradual manner. But, a vast amount of documents and manuscripts can be found in this script, which is of great importance because of diverse subjects addressed by them. Large number of these manuscripts are being preserved, catalogued, digitized, and studied by several institutions today.

There is a lack of material or research available for accurately dating the introduction of this script to these regions and the reasons for doing so when there were several other scripts that were actively used here at the time to write Sanskrit. It is commonly stated by those who have studied this script that due to geographic

isolation of this region created by the Western Ghats to the east and the Arabian Sea to its west, Tigalari further evolved independently from the Chola Grantha Script. Besides this, if we take into account several prominent character constructions, it retains from the Chola Grantha script and the fact that it was a fully formed script when it was used in Sarvamoola-Grantha of Madhwacharya, and it seems possible that Tigalari might have existed at least hundred years before twelfth century.

The regions where Tigalari was used have been centers of learning for centuries and still continues to have high levels of literacy. It naturally follows that the manuscripts written in these regions, majority of which are in Tigalari script, to be of great literary and scientific value. These manuscripts cover a wide range of subjects such as: medicine, various sciences, Vedas, Sutras, Upanishads, mathematical formulae, daily accounts, astronomy, aesthetics, and philosophy to name a few.

Govt. of Karnataka has introduced Tulu and Tigalari in schools of Mangaluru and Udupi districts through Tulu Sahithya Academy. Text books, Web lessons, and manuals are being provided to learn this script. Scholars like Dr. K. P. Rao and Vaishnavi Murthy are working hard to introduce Unicode complaint typeface for Tigalari.

Tigalari has 16 swaras (vowels) including an Anusvara (am) and Visaraga (ah) which do not have an existence of their own and are pronounced with consonants. Letters vocalic rr and ll are rare and only used in Sanskrit texts. We have excluded these four swaras of ‘am’, ‘aha’, ‘rr’, and ‘rr’ from our dataset. Independent swara has their corresponding dependant vowel signs except for ‘a’ which combine with the pure vyanjanas to produce the consonant sound. These vowel signs can be termed as semi-vowels. Tigalari also has 34 vyanjanas (consonants) from ‘ka’ to ‘ḷa’ like similar to the scripts of Kannada and Malayalam. Consonants with semi-vowels together constitute ‘Gunitakshara’. In the orthography of Tigalari, these Gunitakshara are phonetically ordered similar to other Indic scripts. ‘Virama’ symbol is used to present the consonant in its pure form without inherent vowel ‘A’. There are three basic ligatures in Tigalari. First one is a conjunct character which is a combination of a consonant/semi-vowel and another consonant/semi-vowel. This is called as a vattakshara. Here, an independent consonant is vertically stacked below a base character. A large number of vattaksharas can be produced in Tigalari by the combination of two or three consonants. Second is a conjunct character where vowels U and Ru used as a vattu with consonant conjuncts are placed below the vattakshara. Third case occurs when a special character called Rephais followed by a conjunct or consonant and ends with a virama. Table 1 shows the orthography of Tigalari script. Characters are digitally represented in Srihari font of Pada keypad [3].

3 Preparation of Dataset

As no work has been reported yet on Tigalari, major task in OCR will be the creation of character patterns. Handwritten characters are collected from 42 native Tuluvas

Table 1 Tigalari character classes

Vowels (*Swara*)								
a	ā	i	ī	u	ū	ṛ	ṝ	ḷ
e	ai	o	au	am	ah			
Consonants(*Vyanjana*)								
ka	kha	ga	gha	na	ca	cha	ja	jha
ṭa	ṭha	ḍa	ḍha	ṇa	ta	tha	da	dha
pa	pha	ba	bha	ma				
ya	ra	la	va	śa	ṣa	sa	ha	ḷa

between the ages of 18–45. Among the writers, 19 are Male and 23 are Female. They are either graduates or in the final year of graduation. All writers have Tulu as their mother tongue and minimum knowledge of Tigalari script. They were asked to write down the Varnamala (character set) on a form as shown in Fig. 1 for several number of times in different styles. Later, these papers are scanned using Canon E560 scanner and saved in .jpg format. Characters are extracted from the sheets using machine learning segmentation tool. These characters are manually grouped into different classes. 94% of the characters are obtained in this fashion. No data augmentation is used for this dataset.

For remaining 6% of the characters, ancient Tigalari manuscripts are scanned and saved in the earlier mentioned way. Noise removal, preprocessing, and character reconstruction are done on the manuscripts. Phase congruency-based binarization is used to binarize the images of manuscripts, whereas geodesic morphological operators are applied to reconstruct the shape of characters in the script [4, 5]. Similar ML-based segmentation tool is used to extract the text. Characters belonging to the earlier mentioned 45 classes are picked and are added to the database. Nearly 150 palm leaf inscriptions are scanned in this way, and around 5200 characters are constructed.

Fig. 1 Sample of Varnamala character set

92 characters of 46 classes are used to develop the database. Training, test, and validation sets are chosen in a proportion of 3:1:1. All the images are resized to 32 × 32 format. Few random samples taken from six different classes of the dataset are shown in Fig. 2.

Fig. 2 Randomly chosen character images

4 Literature Review

With the spread of computer technology in working places and homes, automatic processing of paper documents is rapidly gaining importance in India. A short description of the encouragement of the progress in OCR of Indian scripts can be seen in [6]. Main scripts in India are Devanagari, Bangla, Gurumukhi, Odiya, Gujrati, Urdu, Kannada, Malayalam, Tamil, and Telugu. All these scripts were acquired by ancient Brahmi except Urdu. Devanagari which evolved from Nagari family of scripts is used to write Sanskrit, Hindi, Marathi, and many other north Indian languages. First work in Devanagari is accounted to Sethi [7]. They proposed an inception model to extract features from hand printed Devanagari characters. Though many works were reported in the later period, it is only in 2006 a full-fledged classifier has been designed by Sharma [8]. Author proposed a quad classifier-based method to recognize offline Hindi characters with a reported accuracy of 80%. Fuzzy model for handwritten Hindi numerals were proposed in [9]. Normalized distances are obtained using box approach. Structural features like horizontal bars and character strokes are employed in [10]. Compound handwritten OCR with Zernike moments is proposed in [11]. Integrated segmentation with hidden Markov model is proposed for Hindi word recognition in [12]. Multiple classifiers with local and global features are used in [13] to detect Devanagari characters. This is the only available OCR in Devanagari using CNN.

Bangla script is used by Bengali, Axomia languages in India and Bangladesh. Urdu is an India-Aryan language written in Nastaliq script. Kahmiri, Gurumukhi, and Shahmukhi are derived from Sharada. It is noteworthy that the first copy of Mahabbharata manuscript discovered was in Sharada script. Southern Brahmi or Tamil-Brahmi gave rise to five Dravidian scripts. Two from the Kadamba script of Halegannada namely the modern Kannada and Telugu. Tamil, Malayalam and Tulu scripts developed through Grantha-lipi.

Bangla is the only Indian language where remarkable amount of work is reported related to OCR. It is second widely spoken language in India. Bangla is an official language of Bangladesh, West Bengal, Tripura, Assam, Jharkhand and Sierra Leone a West African country. In general, Bengali has 11 vowels and 39 consonants. Recently, there are few applications of deep learning techniques for Bangla numeral classification [14, 15]. Halima Begum et al. [16] use artificial neural network to work on their customized dataset with 79.4% accuracy rate. A CNN approach [17]

delivered 85.36% test accuracy. A combination of multilayer perceptron and SVM algorithms have shown 80.9% recognition accuracy on their dataset [18].

Beside these two scripts, very little work is reported in other Indian languages. Unavailability of database is the main reason to be attributed. Sheshadri et al. [19] proposed a multilayer approach to classify Kannada characters. Few others have used traditional image processing approaches with machine learning like SVM [20, 21], divide and conquer [22], wavelet [23], SDTW [24], Tesseract tool [25], and principle component analysis [26] for Kannada or Telugu. Rahiman et al. used structural pattern-based classifier for Malayalam characters [27]. Apart from this statistical [28], gradient [29] and transform domain [30] features are also employed for this task. Even traditional ML and neural network-based algorithms also proved with better results [31–34]. CNN-based model was used for basic Tulu character classification which was an inception model in this language [35].

5 Methodology

Proposed work is inspired by the most popular and efficient DCNN model of transfer learning, called VGGNet. It is considered as one of the best vision models even now that outperforms many of the other networks. In 2015, Vision-Geometry-Group introduced it on ImageNet database for ILSVRC. ImageNet database consists more than 14 million annotated images ranging up to thousand classes. It is the first model to use the smallest receptive field. VGG has a combination of 16 convolutional layers (CL) and fully connected layers (FCL). Many versions of VGG like VGG19 are introduced later. In every CL, it uses a filter size of 3X3. Compared to this, earlier deep learning models used larger receptive fields like 7×7 or 11×11. Small-size filters help broaden network scope with greater discrimination.

The proposed network takes input images that are transferred over a series of layers. The layer represents one processing step which can be a CL, FCL, or a pooling layer. Equation (1) illustrates how CNN runs through layers in forward pass.

$$i_1 \rightarrow l_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_{n-1} \rightarrow l_{n-1} \rightarrow i_L \rightarrow l_n \rightarrow z \tag{1}$$

The input i_1 goes through processing in various layers. i_2 is an output of first layer which is fetched as an input to second layer and so on. Parameters involved in processing of each layer are denoted by tensor $l_1, l_2 \ldots l$n. This operation continues until the completion of all layers in the CNN, which outputs in. Yet one additional layer is introduced for backpropagation, a process that learns good values of parameters in the CNN. Suppose the issue is a grouping of images with classes C. A commonly used strategy is to output in as a C dimensional vector, whose *i*th entry encodes the prediction. If you know all the parameters in the model, then it is ready to be used for prediction, which involves running the network forward. Pseudocode of training the proposed model is given in Algorithm 1.

Algorithm 1 : Pseudo code for training the proposed model

```
1. Input: Model, T, t // Model = Keras sequential model, T = training dataset, t = test
   dataset
2. Output: Trained model
3. Start: // adding convolutional and pooling layers
        Model.add(xConvy), //x = number of filter, y=kernel, Conv=convolution
        Model.add(xAF), //x = convolution output, AF = activation function
        Model.add(xPy), //x = pooling size, y = stride, P = pooling layer
        // adding convolution and pooling layers until optimum test results are
        obtained(iterative investigation)
        Model.add(xFc) //x = number of neurons, FC = fully connected layer
        Model.add(softmax) // final layer for the classification of 46 classes
        Model.compile(Optimizer, accuracy measure)
        Model.fit(T, epochs, validation split)
4. Stop;
```

Like VGGNet, there are numerous small sequential convolutional and FC layers in our model too. Instead of 224 × 224, we used 32 × 32 grayscale images in input layer. After repeated investigation, we have observed that even after extracting a few convolutionary layers, the test results remain the same. A rigorous iterative investigation has been carried out to identify these layers are being removed. Customized FC layers have two dense, a flatten, and a dropout layer. A number of channels in the dense layers are fixed to 1024 with a dropout ratio of 0.5. Output is a softmax layer with 46 classes. Different hyperparameters are tuned and performances are measured. Activation function, optimizers, and learning rates are chosen to obtain the optimal solution. The architecture of this model is shown in Fig. 3.

6 Results and Evaluation

This section deals with the obtained result through the proposed model and their comparison with the available literature. The Tigalari database consists of 92,000 images, and the experimental models were developed in Anaconda programming environment using deep learning API Keras. This was implemented with the support of Google Colab GPU platform. The training, test, and validating sets are taken in the proportion of 3:1:1. Network was trained with a batch size of 500 samples for 15 epochs. On validation set, model showed an impressing accuracy of 98.55% on the validation dataset. It can be concluded from Fig. 4 that the model is neither underfitting nor overfitting. Top five misclassified points with the probability of misclassification are shown in Table 2.

Following three standard measurement systems were employed in the performance evaluation of the model.

Fig. 3 Proposed architecture

Fig. 4 Train and validation accuracy

Table 2 Top five misclassified classes

Actual class	Predicted class	Loss	Probability
ന	മ	2.64	0.07
അ	ആ	3.07	0.05
ഞ	ന	3.14	0.04
ശ	സ	3.30	0.04
ഌ	ഉ	4.70	0.02

Table 3 Confusion matrix of classifier

Confusion matrix	Predicted	
Actual	True positives (TPs)	False negatives (FNs)
	False positives (FPs)	True negatives (TNs)

(i) Confusion matrix: is a specific table (Table 3) which describes the classification performance as shown in Fig. 5.
(ii) Precision, recall: Precision is a ratio of TPs to the sum of TPs and FPs where recall is a ration of TPs to the overall TPs and FNs. These are the classification statistics on class predictions given in Figs. 6 and 7.
(iii) Test accuracy: is a trained model performance on unseen data. Model identifies a given random sample against 46 different classes with an accuracy of 99.8% as shown in Fig. 8.

Validation accuracy of the proposed model is compared with the original VGGNet16 as well as simple seven-layered CNN architecture with similar set of

Fig. 5 Confusion matrix

Fig. 6 Precision matrix

Fig. 7 Recall matrix

parameters. Proposed model converges better than VGG and outperforms CNN as shown in Fig. 9.

7 Conclusion

In this paper, an extensive study on offline recognition of characters in Tigalari documents is proposed. A database comprising of 92,000 Tigalari characters of Varnamala is prepared for the recognition task. A deep learning CNN architecture of 12 layers is developed for character classification. Depth of the convolutional layers is deliberately restricted as the features in text are less compared to the features of semantic

Fig. 8 Test accuracy on a randomly selected character

Fig. 9 Validation accuracy and loss comparison of proposed versus VGG versus DCNN

scene image. This model incorporates CLs, FCLs, activation function, optimizer, maxpooling layer, stride, dropout layer, and classifiers. These hyperparameters are fine-tuned with numerous permutations and combinations to attain maximum efficiency. Obtained result is ascertained with number of evaluation methods like train-validation accuracy, confusion matrix, precision–recall matrix, ROC of multiclass and misclassified classes. Train and validation accuracies are compared with the original model of VGG to establish the effectiveness of the proposed method.

References

1. R. Caldwell, *Comparative Grammar of Dravidian or South Indian Family of Languages* (Trübner &Co., London, 1856)
2. K.Y. Vaishnavi Murthy, Preliminary proposal to encode Tigalari script in Unicode (2017)
3. https://thetulufont.com/
4. S. Bhat, G. Seshikala, Preprocessing and binarization of Inscription images using phase based features. in *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)* (IEEE, 2018)
5. S. Bhat, G. Seshikala, Restoration of characters in degraded inscriptions using phase based binarization and geodesic morphology. Int. J. Recent Technol. Eng. **7**(6), 1070–1075 (2019)
6. U. Pal, B.B. Chaudhuri, Indian script character recognition: a survey. Pattern Recogn. **37**(9), 1887–1899 (2004)
7. I.K. Sethi, B. Chatterjee, Machine recognition of constrained hand printed Devanagari. Pattern Recogn. **9**(2), 69–75 (1977)
8. N. Sharma et al., Recognition of off-line handwritten Devanagari characters using quadratic classifier, in *Computer Vision, Graphics and Image Processing*. (Springer, 2006), pp. 805–816
9. M. Hanmandlu, O. Murthy, Fuzzy model based recognition of handwritten numerals. Pattern Recogn. **40**(6), 1840–1854 (2007)
10. V. Bansal, R.M.K. Sinha, On how to describe shapes of Devanagari characters and use them for recognition. in *Proceedings of Fifth International Conference on Document Analysis and Recognition* (1999), pp. 410–413
11. K. Kale, P.D.S. Chavan, M. Kazi, Y. Rode, Zernike moment feature extraction for handwritten Devanagari (Marathi) compound character recognition. Int. J. Adv. Res. Artif. Intell. **3**(1), 68–76 (2014)
12. S.D. Connell, R.M.K Sinha, A.K. Jain, Recognition of unconstrained on-line Devanagari characters. in *IEEE International Conference on Pattern Recognition* (2000), pp. 368–371
13. M. Avadesh, N. Goyal, Optical character recognition for Sanskrit using convolution neural networks. in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)* (IEEE, 2018)
14. S.M.A. Sharif, N. Mohammed, N. Mansoor, S. Momen, A hybrid deep model with HOG features for Bangla handwritten numeral classification (IEEE, 2016), pp. 463–466.
15. M. Shopon, N. Mohammed, M.A. Abedin, Bangla handwritten digit recognition using autoencoder and deep convolutional neural network. in International Workshop on Computational Intelligence (IWCI) (IEEE, 2016), pp. 64–68
16. Halima Begum et al., Recognition of handwritten Bangla characters using Gabor filter and artificial neural network. Int. J. Comput. Technol. Appl. **8**(5), 618–621. ISSN: 2229-6093
17. Md. M. Rahman, M.A.H. Akhand, S. Islam, P.C. Shill, Bangla handwritten character recognition using convolutional neural network. Int. J. Image Graphics Signal Process. **8**, 42–49, (2015)
18. N. Das, B. Das, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri, Handwritten Bangla basic and compound character recognition using MLP and SVM classifier. arXiv preprint arXiv:1002.4040
19. K. Sheshadri, S.K. Divvala. Exemplar driven character recognition in the wild. in *British Machine Vision Conference* (2012)
20. M. Thungamani, P.R. Kumar, K. Prasanna, S.K. Rao, Off-line handwritten Kannada text recognition using support vector machine using Zernike moments. Int. J. Comput. Sci. Netw. Secur. **11**, 128–135 (2011)
21. T.V. Ashwin, P.S. Sastry, A font and size-independent OCR system for printed Kannada documents using support vector machines. Sadhana **27**, 35–58 (2002)
22. M.M. Prasad, M. Sukumar, A.G. Ramakrishnan, Divide and conquer technique in online handwritten Kannada character recognition. in *Proceedings of the International Workshop on Multilingual OCR* (2009), pp. 1–7

23. R.S.R. Kunte, R.S. Samuel, Online character recognition for handwritten Kannada characters using wavelet features and neural classifier. IETE J. Res. **46**(5), 387–393 (2000)
24. R. Kunwar, P. Mohan, K. Shashikiran, A. Ramakrishnan, Unrestricted Kannada online handwritten Akshara recognition using SDTW. in *Proceedings of International Conference on Signal Processing and Communications* (2010), pp.1–5
25. G.K. Prasad, I. Khan, N.R. Chanukotimath, F. Khan, Online handwritten character recognition system for Kannada using principal component analysis approach: for handheld devices. in *World Congress on Information and Communication Technologies* (2012), pp. 675–678
26. R. Fernandes, A.P. Rodrigues, Kannada handwritten script recognition using machine learning techniques. in *2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* (IEEE, 2019)
27. R. Singh, M. Kaur, Ocr for telugu script using back-propagation based classifier. Int. J. Inf. Technol. Knowl. Manage. **2**(2), 639–643 (2010)
28. M.A. Rahiman, A. Shajan, A. Elizabeth, M. Divya, G.M. Kumar, M. Rajasree, Isolated handwritten Malayalam character recognition using HLH intensity patterns, in *Second International Conference on Machine Learning and Computing (ICMLC)* (IEEE, 2010), pp. 147–151
29. B.S. Moni, G. Raju, Modified quadratic classifier for handwritten Malayalam character recognition using run length count. in *International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT)* (IEEE, 2011), pp. 600–604
30. R.G. Salomon, Brahmi and Kharoshthi. The World's Writing Syst. 373–383 (1996)
31. G. Raju, Recognition of unconstrained handwritten Malayalam characters using zero-crossing of wavelet coefficients. in *International Conference on Advanced Computing and Communications, ADCOM 2006* (IEEE, 2006), pp. 217–221
32. M. Manuel, S. Saidas, Handwritten Malayalam character recognition using curvelet transform and ANN. Int. J. Comput. Appl. **121**(6)
33. K. Manjusha, M.A. Kumar, K. Soman, Reduced scattering representation for Malayalam character recognition. Arab. J. Sci. Eng. 1–12 (2017)
34. N.V. Neeba, C.V. Jawahar, Empirical evaluation of character classification schemes. in *Seventh International Conference on Advances in Pattern Recognition* (IEEE, 2009), pp. 310–313
35. S. Bhat, G. Seshikala, Character recognition of Tulu script using convolutional neural network. in *Advances in Artificial Intelligence and Data Engineering* (Springer, Singapore), pp. 121–131

Handwritten Character Recognition Using Deep Convolutional Neural Networks

R. Shashank, A. Adarsh Rai, and P. Srinivasa Pai

Abstract The automatic detection and recognition of characters in images are an important problem in various applications. The traditional shallow networks in machine learning have limitations in image classification due to their inability to effectively utilize the spatial relationships between the pixels of the image. But the incredible advances in deep learning methods and deep architecture in the recent years have opened doors to the possibility of employing these techniques. In this paper, a deep convolutional neural network (CNN) with minimal preprocessing for the effective classification of handwritten characters has been proposed. The application of this network yielded an accuracy of 99.50 and 94.66% on the test data of MNIST and EMNIST datasets, respectively.

Keywords Convolutional neural networks (CNNs) · Character recognition · Deep learning

1 Introduction

Optical character recognition represents the process of extracting text embedded in images and converting it into a digitized text format. Text data embedded in images can generally be classified into two classes: printed text and handwritten text. While a lot of work has been done for the recognition of printed characters [1–4], including open-source systems like the Tesseract [5], these systems are not suited for general purpose text extraction and tend to perform poorly for handwritten text.

Handwritten character recognition (HCR) tends to be a harder problem, due to the sheer variability of text written by hand. HCR systems find extensive application in industries that generate large corpuses of handwritten documents. These systems

R. Shashank · A. Adarsh Rai · P. Srinivasa Pai (✉)
Department of Mechanical Engineering, NMAM Institute of Technology Nitte, Udupi, Karnataka 574110, India
e-mail: srinivasapai@nitte.edu.in

A. Adarsh Rai
e-mail: adarsh.rai@nitte.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_21

could be used for automatic processing of cheques in banks, to read patient forms in healthcare, and to read grades in universities to name a few.

HCR can either be online or offline. Online HCR involves automatic processing of a message as it is being written, whereas offline HCR deals with text in scanned images [6]. These variants differ mainly in the way the input is fed into the classifier. Previous methods in classifying handwritten characters tend to use handcrafted feature extraction methods [7, 8]. These manually extracted features tend to be limited by the prior knowledge of the user [9], leading to a degradation of the inputs to the classifier, thereby yielding poor results.

Deep learning (DL) is a rapidly developing class of machine learning algorithms, which employ multiple nonlinear hidden layers, allowing it to capture complex relationships in the inputs. This is done by directly extracting features from raw data in a number of nonlinear hidden layers [10]. The features extracted by DL methods like convolutional neural networks (CNNs) tend to be superior to manual engineered features [11]. These techniques have performed better than the traditional neural network models like multilayer perceptron (MLP) [12]. The increase in processing power of processors and graphics processing units (GPUs) has made training a large number of parameters associated with these deep networks possible. This has led to the development of several deep learning frameworks. TensorFlow is one of the most popular one released by Google in 2015 [13]. High-level libraries like Keras, which are built on top of TensorFlow, are also available. These are easier for writing native TensorFlow codes. Thus, this paper proposes a deep convolutional neural network model with minimal preprocessing for the recognition of handwritten characters using Keras. This paper is divided into the following Sect. 2—Reviewing related research in handwritten character recognition, Sect. 3—Convolutional neural network, Sect. 4—Proposed CNN architecture, Sect. 5—Results of the CNN, and finally Sect. 6—Conclusions drawn from the application of the model.

2 Related Work

One of the earliest works of using CNNs for character recognition was probably the LeNet proposed by Yann LeCun et al., back in 1998 [14]. They found that CNNs specifically designed to deal with 2D shapes performed the best when it came to character recognition.

Later with the development of other machine learning and feature engineering processes, the application of methods like K-nearest neighbor and support vector machines were explored [15]. But most of these methods had limitations due to their shallow architecture and manual feature extraction.

The use of deep networks to solve computer vision problems with higher accuracies is increasing. It was found that the deep CNNs perform exceptionally well in problems such as the ImageNet classification [16].

Deep CNNs have shown a marked improvement over previous methods, when used for Arabic handwritten character recognition with minimal preprocessing,

employing batch normalization and dropouts [17]. Techniques exploiting the hierarchical feature extraction process of CNNs have been used successfully in the recognition of handwritten characters in Tamil language [18]. Similar techniques have also been explored for Chinese characters [19] and Bangla character recognition [20].

3 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a class of feed forward deep neural networks that are similar to MLPs, but mainly differ in the way that their weights are organized.

The weights in the convolutional layer have a matrix of weights called kernel or filters, which slide through the inputs performing the convolution operation at every stage. The outputs of these convolution operations generate a feature map, where each feature in the feature map corresponds to the local receptive field of the input, where the kernel performs the operation. The output of the nth convolutional layer with filter size ($M \times M$) is computed as:

$$Y_{i,j}^{n} = B_{i,j}^{n} + \sum_{a=1}^{M} \sum_{b=1}^{M} W_{i,j}^{n} X_{(i+a)(i+b)}^{n-1} \tag{1}$$

where $B_{i,j}^{n}$ is the bias matrix and $W_{i,j}^{n}$ is the filter in the nth convolutional layer.

The learned weights in the kernel are common throughout the inputs of the given convolutional layer. This helps reduce the number of trainable parameters by many folds, thereby reducing the scope for over fitting, which is rampant in MLPs [21].

The use of subsampling or pooling layer further reduces the number of learnable parameters of the network. There are different methods of pooling, the most common method being max pooling. In this there is again a kernel sliding through the inputs of the pooling layer, where the most prominent element within the kernel is obtained. However, in the pooling layer, the local receptive field of the kernel does not overlap. This mechanism provides a method to extract the most prominent features within the kernel size allowing a reduction in overall number of parameters.

Convolutional layers are followed by a nonlinear activation function like the rectified linear unit (ReLU). This is used to overcome problems related to the plateauing of the loss function during backpropagation [22]. Batch normalization, a normalization technique is sometimes used between the convolutional layer and the activation function, and this technique is proven to reduce the training time of deep networks by reducing the internal covariate shifts [23].

A typical CNN architecture consists of multiple convolutional blocks, which consists of several convolutional and pooling layers. Each convolutional block extracts features from the inputs and the complexity of the features extracted increases deeper inside the model [24]. These convolutional blocks act as feature extractor, that extracts the most relevant features that is fed into a classifier, normally

a fully connected layer [25]. The features that are extracted are fine-tuned by backpropagation during training.

The softmax function is applied to the outputs of output layer to determine probabilities for each class. The class with the highest probability would be the predicted class of the input. The probabilities are computed as:

$$x_i = \left(\frac{e^{x_i}}{\sum_{k=1}^{N} e^{x_k}} \right) \tag{2}$$

These CNN models have demonstrated to perform better with an increase in their depth, i.e., number of convolutional blocks [26]. But the– depth of the model is limited by the size of the data available, as overfitting becomes predominant as the depth increases.

4 Proposed CNN Architecture

For this task, Keras has been used to build a deep convolutional neural network and this network has been trained on two datasets, i.e., MNIST dataset [14] and EMNIST Letters dataset [27]. The MNIST dataset consists of handwritten digits from 0 to 9. It has 60,000 training samples and 10,000 test samples. The EMNIST Letters dataset consists of handwritten letters of the English alphabet both capital and cursive. It has 124,000 training samples and 20,800 test samples. The datasets used have letters in the form of binary 28 × 28 pixel images.

The pixels of the images have been normalized by dividing them by 255, i.e., the maximum pixel intensity before using them for training, thereby reducing the range of pixel values to 0–1. All the training samples of the datasets are retained for training and half of the test samples are used for validation of the model and the other half is used for determining the percentage accuracy of the model after the training process has been completed.

Figure 1 shows the basic architecture of the proposed model, and the particulars of the same are explained below.

Batch normalization and parametric rectified linear unit (PReLU) activation is applied after every convolutional and fully connected layer. This activation operation retains the positive pixel values and returns the product of the negative value and a term alpha, which is learnt from the data [28]. PReLU is used instead of ReLU as it can compute negative inputs and yields a lower error rate than the ReLU [29].

The first input layer is followed by three convolutional blocks with 256, 128, and 64 filter maps, respectively. The convolutional layers in these blocks have a kernel size of (3, 3), same padding and have a stride of 1. The max pooling layers that follow have a pool size of (2, 2) and a stride of 2. Thus, the feature map size reduces by half through every convolutional block, as every block has a max pooling layer. The feature maps reduce in size as 28-14-7-3 through these blocks. These blocks

Fig. 1 Basic architecture of the proposed model

are followed by a single convolution layer without padding, so that it converts the two-dimensional array obtained from the convolutional blocks to a one-dimensional vector, so that it can be fed into fully connected layers for the final classification. This layer also has a kernel size of (3, 3). It has 32 filters and the output of this layer will be of the dimension (32, 1, 1). This one-dimensional vector would contain all the features extracted by the convolutional layers preceding it.

Dropout layers are used after some of the layers, as a regularization layer to reduce over fitting and improve the generalization of the networks, by setting a fraction of inputs to zero [30].

The output layer has either 10 or 26 neurons depending on the dataset used. The output layer uses a softmax activation function to represent the probability that a character belongs to a class.

During the training of the model, categorical cross-entropy [31] is used as the cost function for backpropagation, since it is ideal for multiclass classification problems. The AMSGrad optimizer was used to find the minima of the loss function. This optimizer is a modification of the Adam optimizer and deals better with the convergence issues associated with the latter [32].

Figures 2 and 3 show percentage accuracy and cross-entropy loss versus number of epochs for the MNIST dataset. Figures 4 and 5 show the same for the EMNIST Letters dataset.

Fig. 2 Percentage accuracy versus number of epochs for MNIST dataset

Fig. 3 Cross-entropy loss versus number of epochs for MNIST dataset

Fig. 4 Percentage accuracy versus number of epochs for EMNIST Letter dataset

5 Results and Discussion

The whole training section of both the datasets is used for training the network, whereas the testing samples provided in these datasets are divided into two halves: validation and testing. The validation set is used during the training process to know the performance of the model, during training.

Fig. 5 Percentage accuracy and cross-entropy loss versus number of epochs for EMNIST dataset

The network is trained for 25 epochs on both the datasets separately. The testing accuracy for both the MNIST and EMNIST Letters datasets is 99.50% and 94.66%, respectively. The final training and validation accuracies were 99.98 and 99.64% for the MNIST dataset and 97.59 and 94.69% for the EMNIST Letters dataset, respectively. The testing accuracy obtained by the proposed method on the MNIST dataset is superior to the testing accuracy of 98.46% obtained by Younis et al. [33] on the same dataset using a deep neural network with five fully connected layers.

Though the testing and training accuracies for the MNIST dataset are very close to each other, the network seems to have a slight over fitting problem for the EMNIST Letters dataset. The network also reached validation accuracy within 1% of the final value, within the first 5 epochs showing only marginal improvement thereafter indicating a scarcity of the data samples.

6 Conclusions

The network described in this paper successfully classifies English alphabets and numbers using the EMNIST Letters and the MNIST datasets with test accuracies of 94.66% and 99.50%, respectively.

However, the performance of the CNN plateaus out after the initial epochs showing marginal improvement after. The performance of the CNN is limited by the size of the dataset available, as they require large datasets and are prone to over fitting on smaller datasets.

References

1. T. Pavlidis, Recognition of printed text under realistic conditions. Pattern Recogn. Lett. **14**, 317–326 (1993)
2. S.N. Srihari, Recognition of handwritten and machine-printed text for postal address interpretation. Pattern Recogn. Lett. **14**, 291–302 (1993)
3. R. Seethalakshmi, T.R. Sreeranjani, T. Balachandar, A. Singh, M. Singh, R. Ratan, S. Kumar, Optical character recognition for printed Tamil text using unicode. J. Zhejiang Univ. Sci. A **6**, 1297–1305 (2005)
4. H.A. Al-Muhtaseb, S.A. Mahmoud, R.S. Qahwaji, Recognition of off-line printed Arabic text using hidden Markov models. Signal Process. **88**, 2902–2912 (2008)
5. R. Smith, An overview of the Tesseract OCR engine, in *Ninth International Conference on Document Analysis and Recognition* (ICDAR 2007) 2007 Sep 23 (vol. 2, pp. 629–633). IEEE (2007)
6. R. Plamondon, S.N. Srihari, Online and off-line handwriting recognition, a comprehensive survey. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 63–84 (2000)
7. N. Shanthi, K. Duraiswamy, A novel SVM-based handwritten Tamil character recognition system. Pattern Anal. Appl. **13**, 173–180 (2010)
8. N.A. Hamid, N.N. Sjarif, Handwritten recognition using SVM, KNN and neural network. arXiv preprint arXiv:1702.00723 (2017)
9. F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks, a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech. Syst. Signal Process. **72–73**, 303–315 (2016)
10. Y. LeCun, Deep learning & convolutional networks, in *2015 IEEE Hot Chips 27 Symposium* (HCS) (2015)
11. C. Bailer, T. Habtegebrial, K. Varanasi, D. Stricker, Fast dense feature extraction with convolutional neural networks that have pooling or striding layers, in *Proceedings of the British Machine Vision Conference 2017* (2017)
12. S.B. Driss, S. Ben Driss, M. Soua, R. Kachouri, M. Akil, A comparison study between MLP and convolutional neural network models for character recognition, in *Real-Time Image and Video Processing* (2017)
13. S.S. Girija, Tensorflow: large-scale machine learning on heterogeneous distributed systems. https://cse.buffalo.edu/~chandola/teaching/mlseminardocs/TensorFlow.pdf
14. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. Proc. IEEE **86**, 2278–2324 (1998)
15. P. Kumar, N. Sharma, A. Rana, Handwritten character recognition using different Kernel based SVM classifier and MLP neural network (a comparıson). Int. J. Comput. Appl. **53**, 25–31 (2012)
16. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. Commun. ACM **60**, 84–90 (2017)
17. K. Younis, A. Khateeb, Arabic hand-written character recognition based on deep convolutional neural networks. Jordanian J. Comput. Inf. Technol. **3**, 186 (2017)
18. P. Vijayaraghavan, M. Sra, Handwritten tamil recognition using a convolutional neural network, in *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)* (2014)
19. W. Yang, L. Jin, Z. Xie, Z. Feng, Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge, in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (2015)
20. S. Sen, D. Shaoo, S. Paul, R. Sarkar, K. Roy, Online handwritten Bangla character recognition using CNN, a deep learning approach. Adv. Intell. Syst. Comput. 413–420 (2018)
21. S. Lawrence, C.L. Giles, Overfitting and neural networks: conjugate gradient and backpropagation, in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium* 2000 Jul 27 (vol. 1, pp. 114–119). IEEE

22. K. Hara, D. Saito, H. Shouno, Analysis of function of rectified linear unit used in deep learning, in *2015 International Joint Conference on Neural Networks (IJCNN)* (2015)
23. S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift (2015)
24. N. Eikmeier, R. Westerkamp, E. Zelnio. Development of CNNs for feature extraction, in *Algorithms for Synthetic Aperture Radar Imagery* XXV, vol. 10647, p. 106470C. International Society for Optics and Photonics (2018)
25. M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in *European Conference on Computer Vsision*, pp. 818–833. Springer, Cham (2014)
26. S. Liu, W. Deng. Very deep convolutional neural network based image classification using small training sample size, in *2015 3rd IAPR Asian conference on pattern recognition (ACPR)*, pp. 730–734. IEEE (2015)
27. G. Cohen, S. Afshar, J. Tapson, A. Schaik, EMNIST: Extending MNIST to handwritten letters, in *2017 International Joint Conference on Neural Networks (ICJNN)*, pp. 2921–2926 (2017)
28. K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing Human-level performance on ImageNet classification, in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034 (2015)
29. B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network, in arXiv:1505.00853v2[cs.LG]
30. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(56), 1929–1958 (2014)
31. P. Golik, P. Doetsch, H. Ney, Cross-entropy vs. squared error training: a theoretical and experimental comparison, in *Interspeech*, vol. 13, pp. 1756–1760 (2013)
32. S.J. Reddi, S. Kale, S. Kumar, On the convergence of adam and beyond. arXiv preprint arXiv:1904.09237. 2019 Apr 19
33. K.S. Younis, A.A. Alkhateeb, A new implementation of deep neural networks for optical character recognition and face recognition. *Proceedings of the new trends in information technology*. 2017 Apr 25 pp. 157–162

Implementing Face Search Using Haar Cascade

Ramyashree and P. S. Venugopala

Abstract It is a known fact that human beings are able to distinguish distinct people primarily by observing the human face. Visual search engines are the ones that, when given the name, search for celebrity faces in a video file. This will be retrieval based on the data related to the file. The goal of this paper is to enhance the usability of video searches. Users would be able to input a video and a picture (passport size) of the person they are intended to search for, and then select the frames containing the given image from the uploaded file. This would be helpful for video capture purposes. This is especially useful for the people who are interested in a part of the video in which either they themselves or the person of their interest is available. This is achieved by removing all frames from the video and saving the extracted video in different directories. Facial detection method is used to recognize the faces in the frame, and these identified frames are used as training models for the next level of facial recognition. The facial recognition technique is used to differentiate the reference image from all the frames, to transform the selected frames to a video, and to play it for the user. This application faces many challenges like occluded face detection and recognition. Also, frontal view of the face will be necessary for this application to work successfully. These challenges can be taken into account as the future work for this particular application.

Keywords Face detection · Haar cascade · PCA · Viola–Jones technique

1 Introduction

In the current situation, facial identification and face recognition have become a significant area in machine perception science and thus the most effective applications

Ramyashree
Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi, Karnataka, India
e-mail: ramyashree.cs@sode-edu.in

P. S. Venugopala (✉)
NMAM Institute of Technology Nitte, Udupi, Karnataka 574110, India
e-mail: venugopalaps@nitte.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_22

are comprehension and analysis of the image [1]. Face detection is the essential testing area with the most heterogeneity. The implementations include video conferencing, image retrieval based on information and automated authorization, etc. The classifier will isolate the face parts from the context and will be trained using a series of non-facial sample results. Positions and sizes are extracted and scaled to the same sample size when a replacement input image is given. This brute-force search technique is used in most face recognition methods [2].

The purpose of this work is to help those users who wish to view only those parts of the video in which the person of their interest is present. The application will result in a series of frames containing the picture of the requested person which is then converted into a video and presented to the use. In this system, the user has to select an image that has to be searched in a video [3]. The user must also select a video in which the search has to be performed. The important phases are:

Face detection: The first phase is converting the video into frames and then performing face detection on all these frames [4]. Implementation of face detection using Viola–Jones technique is one of the approaches. This technique of face exploration is designed to prepare faces extremely rapidly while maintaining high recognition speeds [5]. There are three main responsibilities. The first is the display of another image, called a fundamental frame, which enables the highlights used by the locator to be processed easily. The second is a quick and effective classifier that is generated using the AdaBoost learning equation to select a few specific visual highlights from a wide range of potential highlights [6]. The third pledge is a technique for entering classifiers in a route that requires image districts to be disposed of immediately while focusing time on promising face-like regions. The faces thus identified are used as training images for the purpose of facial recognition [7].

Face recognition: Principal component analysis (PCA) is one of the approaches that is used for the recognition purpose [8, 9]. The frames which contain the required face are then extracted and then converted back into a video [10]. The method of initialization comprises the following operations:

(i) Get the initial set of face images referred to as a training set.
(ii) Calculate from the training collection the eigenfaces, holding highest eigenvalues only. These pictures of M describe the space faces. The eigenfaces, as new faces are met, it is possible to upgrade or recalculate.
(iii) In this M-dimensional space, calculate the distribution for by projecting his or her face pictures, each known individual into this space face [11].

The Haar cascade classifier is based on the Haar wavelet process of functionally analyzing pixels in the image into squares [12]. This utilizes "integral image" principles to measure the detected “features.” Haar cascades use the AdaBoost learning algorithm to pick a small number of important features from a wide collection to give classifiers an effective result and then use cascading techniques to detect the face in an image [13].

2 Literature Survey

The face detection, face identification, and face recognition have been carried out by several researchers.

Sakle [1] has suggested the different strategies of face detection. Confidentiality and identity have become the core problem in the modern world today. Face recognition plays a role major function in verification and recognition. In this paper, multiple current facial recognition techniques are evaluated and it was addressed.

Prasad [4], a report on the identification of features in real images is explained and also provides a brief understanding of the issue in identification of features in real images. The aims of this paper are to describe the definition and scope of image processing, to address the different steps and methodologies involved in the typical processing of images, and to apply image processing tools and processes to research frontiers.

Chora [14], it describes the extraction of the feature and different techniques used for this. One of the main areas of technology that continues to evolve and expand is facial recognition. This paper, together with its pros and cons, provides the different strategies. Three major categories classified here as feature-based approach, hybrid approach and holistic approach, one of the key methods that many researchers may use.

Menser and Muller [7] suggested one algorithm to identify faces with dynamic background pictures of color. Use color details in a face detection system focused on study of key components (PCA). Using a color analysis, a skin color likelihood image is created and the PCA is performed on this new image, instead of the luminance image.

Perveen et al. [10], in this article, facial expression recognition (FER) framework supports estimated facial expression space by facial features and Gini indices for facial expression identification. Demeanor awareness also focuses on facial expression information regions, so that eyes, eyebrows, and mouth are extracted from the input image. When meeting a human face, an isolation is made of a feature that helps to distinguish the facial features. The values of the variable factors are defined in order to explain one of the six standard facial expressions. The clustering algorithm is applied to the JAFFE range of 30 images, each of which has six normal facial expression images.

Stojmenovic [15] explains about Viewdle concepts in face recognition. Viewdle feature for face recognition is explained in Video Search Engine, powered by Reuters Labs. It analyzes sequence of data clips and filters for names. Furthermore, by translating speech to text it adds contextual information to the facial recognition results.

Ghazi Mohammed Zafaruddin et al. [16] explained the applications of Face Search Engines. This search engine is as benign as it can be. It looks for faces that are based on picture tags. You will then ask the search engine for a name and return all matching images of faces marked with the name of the user. But this search engine does not appear to be very accurate at the retrieval of the images.

Nikolskaia et al. [2] about PicTriev function explained and go a step forward by actively looking for something similar faces. Then what you do is upload a portrait shot or other face picture in the format of.jpg or.jpeg, with a scale not exceeding 200 KB, and the search engine will return similar photographs found online. This will run a demo with a range of well-known images too. It fits particularly well with pictures of celebrities.

Pinto et al. [17] explain the approaches. There are several approaches to face recognition that are commonly known as systematic and feature oriented approaches. Currently there are relatively few studies that compare these two methods. Nowadays, there is a huge rise in facial recognition research, largely because of the numerous negative incidents taking place. In the areas of image processing and patter recognition, human face identification has become a difficult issue. This paper proposes a modern human face recognition algorithm using the primitive Haar cascade algorithm combined with three additional weak classifiers.

3 Methodology

The implementation is carried out using MATLAB. Viola–Jones algorithm is used in this work for face detection. The classic technique for image processing is to ramp the feature vector dynamically to varying lengths and then run the fixed-length detector through those images [18].

Although there are numerous face search engines, most of them are used for online purposes. Also, these search engines are used to identify the faces in still images rather than a video. So this led to the idea of detecting and recognizing faces in a video. This kind of application would be helpful in tracking a person and for video surveillance purposes [19].

In this work, there are four main steps. The first step is converting the given input video into frames. These frames are then stored in a separate folder. In this step, upload an image which has to be recognized in the video frames. The second step is face detection phase. Each of the frames is checked for the presence of face using face detection algorithm. The detected face from each of the frame is stored in a separate folder. The third step is face recognition on the frame from which the face has been detected. For recognition purpose, the face that are detected are used and the uploaded photograph is used as the test image. If there is a match then that particular frame is extracted and saved in a separate folder [16]. The final step is extracting all these selected frames from the folder and converting them into a video. The application then displays the final video which consists only the required frames of the given input video.

MATLAB script is used for converting video to frames. The script reads the input video in avi format [20]. The frame rate of the video is taken and using a counter the video is traversed till the end of file is reached. For each counter increment, a frame is extracted and saved in the desired output folder.

3.1 Face Detection

Here, the implementation is done using MATLAB. It includes control flow statements, features of object-oriented programming, etc. This will produce very speed and accurate result due to feature-based method [14]. When doing this work, several limitations were found out. To overcome this enhancement, techniques are applied using Viola–Jones algorithm.

3.2 Face Recognition

This is the third main step in this work. There are several algorithms for face recognition. Popular recognition algorithms include principal component analysis which uses eigenfaces, linear discriminate analysis, elastic bunch graph matching fisher face, the hidden Markov model, and the neuronal motivated dynamic link matching.

3.3 Converting the Frames into Video

This is the reverse process of converting video into frames. The stored frames are retrieved one at a time, and each frame is inserted into the video. The result is the video with only the required frames containing the person of your concern.

3.4 Algorithm

Step 1: Read the video
Step 2: Read the test image
Step 3: Convert the input video into a number of frames
Step 4: Perform face detection on each and every frame
Step 5: Select the frames in which the faces have been detected and discard the frames in which there has been no detection of the faces
Step 6: Perform face recognition on the selected frames
Step 7: Store only those frames in a different folder which has face match with the test image and discard the remaining frames
Step 8: Extract these frames from the folder and convert it back into a video
Step 9: Return the extracted video as the output

The steps that are used are shown in following Fig. 1.

Fig. 1 Flowchart of video searching

4 Implementation Details

Viola–Jones algorithm [15] is used in this work. The typical approach to image processing will be to rescale the image data to various sizes, and then run the fixed size detector through these images. This method turns out to be very time-consuming due to the estimation of the various scale pictures. However, Viola–Jones developed an invariant scale locator that needs a comparable number of counts regardless of the distance. This finder is designed using a supposed important image and some straight rectangular highlights reminiscent of the Hair wavelets.

Fig. 2 Integral image

Fig. 3 Calculation of sum

4.1 The Scale Invariant Detector

Converting a given image that is input to the integral image is the very initial stage of Viola–Jones algorithm. This can be done using some calculations that is initially calculated the sum of the pixel that is above and left of the particular pixel. Then set every pixel to the above calculated sum. This is shown in Fig. 2.

With the four values of any given rectangle gives the values like shown in Fig. 3.

$$\text{Rectangle sum} = D - (B + C) + A$$

From Fig. 3, it is noted that the rectangle A is covered with the two rectangle B and C. So the calculated sum of A should be added. Viola–Jones algorithm decided that, it gives the acceptable outcome when the constancy is about 576 pixel. So totally using this 160,000 features can be build.

4.2 The Modified AdaBoost Algorithm

AdaBoost algorithm is established by Freund Schapiro in 1996 [17]. So the revised adaptation of AdaBoost algorithm is used in the Viola–Jones technique. One of the popular machine learning algorithms is the AdaBoost algorithm. This has the ability

Fig. 4 Cascade classifier principal component analysis algorithm

to build a very active classifier. The threshold value and polarity are the main features of this type of algorithm. The simple brute-force approach is used in this technique.

4.3 The Cascaded Classifier

The very determinant in the Viola–Jones algorithm is scanning. For the given image, scan with various times with different size. This process should be done for a small number of faces in images too. Using this technique, the part which is not a face can be subtracted, and this method is very fast approach. Strong classifier is contained in every step of cascade classifier. The task of this step is to identify that whether the subwindow contains a non-facial part or any facial part. Non-face parts are eliminated during this stage, and image contains a face that is moved to next step. This is demonstrated in the following Fig. 4.

PCA [10] includes a scientific method that changes various perhaps co-related factors into few co-related factors called chief parts. The most relevant section, though, represents as much of the changeability of the information as could be predicted, and each subsequent segment reflects as much of the remainder of the fluctuation as could reasonably be anticipated. Steps involved in this are prepare the details, deduct the mean, calculate the covariance matrix, calculate the own vectors and own values, and choose the main part.

5 Results and Discussion

Recognition has recently attracted attention and is beginning to be applied in variety of domains predominantly for security. There are some image search engines that search for the image of a particular person and there are no such video search engines.

So this led to the idea of video searching. User interface is developed using MATLAB to provide the input. An image and a video are the inputs for this application. Output is the part of the video in which the interested person is present. This application is ideal for those wanting to see only portions of their interest in the stream. The experimental results are shown in following Fig. 5.

	After selecting a video
	After selecting both the inputs
	Output after searching the video for test image

Fig. 5 Experimental results

Table 1 Performance of facial recognition

Number of samples	Recognition accuracy (%)	Execution time (s)
1	100	0.02
3	100	0.02
5	100	0.05
7	98	0.1
10	95	0.2
20	91	0.5

In the initial step, the user has to select the video, after selecting the video user have to select the test image from the dataset. After selecting both the inputs, system will show the presence of test image in the video and it will give suitable error messages when test image does not contain a face or test image is not present in the video. Accuracy percentage is calculated based on the classification accuracy rate (CAR). Based on the total number of samples and correct prediction, percentage was calculated.

$$\text{CAR} = (\text{Total Number of Faces} - \text{a total number of False Recognition}) \\ /(\text{Total number of Faces}) * 100\%.$$

Performance of facial recognition is calculated based on the number of sample as shown in Table 1.

The bar graph in Fig. 6 shows the data segment and compares the data using solid bars to represent the quantity. From this we can observe that, when the number of samples are less then percentage of accuracy is more and execution time is also less compared to more number of samples.

The above experiment is compared with local binary patterns (LBP). By this we observe that quality of analysis is high and false positive rate is always less in Haar cascade approach. The comparison between this samples is shown in Table 2.

The graphical representation is shown in Fig. 7.

6 Conclusion and Future Work

Face recognition has recently attracted attention and is beginning to be applied in variety of domains predominantly for security. Although there are some image search engines that search for the image of a particular person there are no such video search engines. So this led to the idea of video searching. Input is an image and a video and output is a part of the video in which the interested person is present. This application is ideal for those appearances wanting to see only portions of their interest in the stream. For photojournalists and photographers, for surveillance systems, and

Fig. 6 Graphical representation of accuracy rate

Table 2 Comparison between Haar and LBP

Number of samples	Execution time (s) in Haar cascade	Execution time (s) in LBP
1	0.02	0.10
3	0.02	0.10
5	0.05	0.128
7	0.1	0.435
10	0.2	0.672
20	0.5	0.925

for tracking a person in a video, this program may also be essential. The above experiment is compared with local binary patterns (LBP). By this we observe that quality of analysis is high and false positive rate is always less in Haar cascade approach

Face detection works well with photographs or video with a very small audience, but it is not that good as the number of users rises. And face recognition operates on the probabilistic approach, because there is a possibility to identify a similar face to that of the input image. Thus, the structure can be generalized to ensure the precise matching of images by selecting threshold values, which is an art in itself. As if the task is now taking a long time to process and return the result, this drawback can be taken into account and the performance can be increased by using parallel processing techniques. This application faces many challenges like occluded face detection and recognition. Also, frontal view of the face will be necessary for this application to

Fig. 7 Execution time using Haar cascade and LBP

work successfully. These challenges can be taken into account as the future work for this particular application.

References

1. M. Sakle, Study and analysis of different face detection techniques. Int. J. Comput. Sci. Inf. Technol. **5** (2014)
2. K. Nikolskaia, N. Ilya, A. Minbaleev, Development of the face recognition application using neural networks technology, in *2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)* (Sochi, Russia, 2019), pp. 87–91. https://doi.org/10.1109/ITQMIS.2019.8928392
3. Z. Shaaban, Face detection methods, recent researches in applied informatics and remote sensing. Int. J. Comput. Appl. (0965–8687) **55**(1) (2010)
4. D.K. Prasad, Survey of the problem of feature detection in real images. Int. J. Image Process. (IJIP) **6**(6) (2012)
5. Y.K. Singh, V. Hruaia, Detecting face region in binary image, in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (2015, 10–12 December)
6. R. Zwiggelaar, B. Rajoub, M.H. Yap, H. Ugail, A short review of methods for face detection and multifractal analysis, in *2009 International Conference on CyberWorlds. 978–0–7695–3791–7*. https://doi.org/10.1109/CW.2009.47
7. B. Menser, F. Muller, Face detection in color images using principal component analysis. Institute for Computer Science, Social Informatics and Telecommunication Engineering (2012)
8. V. Bhandiwad, B. Tekwani, Face recognition and detection using neural networks, in *International Conference on Trends in Electronics and Informatics ICEI* (2017) 978-1-5090-4257-9
9. A. Robert Singh, A. Suganya, Efficient tool for face detection and face recognition in color group photos. 978-1-4244-8679-3

10. N. Perveen, S. Gupta, K. Verma, Facial expression recognition system using statistical feature and neural network. Int. J. Comput. Appl. (0975–8887) **47**(1) (2012, June)
11. I. Paliy, Face detection using Haar-like features cascade and convolutional neural network, in *2008 International Conference on "Modern Problems of Radio Engineering, Telecommunications and Computer Science" (TCSET)* (Lviv-Slavsko, 2008), pp. 375–377
12. V. Mohan, R. Deepa, M. Deepa, S. Somannavar, M. Datta, A simplified Indian diabetes risk score for screening for undiagnosed diabetic subjects. J. Assoc. Phys. India **53**, 759–763 (2005)
13. D. Tyas Purwa Hapsari, C. Gusti Berliana, P. Winda, M. Arief Soeleman, Face detection using Haar cascade in difference illumination, in *2018 International Seminar on Application for Technology of Information and Communication* (Semarang, 2018), pp. 555–559. https://doi.org/10.1109/ISEMANTIC.2018.8549752
14. R.S. Chora, *Image Feature Extraction Techniques and Their Applications for CBIR and Biometric System*
15. M. Stojmenovic, Mobile cloud computing for biometric applications, in *2013 16th International Conference on Network-Based Information Systems* (Melbourne, Australia, 2012), pp. 654–659. https://doi.org/10.1109/NBiS.2012.147
16. G.M. Zafaruddin, H. S. Fadewar, Face Recognition: A Holistic Approach Review, 4799-6629-5
17. N. Pinto, J.J. Di Carlo, D.D. Cox, How far can you get with a modern face recognition test set using only simple features? in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2009)
18. A. Rastogi, A. Pal, B.S. Ryuh, Real-time teat detection using Haar cascade classifier in smart automatic milking system, in *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (Penang, 2017), pp. 74–79. https://doi.org/10.1109/ICCSCE.2017.8284383
19. Ramyashree, P.S. Venugopala, Proposal for enhancing face detection in group photos, in *2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)* (Mangalore, India, 2018), pp. 113–118. https://doi.org/10.1109/iCATccT44854.2018.9001937
20. D.K. Ulfa, D.H. Widyantoro, Implementation of Haar cascade classifier for motorcycle detection, in *2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)* (Phuket, 2017), pp. 39–44. https://doi.org/10.1109/CYBERNETICSCOM.2017.8311712

Deep Learning Photograph Caption Generator

Savitha Shetty, Sarika Hegde, Saritha Shetty, Deepthi Shetty, M. R. Sowmya, Reevan Miranda, Fedrick Sequeira, and Joyston Menezes

Abstract Generating caption for an image is challenging problem present in artificial intelligence. For a given photograph text, descriptions must be generated. So we need to have computer vision method for understanding the exact content present inside image and also natural language processing field to turn image understanding into proper words in right sequence. Methods of deep learning have successfully demonstrated the ability to solve this problem statement. The advantage of deep learning methods is that it is possible to define an end-to-end model for predicting the caption whenever picture is given, rather than needing a pipeline of specifically designed models or sophisticated data preparation. Our paper focuses on preparing the text and image data for training our deep learning model, we need to design and also train our deep learning caption generation model, and we need to evaluate the trained caption generation model, using it to caption entirely new photographs. Finally, we are hosting a webserver to which the user can upload images and fetch the generated captions.

Keywords Caption generation · Deep learning · End-to-end model · Webserver

S. Shetty (✉) · S. Hegde · R. Miranda · J. Menezes
Department of Computer Science and Engineering, NMAM Institute of Technology, Karkala, India
e-mail: shettysavi1@nitte.edu.in

S. Hegde
e-mail: sarika.hegde@nitte.edu.in

S. Shetty · F. Sequeira
Department of Computer Applications, NMAM Institute of Technology, Karkala, India
e-mail: shettysaritha1@nitte.edu.in

D. Shetty · M. R. Sowmya
Department of Computer Science and Engineering, NMIT, Bangalore, India
e-mail: deepthi.shetty@nmit.ac.in

M. R. Sowmya
e-mail: sowmya.mr@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_23

1 Introduction

Caption generation requires both the computer vision methods to understand image content and language model, from natural language processing field to turn image understanding into words in the right sequence. The aim of computer vision problems at an abstract level is, we are using image data to infer some information about the world. It is considered as multidisciplinary field, which is also a subfield of machine learning and artificial intelligence. It involves the use of some special methods and usage of general learning algorithms. If input dataset is provided, then CNN has the ability of humans to do the classification of images [1].

Generation of natural language is the process of transforming structured data into natural language. It is also used to generate long-form content for automating customer reports for various organizations and also to produce the custom contents for web or mobile application. Description and recognition of image are fundamental challenge of computer vision [2].

Deep learning focuses on state-of-the-art results on various problems related to the generation of captions. Various deep learning algorithms try to exploit the unknown structure in distributing the inputs to further discover the good representations. At multiple levels, the features which are learned at higher levels define features at lower levels. The Internet is composed of images and text. Indexing and searching text are relatively straightforward but algorithm which is used should know what is the content of the image in order to index and search images. The images and video content have remained opaque for the longest time, and it is properly described using meta descriptions which are provided by person who uploads them. In order to get the proper information from the image, computer must understand content of image and generate descriptions in human understandable format. The paper focuses on utilizing the deep learning architectures, deep neural networks, recurrent neural networks, convolutional neural networks for generating captions for various images. The convolutional neural network is used for the computer vision. The recurrent neural network is used for language modeling, and the deep neural network is used to find proper mathematical manipulations to turn input into output. Web technologies are used to present generated captions to the client.

2 Related Work

Raffaella et al. [3] proposed a system which classifies existing approaches based on, how they conceptualize the problem, that is models that cast description as either a generation problem or visual or a multimodal representation space retrieval problem. It provides detailed review of various existing models and also highlights their pros and cons. It also gives an overview of various benchmark image datasets used and various evaluation measures which are developed to evaluate the quality of the

machine-generated descriptions of images. Finally, it extrapolates future directions in the field of automatically generating proper descriptions of images.

Marc et al. [4] explained that in neural image caption system, and it is typically seen as a primary 'generating' component. It suggests that features of should be 'injected' to RNN. It is dominant view in case of the literature. RNN is used to encode all earlier generated wordings. RNN is used to encode various linguistic characteristics, and final representation is merged with features of image in next stage. After comparing two architectures, we came to know that late merging will outperform injection. RNN is seen as encoder than generator.

Marctanti et al. [5] explained that whenever a recurrent neural networking model is used for generating caption, information of the image is fed into neural network by including it in that RNN with a condition that our model by 'injecting' the image feature, layer following that RNN condition that particular language-related model by 'merging' all image feature. Both the options are attested in our literature, there are no systematic comparisons among two. The authors say it is not necessary to know whether performance first architecture will be used or not. This merge architecture has various advantages, since conditioning by merging will make RNN's hidden state vector will be shrunk by 4 times. Linguistic and visual features need not been coded together because that will result in very huge model generation.

Kelvin et al. [6] proposed that there is attention-based model which will learn automatically to describe contents of image. It also describes how a person is able to train this model in deterministic manner by the using backpropagation technique, stochastically by maximizing the lower bound. It describes that if we use visualization then model is able to fix its gaze by considering minute objects. But it generates proper words as in the output sequence. Validation is done by using three benchmark datasets, namely MSCOCO, Flickr30k, Flickr8k. Ramnath et al. [7] proposed auto-caption system which automatically suggests a caption for the photograph for the user.

3 Proposed System

We have used Flickr8k open-source dataset consisting 8092 photographs in JPEG format and five descriptions for each photograph to build the vocabulary and train the model. The dataset has a predefined training data with 6000 images, our development dataset with 1000 images, test dataset with 1000 images. Firstly, the features of images are extracted using VGG-16 model without its last classification layer. VGG-16 is a convolutional neural network model for recognizing and classifying images [5]. Since we only need the features and not classification, we simply remove the final layer in that model. Hence, each image will be represented as a 4096-element featured vector. Descriptions of images are cleaned, and a vocabulary of words is built. The description is encoded and will have to be split into many words. Model will provide with previous word, the image and next word as training. Then model is

Fig. 1 High-level design architecture

used to generate descriptions and words which are generated are concatenated and provided recursively for generating caption of the image.

The input text is given to word embedding layer followed by LSTM. The photograph feature is given into dense layer to obtain 256 element representation. Both input models produce a vector with 256 elements. In addition, both input models use 50 percent dropout regularization. This will reduce overfitting of training dataset, since configuration of the model learns faster.

Decoder model uses an addition operation to merge vectors from 2 input models. This will be fed to 256 neuron layer which is dense, and after that, next to final output dense layer which will make softmax prediction for our very next word in that sequence over the entire output vocabulary. In this way, we obtain captions for any given image. The web server is hosted using express module of node JS to handle client requests. When the user submits a form by uploading an image, a python child process is spawned which generates the caption. Upon receiving the caption from child process, the caption is rendered on webpage to the user.

Figure 1 gives an abstract view of our work. The web server listens to client requests. When the client uploads an image to be captioned, it first uses the modified VGG-16 model to extract image features. The image features and a start token are fed to the caption generation model. The caption generation model generates the next word of caption, and it is appended and recursively fed to the caption generation model until it generates end token. Once the complete caption is generated, it is rendered on a web page to the client. The design of caption generation model and modified VGG-16 model is discussed in low-level design architecture.

4 Methodology

Figure 2 describes each layer of modified VGG-16 model. Final layer in originally used VGG-16 model classified the image into categories. The final layer is removed, and hence, output of fifteenth layer is taken as the final output. The output obtained is a 4096 feature elements vector of the image, which will be fed to caption generation model.

Fig. 2 VGG-16 model without final classification layer

Fig. 3 Merge model of caption generator

Caption Generation Model

Figure 3 explains the merged model used for caption generation [6]. The model can be described in three parts:

Photograph Feature Extractor: The photograph feature extractor expects features of the input photograph to have a vector of 4096 elements. A dense layer processes these to produce a photograph representation of 256 elements.

Sequence Processor: Here word embedding layer is used to handle our text input and it is followed by LSTM, recurrent neural network layer. Sequence processor is expecting input sequences with predefined length of thirty-four words to be fed into our embedding layer which is being masked to ignore associated padded values. It is followed by an LSTM layer which has 256, memory units. Both of input models which are used to produce a vector 256 elements. Both our input models are using regularization with 50% dropout. This is generally used for reducing the over fitting of our training dataset, since our model configuration learns too fast.

Decoder: Our decoder model uses an addition operation to merge the vectors from input models. It is fed to dense layer having 256 neurons and to one final output dense layer which will make a softmax prediction for next used word using sequence over our output vocabulary. Figure 4 visualizes structure of network which better

Fig. 4 Caption generation model

helps to understand two streams presenting input. It provides low-level structure of the model discussed in Fig. 3, specifically the shapes of the layers.

Model definition and training: The steps for defining the model and training using the dataset are:

Prepare image data:

For each image in the dataset:

1. Load the VGG-16 model provided by Keras**:** Remove the sixteenth layers of the model to make penultimate layer the output layer.
   (a) Load the image
   (b) Convert the image pixels into numpy array
   (c) Reshape data for the model
   (d) Prepare the image for input to model
   (e) Extract features using the model
   (f) Store the features into a dictionary with the corresponding image ID.
2. Serialize the dictionary into a file.

Prepare Text Data

1. Read the descriptions file and convert it into dictionary form. Image ID will be the key, and value will be an array of five descriptions each.
2. Clean the descriptions.

   (a)Make all words lower case
   (b)Remove all punctuations
   (c)Remove single characters
   (d)Remove all words which have numbers in them.

3. Store dictionary of image ID and descriptions into a file.

Loading the Data

1. Load the list of training and development image identifiers provided by dataset.
2. Load the image features and cleaned descriptions for corresponding identifiers inlist.
3. Add 'start seq' and 'end seq' tokens to the descriptions.
4. Encode the description to create a mapping of each word to unique integer values. Tokenizer class provided by Keras does the encoding of loaded description.
5. Serialize the Tokenizer for future use.
6. Create input–output pairs for training and validation.

Define and Fit the Model

1. Define the model using Keras
2. Train the model using input–output pairs on Google Colab [7] for 20 epochs [8].
3. Our model learns very fast, and it quickly overfits our training dataset. Hence, skill of trained model should be always monitored on our holdout development dataset. We have to save whole model to a particular file whenever model skill on our development dataset is improving at end of each epoch.

Image Captioning

The trained model and Tokenizer obtained can be used to generate new captions for clients. The steps are listed below:

1. Setup a Web server using express module in Node.js which displays an image upload page to the clients
2. Upon image upload by the client, save the image using express file upload module and spawn a python child process
3. The python child process will:

   (a) Extract image features using modified VGG-16model
   (b) Seed the generation process
   (c) Iterate:

i. Integer encode input sequence using Tokenizer
ii. Pad the input sequence
iii. Predict next word using the model
iv. Convert probability to an integer
v. Map integer back to word using Tokenizer
vi. If word cannot be mapped then stop or else append it as input for the generation of next word
vii. Stop the process if end of sequence is predicted.

(d) Flush the generated caption onto standard output.

4. Render the generated caption on a webpage to the client using Embedded JavaScript Template [9].

5 Results

Figure 5 shows the features of images in dataset being extracted using modified VGG-16 model to train the caption generation model. Figure 6 shows the descriptions in dataset before processing. Figure 7 shows the cleaned descriptions used to train the caption generation model.

Figure 8 shows the data being loaded for training and validation. The Tokenizer is built here, and vocabulary size indicates the number of unique mappings. Input–output pairs are created from the data for training and validation. Figure 9 shows the caption generation model being trained on training data and validated against development data. Figure 10 shows webpage rendered by web server upon client request.

Figure 11 shows the web page rendered by web server to the client with final generated caption.

Fig. 5 Image feature extraction

Fig. 6 Descriptions in dataset

Fig. 7 Cleaned descriptions

Model evaluation is done by generating all descriptions for available photographs which present in test data and then evaluating each of these predictions using standard cost function. The predicted and actual descriptions are collected and then evaluated with corpus BLEU score which is used to summarize how close generated text is to text which is expected. Translated text is evaluated using bilingual evaluation understanding score. Generated description is compared against all reference descriptions for all photographs in the dataset. BLEU scores are calculated for 1, 2, 3, 4 cumulative *n*-grams.

The score can range from 0 to 1, where a score closer to 1 is always better and score close to 0 is always worse. The model scores fit perfectly within our expected range of skillful model on the problem.

Fig. 8 Loading the data

Fig. 9 Training on Google Colab

6 Conclusion and Future Work

Our paper on training a deep learning model using text and image, and using it to caption new photographs. A web server was developed to listen to client requests, where the clients can upload images on the webpage and get them captioned. The convoluted neural network used here is a small 16-layer VGG model. Larger models that might offer better performance can be explored, such as ResNet. Word vectors pre-trained on training dataset or large corpus of text such as articles and Wikipedia

Fig. 10 Image uploading page

Fig. 11 Generated caption

can be used rather than learning them as a part of fitting the model, to improve the performance. The webpage for the client can be made more user-friendly, and the model can be trained on a larger dataset.

References

1. K. Kafle, C. Kanan, Visual question answering: Datasets, algorithms, and future challenges. Comput. Vis. Image Underst. **163**, 3–20 (2017)
2. J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 2625–2634
3. Raffaella Bernardi Plank, et al., *Automatic Description Generation from Images: A Survey of Models, Datasets, and Evaluation Measures* (2016)
4. Marc Tanti Camilleri, Albert Gatt, P. Kenneth, *What is the Role of Recurrent Neural Networks (RNNs) in an Image Caption Generator?* (2017)
5. A. Marctanti, C. Kennethp, *Where to put the Image in an Image Caption Generator* (2018)
6. Kelvin Xu Bengio, et al., *Show, Attend and Tell: Neural Image Caption Generation with Visual Attention* (2015)
7. K. Ramnath, S. Baker, L. Vanderwende, M. El-Saban, S.N. Sinha, A. Kannan, ... A. Bergamo, Autocaption: automatic caption generation for personal photos. in *IEEE Winter Conference on Applications of Computer Vision* (IEEE, March 2014), pp. 1050–1057
8. K. Simonyan, A. Zisserman, *A Very Deep Convolutional Networks for Large-Scale Image Recognition* (2015)

Streaming of Multimedia Data Using SCTP from an Embedded Platform

E. S. Vani and Sankar Dasiga

Abstract With the Internet of things which is going to be happen in near future, the several of the functionality that currently being done by large computer systems are also going to become a requirement for embedded platforms. As we all know that nowadays, Internet is also being used for multimedia streaming. As such, the streaming of multimedia data is also envisaged to be a requirement for embedded platform. In our project, we have taken an embedded platform which is very cost effective and having an ARM processor and has got ARM A profile. With this platform, we are implementing an application using SCTP which we are attempting to transmit multimedia data both video and audio from an embedded platform, and we are comparing the performance of SCTP and TCP.

Keywords Streaming · ARM processor · SCTP · TCP

1 Introduction

For many of our day-to-day needs, embedded systems offer specific and cost-effective solutions and they have become an integral part of our lives. Nowadays, embedded technology is in prime, and the wealth of knowledge available in embedded technology is mind blowing. Also, nowadays for hands-on study, a number of development platforms are available along with tools, libraries, and peripheral devices.

Streaming is one of the data transfer technique [1]. Streaming means the end user can make use of the data when the data is on its way to the user. The user can view the data as and when he receives without waiting for the whole data to be received first. He

E. S. Vani (✉)
Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: Vani.es@nmit.ac.in

S. Dasiga
Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_24

can view it directly without downloading the whole thing [2]. Streaming is the real-time transmission of data. The data may be live or stored media. Compared to other data, multimedia data is typically large in size, and for its reasonable use it requires significant bandwidth. The multimedia data may be audio, video, slideshows, etc.

For inserted applications, the microcontroller having ARM processor has turn into a default processor. This assignment means to execute sight and sound spilling on ARM processor-based embedded platform.

The first objective of this work is to transfer multimedia data using TCP and also using SCTP protocols. Then, compare the performance of both. The second objective is to stream the multimedia data from an embedded platform using SCTP through multiple communication links simultaneously.

With IOT, the embedded systems are expected to be on the Internet and communicating with zero user intervention. These embedded systems have variants of 802.11 standards such as Wi-Fi, Ethernet, Bluetooth, and NFC [3].

The motivation for this project is to interconnect through multiple 802.11 when available, whereby stream the information such as multimedia with a protocol such as SCTP that support simultaneous communication through multiple links. These multi-channels should be selectable dynamically [4].

To establish connectivity through all, the available communication links to get maximum throughput [5].

2 System Specifications

2.1 Hardware Specifications

2.1.1 Raspberry Pi

With the intension of promoting the teaching of basic computer science, the Raspberry Pi foundation in the UK has developed the Raspberry Pi which is a series of credit card sized single board computer. The original Raspberry Pi includes an SRM1176JZF 700 MHz processor which is based on BroadcomBCM2835 system on chip. USB port is directly connected to SoC in model A and A+. Model B has two ports and model B+ has five-point hub.

2.2 Software Specifications

2.2.1 Linux

The Linux is one of the open-source operating system which is freely distributed, cross-platform based on Unix. It is similar to windows and OS X. The Linux kernel is

the core of the operating system. Linux has begun as a server OS "From wristwatches to supercomputers" is the popular description of Linux capability. It brings up a streamlined Linux environment. It allows the user to turn on their devices within a matter of seconds. The natural fit for Linux is the growth of cloud computing.

2.2.2 The NOOBS Installer

The Raspberry Pi package comes with only Raspberry Pi board and nothing else comes with it. It does not come with OS. We have to install an OS on it. The SD card loaded with the OS from the computer is inserted in the Pi which becomes the primary boot device. To ease the installation of OS on Pi, Raspberry Pi foundation has made a software called NOOBS (New Out of Box Service). The NOOBS installer can be downloaded from the official Web site. To install NOOBS on the SD card, a user needs to connect the SD card to the computer and just run the setup file. Then, the card is inserted to the Raspberry Pi board. When booting it first time, a user can select an OS from a list of OS to install.

2.2.3 Raspbian

Raspbian is a free operating system which is based on Debian for Raspberry Pi hardware. It provides more than a pure OS. Raspbian comes with 35000 packages. For easy installation on Raspberry Pi, Raspbian comes with pre-compiled software bundled nicely.

3 Communication Protocols

3.1 Transmission Control Protocol (TCP)

TCP is one of the transport layer protocol and is a core protocol of the Internet protocol suite. It provides reliable, ordered message delivery. File transfer, email, and other Internet applications rely on TCP. It provides its service at the transport layer which is between application layer and Internet protocol layer. Host-to-host connectivity is provided by TCP. The transmission details and handshaking are handled by TCP at the transport layer. TCP detects the problems like packet loss, unordered message delivery. TCP requests retransmission of data and also rearranged the unordered message. It passes the data to the receiving application only when it reassembled the sequence of data transmitted originally. This protocol is unsuitable for real-time applications like video over. TCP provides a guaranty that the message received is identical and also is in correct order. In order to guaranty the reliability of data transfer, it uses technique of positive acknowledgment.

Fig. 1 Multi-homing

3.2 Stream Control Transmission Protocol (SCTP)

SCTP is one of the important transport layer protocol [6]. It provides the features of both TCP and UDP, i.e., it is message-oriented like UDP and reliable and congestion control like TCP. The message can be fragmented into a number of data chunks by SCTP. In SCTP, the message has been sent by the sender in only one operation. The receiver receives the exact message in one operation.

The two important features of SCTP are:

1. Multi-homing
2. Multi-streaming

3.2.1 Multi-homing

The multi-homing means an endpoint can have more than one IP address associated with it. The connection in SCTP is called as an association [7]. Among multiple IP addresses, one is taken as primary and the remaining is taken as secondary or redundant IP address. In the event of failure of primary path, the data transfer takes the secondary path. So, it provides redundancy. Figure 1 represents multi-homing.

3.2.2 Multi-streaming

Another important feature of SCTP is multi-streaming [8]. It means a stream of messages can be sent within a single connection. It is the capability of SCTP in parallel transmission of several independent messages. Rather operating on bytes, it operates on messages. The packet loss in one stream does not affect the packet loss of another stream. And, this feature of SCTP avoids head of line blocking (HOL). The Fig. 2 represents multi-streaming.

Fig. 2 Multi-streaming

4 Design and Implementation

4.1 Block Diagram

Figure 3 shows the experimental setup used for the project. One of the first steps to be done in the setup is to assign an IP address for the Ethernet interface as relevant for the configuration by using ifconfig. For example, ifconfig etho 192.168.200.200.

The system to the left is connected to the Raspberry pi using an Ethernet cable. The systems to the right interact with the board using Wi-Fi or Ethernet. Once the setup is created, the Raspberry Pi receives the multimedia files transmitted from the system

Fig. 3 Block diagram of the system

connected to its left using SCTP and it will stream it to the other side through Wi-Fi. So in this experimental setup, any number of systems can be connected simultaneously, and the streaming happens simultaneously. With this, we have demonstrated an embedded platform serving as a kind of router and also as a streaming device. Because we are using an embedded platform with limited memory and with limited CPU capability, the idea here is streaming of multimedia files simultaneously is of limited size.

Fig. 4 Flowchart for multimedia streaming

4.2 Flow Chart

Figure 4 is a flowchart for multimedia streaming which gives an overview of the sequence of steps involved in the multimedia streaming implementation. The first step is the initialization of SCTP connections, then setup the buffer, and then Raspberry Pi is provided with the multimedia data. Fill up the buffer and then stream the data to different systems.

5 Experimental Results

The multimedia data has been streamed using TCP and also SCTP. This has been done in both desktop and also in embedded platform. Also streaming is done through both wired and wireless. The time taken for the transmission of data of different sizes has been noted down for all the above-mentioned cases. The readings and the corresponding graphs are as below (Tables 1 and 2).

The corresponding graph for Tables 1 and 2 is shown below.

Table 1 Readings of streaming in desktop computers

(a) Using TCP through wired connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	3.953	3.585	3.457
(b) Using TCP through wireless connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	14.753	14.409	13.489
(c) Using SCTP through wired connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	1.388	1.26	1.208
(d) Using SCTP through wireless connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	12.383	11.46	10.868

Table 2 Readings of streaming in an embedded platform

(a) Using TCP through wired connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	7.388	6.544	6.602
(b) Using TCP through wireless connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	19.666	19.955	18.41
(c) Using SCTP through wired connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	1.416	1.408	1.613
(d) Using SCTP through wireless connection			
No. of bytes (MB)	15.7	14.4	13.8
Time (s)	13.260	16.376	16.214

1. Streaming in desktop computers:

A) Using TCP through wired connection

B) Using TCP through wireless connection

C) Using SCTP through wired connection

D) Using SCTP through wireless connection

2. Streaming in an embedded platform:

A) Using TCP through wired connection

B) Using TCP through wireless connection

C) Using SCTP through wired connection

D) Using SCTP through wireless connection

6 Conclusion

In our project, we have compared the performance of both TCP and SCTP in the transmission of multimedia data. Also using an embedded platform, we have successfully transmitted multimedia data on multiple links simultaneously but with the restrictions for the reason that the available memory and computational power on the embedded platform is limited.

So, the conclusion is that the streaming of bigger and more complex multimedia data may not be suitable for embedded platform. Yet we envisaged that transmission of short and less resolution multimedia would be a necessity with IOT which is going to be happen. Embedded systems are becoming more powerful with computational power and available memory; they can also be a candidate for transferring large and more complex multimedia files.

References

1. K.-H. Kim, K.-M. Jeong, C.-H. Kang, S.-J. Seok, A transmission control SCTP for real-time multimedia streaming. Comput. Netw. 54, 1418–1425 (2010)
2. S. Ladha, P. Amer, Improving multiple file transfers using SCTP multistreaming. in *23rd IEEE and Communication Conference (IPCCC)*, pp. 13–22
3. R. Rajamani, S. Kumar, N. Gupta SCTP versus TCP: comparing the performance of transport. in *Protocols for Web Traffic* (Computer Sciences Department, University of Wisconsin-Madison)

4. A.L. Caro Jr., *SCTP: A New Internet Transport 1. Layer Protocol* (University of Delaware)
5. Y. Cao, C. Xu, J. Guan, H. Zhang, Receiver-driven SCTP-based multimedia streaming services in heterogeneous wireless networks
6. I.A. Rodriguez, Stream control transmission protocol—the design of a new reliable transport protocol for IP network
7. A. Jungmaier, E.P. Rathgeb, M. Schopp, M. Tüxen, SCTP—A multi-link end-to-end protocol for IP-based networks. Int. J. Electron. Commun. (AEU) **55**(1), 46−54 (2001)
8. S.J. Wee, J.G. Apostolopoulos, Secure video streaming for wireless networks. in *IEEE International Conference on Acoustics, Speech and Signal Processing*

A Fast Block-Based Technique to Detect Copy-Move Forgery in Digital Images

Vaneet Kour, Preeti Aggarwal, and Ravreet Kaur

Abstract Images have a use in almost each and every field of the world, from institutions to the court of law. This huge acceptance of the images in every field gave birth to numerous image editing tools and hence, various forgeries. Copy-move is a very popular forgery technique. Generally, keypoint-based and block-based methods are used for the copy-move forgery detection (CMFD). In this paper, a block-based approach for CMFD is implemented in which quantized discrete cosine transform (DCT) coefficients are used to detect forgery over an image. As, many forgers apply various post-processing like JPEG compression, blurring to make forged image traceless. Thus, these quantized DCT coefficients help in detecting the approximate match between the regions by setting user defined threshold parameter.

Keywords CMFD · Copy-move · Block-based · CoMoFoD

1 Introduction

Sophisticated tools for image manipulation have increased very much and are easily available. Editing is very easy to do because of such kind of tools but it is difficult to detect. Consider a scenario, an image needs to present at the court as an evidence. But before presenting, it has been edited which alters the original content of the image. Thus, this image can prove a criminal person as an innocent or an innocent as a criminal. Detecting forgery becomes a basic need of today. Therefore, forensic investigation is required for the images. Forensics can be mainly divided into two categories: analog forensics and digital forensics [1].

Analog forensics is all about physical evidences such as validating evidences and finding traces. It is also known by the name of classical forensics, e.g., forensic

V. Kour (✉) · P. Aggarwal · R. Kaur
University Institute of Engineering and Technology (UIET), Panjab University, Chandigarh, India

P. Aggarwal
e-mail: pree_agg@pu.ac.in

R. Kaur
e-mail: ravreetkaur@pu.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_25

investigation of the murder weapon which was found at the crime scene comes under analog forensics.

As the name suggests, digital forensics is the investigation or exploration of the digital evidences like images, videos, audio clips, etc. It can be grouped into two categories: computer forensics and multimedia forensics. Computer forensics is used when the evidence need to be extracted from the computer through which the malicious or criminal code of conduct took place. Audio, video, and image forensics come under the multimedia forensics. Assuring the authenticity and integrity of the multimedia is the principle goal of the multimedia forensics. Hence, image forensics is a part of multimedia forensics. Thus, image forgery refers to changing the content of the original image and producing transformed image as the original image with illegal intentions. There are three popular ways of image forgery: copy-move, image splicing, and image retouching [2].

Copy-move is copying the content from the image and pasting within the same image to hide the original information or in order to representing the fake information. Figure 1 is the example of copy-move forgery in which bird is copied from the original image and very cleverly pasted into the forged image. The images shown in Fig. 1 are taken from CoMoFoD_small dataset. Image splicing refers to copying the multiple region from two or more images and pasting into the fake image which is to be created. Image splicing is also known as cut-paste image forgery [3]. Image retouching is basically the enhancement of an image. It is done in order to draw the attention of the viewers to a particular object in an image or make the image more attractive [4]. Forgers apply various attacks and post-processing after the forgery to make the forged image more natural which makes detection of the forgery more challenging.

Fig. 1 **a** Original image, and **b** copy-move forgery

In this paper, a method to detect copy-move type of forgery is implemented. The rest of the paper is organized as follows. Section 2 is about the state of the art on CMFD techniques. The method implemented in this paper is described in Sect. 3. Section 4 shows the copy-move forgery detection results based on proposed method. Section 5 is the conclusion and discussion section.

2 Related Work

Copy-move forged images contain duplicate regions. Thus, the main idea behind detection is to detect duplicate region within the image. CMFD method involves four main steps: preprocessing, feature extraction, feature matching, and taking decision [3]. Preprocessing steps generally involve color conversions. There are mainly two approaches followed to detect the copy-move forgery: keypoint-based and block-based. If the detection is block-based, then dividing the image into blocks also comes under preprocessing step. Afterward features vectors are computed by extracting the features from the image or blocks. Feature matching is done in order to check the similarity within the image. The final step is taking the decision whether the image is authentic or not. In keypoint-based methods, the keypoints are extracted using a particular feature like SIFT, SURF, KAZE, LBP, etc., and then the matching is performed between the keypoints extracted to detect the copied region. In 2010, the method for detecting copy-move forgery based on SURF features was proposed by Xu et al. [5]. KAZE and SIFT features are used for keypoint extraction in the method proposed by Yang et al. [6]. Roy et al. [7] proposed a method for the CMFD in the presence of SGOs using the RLBP features. In 2019, Lin et al. [8] presented a novel approach for the CMFD. They had used combined features in which they extracted SIFT and Local Intensity Order Pattern (LIOP) at first. For the feature matching purpose, they used transitive matching.

Blocks-based approaches first divide the images into blocks and then compare these blocks with each other to detect the copy-move forgery. Block-based approaches using DCT are presented in various works [9–12]. Hu et al. [10] proposed an improved lexicographical sort algorithm for copy-move forgery detection. Authors first divided the image into 8×8 blocks and then applied discrete cosine transform (DCT). Then, they cluttered DCT coefficients according to their frequencies. At the end, they used eigenvalues for the block matching [10]. Sondos M. Fadl et al. [12] proposed a fast and efficient method to detect copy-move forgery. This method is a block-based method which firstly divide the image into fixed size overlapping blocks. Then, authors had applied DCT to each block. After that, they used fast K-means clustering technique. This technique makes clusters by putting the blocks into different classes.

In the maximum block-based approaches, the square shape is used for the blocks. However, Zunliu Zhang et al. [13] used circular blocks to extract DRHFM in their method which shows good results in detecting forgeries under geometrical transformations. Dixit et al. used statistical image properties. They used mean and variance

as features [14]. Fadl et al. [15] used frequency as a feature. The frequency is calculated using the Fourier transform over the polar system representation of the blocks of the image.

Keypoint-based methods are generally faster than the block-based methods. However, block-based approaches are known for producing good results. In this work, a block-based method using DCT is implemented which is described in the following section.

3 Proposed Methodology

The idea provided by Fridrich et al. proposed in [9] is applied with modification in this paper. The modification is done in order to reduce the false positives in the final output. The pairs of the matching blocks that create more false positives are discarded. When a region is copied and pasted within the same image, there is a correlation between the copied and pasted region. Due to the various post-processing on the forged image, these two regions, i.e., copied and the pasted, may not exactly match but approximately. Therefore, the method for the approximate match is implemented in the paper. The flowchart of the implemented algorithm is shown in Fig. 2.

The algorithm takes grayscale image for processing but if the image is RGB, then grayscale conversion for the image is done by using the following standard formula:

$$\text{Grayscale Image} = 0.299\text{R} + 0.587\text{G} + 0.114\text{B} \tag{1}$$

In Eq. (1), R, G, and B represent the red, green, and blue components, respectively.

Afterward, the image is divided into B $\times$ B overlapping blocks. For dividing the image into blocks, a sliding square window of size B is taken from the upper left corner of the image. The pixels coming under each window are considered as a single block. To make the blocks overlapping with each other, the window is shifted by 1 pixel toward the right and hence this procedure is followed till the last pixel of an image [9]. DCT coefficients are calculated and quantized for each block. The quantization steps depend upon the Q-factor which will be set by user. Q-factor is the quality factor which is similar to the JPEG compression's quality factor. Higher the value of Q-factor, finer will be the quantization and hence, much closer match between the blocks to be declared as similar. On the other hand, low value of Q-factor results in some false matches. After the quantization, a matrix 'A' is created for storing these quantized coefficients. These are stored in row-wise manner. Each row represents a single block of an image. It is to be noted that, here quantization is the key to find the approximate matching between the blocks. However, if there is a requirement to find only the exact matching between the blocks, then there is no need to do quantization and hence the step of doing quantization can be skipped.

As for now, the matrix 'A' is created in which each row represents a block. To find the correlation between the blocks, matching needs to be done. But before performing the matching, one more step is required which is lexicographical sorting.

Fig. 2 Flowchart of the implemented method

It will create order between the rows which makes the matching process simpler and less time consuming.

The next step is to look for the matched blocks. As the matrix A is sorted, therefore, the algorithm only looks for the two consecutive matching blocks. But remember, algorithm can result into more false matches because DCT coefficients are used in this detection method instead of exact pixel values. For that, all the matching pairs of the blocks required to be remembered for further detection process. Hence, algorithm creates a separate list for storing the positions of the matching blocks. For each matching pair, a shift vector is calculated by using the locations of the blocks. These locations are actually the coordinates of the upper left pixel of the block. Let (M_1, M_2) and (N_1, N_2) are the locations of the two matching blocks. In other words, (M_1, M_2) are the coordinates of upper left (first) pixel of first block and (N_1, N_2) are the coordinates of the upper left (first) pixel of the other block in the matching pair. Then shift vector is calculated as shown in Eq. (2):

$$S = (M_1 - N_1, M_2 - N_2) \tag{2}$$

However, S or $-S$ represents the same shift. Thus, normalization for the $-S$ can be done by multiplying it by -1. The count of the shift vectors is stored in the shift counter 'C'. Counter is incremented by 1 for each matching pair as shown in Eq. (3):

$$C(S_1, S_2) = C(S_1, S_2) + 1 \tag{3}$$

At the last, counter represents the frequencies of the shift vectors. For taking decision, threshold 'T' plays a role which is specified by user. Then the following condition shown in Eq. (4) is checked for all $r = 1, 2, 3, \ldots k$:

$$C(S^{(r)}) > T \tag{4}$$

$S^{(1)}, S^{(2)} \ldots . S^{(k)}$ are the normalized shift vectors calculated by the algorithm. If the condition represented in the Eq. (4) is met, then the matching blocks corresponding to that shift vectors are highlighted with the same color, hence detects copy-move forgery.

Threshold need to be chosen wisely; very small threshold can lead to high number of false positives whereas by setting very high threshold can result into missing of the some closely matched blocks.

4 Experimental Results

The above algorithm is implemented using MATLAB R2018a, RAM 4 GB, and Processor 2.50 GHz. The block size chosen is 16×16 for the testing of the algorithm. As the block size is quite large, the reason behind taking this block size is that there will be low chances of the false matches. This algorithm is tested on the several images of the CoMoFoD dataset. The results are shown in the Figs. 3, 4, and 5. These figures clearly show that algorithm correctly detects the copied regions which are highlighted in the resultant image. For example, Fig. 3b is the copy-move forged

Fig. 3 **a** Original image, **b** copy-move forged image, and **c** forgery detection result

Fig. 4 **a** Original image, **b** copy-move forged image, and **c** forgery detection result

Fig. 5 **a** Original image, **b** copy-move forged image, and **c** forgery detection result

image in which the poles are missing. Figure 3c is the detection results which shows that two regions are duplicate in this image, thus proving Fig. 3b copy-move forged.

The quality parameter and threshold chosen were 0.25 and 25, respectively. The criteria behind choosing these parameters are already explained in Sect. 3. In short, the parameters settings are done in order to avoid the false matches. Although, the algorithm is tested by taking the other threshold and quality parameters. Not much differences are noticed except the images containing the flat surfaces like wall, plain background, etc. Figure 6 represents the time taken by the algorithm of 10 random images to show the results.

The average time taken by the algorithm to run on a single image is 36 s. However, it can vary according to system specifications.

5 Conclusion and Discussion

CMFD is a quite challenging task because the algorithm needs to put a label on an image as forged or original. Therefore, in this research, a method based on DCT is

Fig. 6 Performance time

implemented. The foremost step of the algorithm is to divide the grayscale image into 16 × 16 blocks. DCT coefficients are calculated and quantized which are then stored in the matrix form. A lexicographical sorting is performed on the matrix before comparing the rows which represents the blocks of an image. After the matching, the copied blocks are highlighted in the output image. Although the proposed algorithm took very less time approximately 36 s to provide the result for one image, it can generate some false positives for the uniform areas for example flat surfaces like sky.

References

1. S. Sadeghi, S. Dadkhah, H.A. Jalab, G. Mazzola, D. Uliyan, State of the art in passive digital image forgery detection: copy-move image forgery. Pattern Anal. Appl. **21**(2), 291–306 (2018)
2. M.H. Alkawaz, G. Sulong, T. Saba, A. Rehman, Detection of copy-move image forgery based on discrete cosine transform. Neural Comput. Appl. **30**(1), 183–192 (2018)
3. L. Zheng, Y. Zhang, V.L.L. Thing, A survey on image tampering and its detection in real-world photos. J. Vis. Commun. Image Represent. **58**(December), 380–399 (2019)
4. M.A. Elaskily, H.A. Elnemr, M.M. Dessouky, O.S. Faragallah, Two stages object recognition based copy-move forgery detection algorithm. Multimed. Tools Appl. **78**(11), 15353–15373 (2019)
5. B. Xu, J. Wang, G. Liu, Y. Dai, Image copy-move forgery detection based on SURF in *Proceedings of 2010 2nd International Conference on Multimedia Information Networking and Security (MINES 2010)*, pp. 889–892, (2010)
6. F. Yang, J. Li, W. Lu, J. Weng, Copy-move forgery detection based on hybrid features. Eng. Appl. Artif. Intell. **59**(October), 73–83 (2017)
7. A. Roy, A. Konda, R.S. Chakraborty, Copy move forgery detection with similar but genuine objects. in *Proceedings of International Conference on Image Processing ICIP*, vol. 2017, (September 2018), pp. 4083–4087
8. C. Lin et al., Copy-move forgery detection using combined features and transitive matching. Multimed. Tools Appl. **78**(21), 30081–30096 (2019)
9. J. Fridrich, D. Soukal, and J. Lukáš, Detection of copy-move forgery in digital images. in *Proceedings of Digital Forensic Research Workshop* (Aug 2003)

10. J. Hu, H. Zhang, Q. Gao, H. Huang, An improved lexicographical sort algorithm of copy-move forgery detection. in *Proceedings of 2nd International Conference on Network Distribution and Computing ICNDC 2011* (2011), pp. 23–27
11. S.M. Fadl, N.A. Semary, A proposed accelerated image copy-move forgery detection (Dec 2014), pp. 253–257
12. A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, G. Bebis, H. Mathkour, Passive detection of image forgery using DCT and local binary pattern. Signal Image Video Process. **11**(1), 81–88 (2017)
13. J. Zhong, Y. Gan, J. Young, L. Huang, P. Lin, A new block-based method for copy move forgery detection under image geometric transforms. Multimed. Tools Appl. **76**(13), 14887–14903 (2017)
14. A. Roy, R. Dixit, R. Naskar, R.S. Chakraborty, Copy-move forgery detection exploiting statistical image features. Stud. Comput. Intell. **755**, 57–64 (2020)
15. S.M. Fadl, N.A. Semary, Robust Copy-Move forgery revealing in digital images using polar coordinate system. Neurocomputing **265**, 57–65 (2017)

Bottlenecks in Finite Impulse Response Filter Architectures on a Reconfigurable Platform

Kunjan D. Shinde and C. Vijaya

Abstract Digital filters are primary and important part in digital signal processing, often used in the scope of separation or restoration of signals. Due to inherent properties and advantages of finite impulse response (FIR) filter, it is more preferred over infinite impulse response (IIR) filter. In the present work, bottlenecks of various FIR filter architecture are discussed to estimate and understand the impact of these architectures on reconfigurable platform. The RTL implementation of various architectures is focused while justifying the critical path and the computation complexity involved. Use of symmetric coefficients (SC) is a key in reducing the area constrains of filter, and further use of data representation for coefficient is addressed with CSD and its impact is discussed. Impact of pipeline and parallel architectures is discussed, and their performance is evaluated for reconfigurable platform.

Keywords FIR filter · FPGA · Reconfigurable platform · Bottlenecks · FIR filter architectures · Direct from transposed form · Multiplier-less architecture · Canonical signed digit (CSD) · Pipeline architecture · Parallel architecture

1 Introduction

Gadgets around us are more sophisticated than earlier, powered on battery and process real-world signals while interacting with the environment. Digital signal processing (DSP) is important as it improves the overall quality of signal that we are interested to study and evaluate. Nowadays, it is equally important to develop an effective algorithm for signal processing and efficiently implement the same without degrading the performance.

Most of the signal processing algorithms have explicit parallelism or the parallelism can be explored and used. In such cases, the implementation platform is unable to support the parallelism offered by the algorithms. Reconfigurable platform using FPGA is often preferred for exploring and implementing parallelism found

K. D. Shinde (✉) · C. Vijaya
Department of E&CE, SDM College of Engineering & Technology, Dharwad, Karnataka, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_26

in algorithms, prototyping in ASIC, lower time to market, low NRE cost, functional verification of the design at early cycle, designer works at RTL-level while optimizing and modifying the designed RTL to meet the requirements. Change in targeted hardware has no impact on RTL described, as the tool will regenerate the necessary net list describing the design. Upgrading the signal processing algorithms on FPGA platform is easy as they are reconfigurable.

In [1], the author has given detailed study on VLSI digital signal processing algorithms and their impact on VLSI platform. Several methods are presented for exploring various architectures. In [2], the authors have explained in detail the study on digital filter design using various methods and representation of FIR filter using DF and TF is provided. In [3], the authors have designed FIR filter and improved its performance using CSD representation and optimized using common subexpression elimination methods, the number of logical operators (addition units) required is reduced while preserving the logical depth (number of adder stages) required to implement constant multiplication. In [4], the authors have applied pipeline architecture to transpose for FIR filter of two-stage/level and has obtained better results in their work. Simulations on ECG signal is carried out and analysis on critical path is made, filter designed is of order 16, 32, and 64, and coefficients are represented in 16 bits for reconfigurable platform. In [5], parallel FIR filter structures using iterated short convolution is proposed by the authors; this architecture can be used while exploring the parallel architectures. In [6], FIR filter architecture is presented with high performance for fixed and reconfigurable applications. Data flow graph for TF is used to identify the replication of computations made in successive iterations and has presented a structure with pipeline adder unit. Authors quote that their work consumes less ADP and EPS than DF structure. In [7], the authors have designed FIR filter with odd length and used fast FIR algorithm for parallel architecture. The authors have obtained area-efficient results for the implementation. In [8], the FIR filter is designed for block memories and obtained better performance in computational speed and power consumption while comparing with conventional architectures for ASIC and FPGA applications. In [9], the block FIR is designed using multiple constant multiplication for direct form FIR structure as block processing cannot be used for transposed form. Improvement in sampling rate and EDP is observed while consuming less area. In [10], authors have investigated pipeline and parallel architecture for FPGA Platform and observed significant reduction in area by using fast FIR filter architecture.

2 Architectures for FIR Filter

Digital filters are important class of LTI system. The properties of such filters can be represented by their impulse response or frequency response or using difference equation [1]. In the present work, difference equation is used as it provides the implementation details needed for VLSI platform. Whereas, the impulse response and frequency response gives the filter properties in time and frequency domain,

respectively. Digital filter can be represented as given in Eq. 1.

$$y(n) = \sum_{k=0}^{N-1} b_k x(n-k) - \sum_{k=1}^{M} a_k y(n-k) \tag{1}$$

If $a_k = 0$ for $1 \leq k \geq M$, the Eq. 1 gets reduced to N tap of FIR filter as given in Eq. 2 [1]. The computations in FIR filter are non-recursive as there is no feedback from the output.

$$y(n) = \sum_{k=0}^{N-1} b_k x(n-k) \tag{2}$$

Expanding the above equation

$$y(n) = b_0 x(n) + b_1 x(n-1) + b_2 x(n-2) + \cdots b_{N-1} x(n-N-1) \tag{3}$$

where $y(n)$ is output sample, N is filter order, b_k is filter coefficients, and $x(n)$ is input sample.

Equation 3 gives the expanded version of Eq. 2 where the output sample is computed by multiplying filter coefficients with past and present values of input samples. The impulse response of the filter has finite duration and the output depends on past and present values of inputs samples, due to which the filter is causal and stable. In applications where the phase or shape of the signal is to be addressed, FIR filters are preferred.

Filter coefficients can be symmetric (SC) or asymmetric based on the type of application the filter coefficients are selected [2]. FIR filter requires higher order to obtained desired response, the area and computation complexity has direct impact on the filter designed, use symmetric coefficients is a key and plays a vital role on reconfigurable platform. Implementation of coefficients can be of various forms and the same can be explored for its impact on the filter design [8].

In general, Nth order FIR filter needs $N-1$ delay elements, N multiplication, and $N-1$ addition units. Architectures for FIR filter gives details on the arrangement of these blocks and improvement in the filter performance by rearranging/replicating the blocks. Critical path in a FIR filter is the longest path for data to travel from input to output node. In general, the critical path for Nth order FIR filter is $\mathbf{T_M + (N-1)T_A}$ where the computation time of adder unit to be $\mathbf{T_A}$ and multiplication unit to be $\mathbf{T_M}$. The clock cycle is represented by $\mathbf{T_{clk}}$.

The following are the FIR filter architectures.

2.1 Conventional Architecture (Direct Form and Transposed Form)

The realization of Eq. 2 by developing signal flow diagram in a simple and straightforward way gives the direct form realization. Here the coefficients appear as gain units of the function, and number of delays are equal to filter order. Figure 1 gives the block diagram of generalized direct form structure of FIR filter.

Using Tellegen's theorem or Mason's gain formula, for a LTI system, its signal flow graph is inter-reciprocal with its transpose. It may be noted that in transposed form, the input sample gets multiplied with all the coefficients at a time [2]. This structure is later utilized to implement constant multiplication which results in new architecture (Fig. 2).

FIR filters are stable and can be practically realized to meet the requirements. The first bottleneck in the design process is the order of the filter as this increases the computation complexity and area constrain.

The critical path for tap 3 FIR filter represented in DF structure is $\mathbf{T_M + 2T_A}$, and the critical path for TF is $\mathbf{T_M + T_A}$ [1]. The critical path in transposed form is dependent on the effective realization of constant multiplication. Shift and add method is more commonly used along with its variants. Pipeline for TF can further increase the performance by reducing critical path [6]. As quoted in results section, the TF gives better performances as compared to DF, by the use of symmetric coefficients. Further optimization in computational complexity and arithmetic involved is seen.

Fig. 1 Block diagram of generalized direct form structure of FIR Filter

Fig. 2 Block diagram of generalized transposed form structure of FIR filter

Let $\mathbf{T_{M,\,max}}$ be the term for maximum delay through the constant multiplication and $\mathbf{T_{A,\,Tree}}$ be the maximum delay of adding the $N - 1$ products (or $(N - 1)/2$ partial products and one shift register delay in case of symmetric coefficients) of constant multiplication to generate the final output. This results in the critical path delay of $\mathbf{T_{M,\,max} + T_{A,\,Tree}}$.

Use of fast/parallel adders can resolve the speed metrics but may increase in area and power metrics of the designed system. Precision of the input sample and coefficients also act as a bottleneck for the architectures. As the precession increases, more effective is the signal interpreted while further increasing the design complexity.

2.2 Multiplier-Less Architecture

The transposed form structure of FIR filter results in constant multiplication as the input sample getting multiplied with all the coefficients. The coefficients are constants for a given design and it can be implemented using a series of shift and addition blocks.

Use of common subexpression elimination can reduce the logical depth and number of adders required [3, 9] to implement the constant multiplication, and shift operations can be hardwired. The resultant is new architecture known as multiplier-less FIR filter architecture. The generalized structure for multiplier-less architectures is given in Fig. 3.

Canonical signed digit (CSD) is a special manner for encoding a value in a signed-digit ternary representation, which itself is non-unique representation and allows one number to be represented in many ways. Generally, digital signal processing algorithms require a large number of multiplications. Depending on the number of bits used to represent the input samples and the filter coefficients, these multiplications can become a time-consuming process. CSD is an interesting solution in implementing efficient multipliers. Figure 4 gives the generalized CSD-based FIR filter architecture.

The multiplier-less architecture resembles TF structure. The delay $T_{\mathrm{M,\,max}}$ is replaced by $\mathbf{T_{CSD}}$, where $\mathbf{T_{CSD}}$ is the delay for implementing constant multiplication using shift and add method with data represented in CSD notation. The critical

Fig. 3 Block diagram of TF structure with constant multiplication

Fig. 4 Block diagram of TF structure with CSD-based constant multiplication

path for this architecture would be $\mathbf{T_{CSD} + T_A}$, tree and hence the bottleneck. The key to this is an efficient design of constant multiplication with CSD by eliminating common subexpressions and reducing its critical path.

2.3 Pipelined Architecture

Pipelining transformations lead to reduction in critical path [1]. In this architecture, pipeline registers are introduced in between paths as shown in Fig. 5, by adding pipeline register. The effective critical path is reduced with small increase in area due to pipeline register [4, 10]. Bottleneck of this architecture is that we cannot insert pipeline registers in every stage of FIR filter. Further by increasing the pipeline registers, the design may not reduce the critical path beyond some value. So it is important to identify the limit and effectively use pipeline registers.

A register is added at identified feed-forward cut; at every new interval, the data in this register will be updated and the previous data is forwarded within the network as shown in Fig. 5. In the present work, the pipeline is applied to the DF representation of FIR filter; the same concept can also be extended to transpose from FIR filter.

Fig. 5 Generalized structure of pipeline FIR filter

Lets consider a tap 3 FIR filter using DF structure with critical path of $\mathbf{T_M + 2T_A}$. With this structure, two-level pipeline is the limit, and the computation this cannot be reduced beyond $\mathbf{T_M + T_A}$ as bottleneck of this design, and hence, the minimum critical path achievable with pipelining.

$$\mathbf{T_{sample} \geq TM + TA} \tag{4}$$

where $\mathbf{T_{sample}}$ is the iteration period of pipeline architecture and $\mathbf{T_{sample} = T_{clk}}$.

2.4 Parallel Architecture

Parallel processing architectures compute multiple outputs at a given clock cycle; hence, the effective processing speed is increased by the level of parallelism [1].

The conventional architectures are single input single output (SISO) systems. In parallel architectures, multiple input multiple output (MIMO) is created by replicating the functional block as indicated in Fig. 6, where the inputs are $x(3k)$, $x(3k+1)$, and $x(3k+2)$ and generating the outputs as $y(3k)$, $y(3k+1)$, and $y(3k+2)$. Most of the reconfigurable platforms have limited IO pins; it is difficult to support MIMO architecture, and hence, serial to parallel and vice versa is needed. Availability of large area and IO pins may be the bottleneck for the design. Whereas the design becomes faster by L fold as given in Eq. 5 [1] (Fig. 7)

$$\mathbf{Tparallel} = \frac{1}{\mathbf{L}}\mathbf{T_{clk}} \geq \frac{1}{\mathbf{L}}\mathbf{CriticalPath} \tag{5}$$

Fig. 6 Parallel FIR filter for tap 3 with DF structure

where $\mathbf{T_{parallel}} \neq \boldsymbol{T_{clk}}$ as multiple outputs are computed in single clock cycle.

3 Requirements of FIR Filter: A Reconfigurable Platform Perspective

In the present work, band pass filter is designed for base band signals of frequency range 300–3400 Hz using hamming window. As it is a voice signal band, the sampling rate is set to 48,000 Hz. The filter coefficients are obtained from FILTER DESIGNER tool available with MATLAB; the above specifications are used for the design of tap 3 and tap 11 FIR filter.

The following are the filter parameters that may have a major impact on the FIR filter design on reconfigurable platform.

3.1 Filter Order

The notation "*N*" represents the order of the filter, which defines the length of the sequence to be sampled at a given interval of time. FIR filter must have higher-order to achieve the ideal response; higher-order means more number of taps which in turn means more data to be processed.

3.2 Precession of Data to Represent the Sample/Coefficients

Real-world signals and filter coefficients are often fractions, and hence, any standard data type can be used to represent it. In the present work, factions are scaled, and its equivalent binary is used. The impact of scaling the filter coefficients is given in Table 1.

The accuracy of scaled coefficients is improved by increasing the scaling factor at a cost of increase in computation complexity and area, while reducing the round-off error due to scaling. A scaling factor of 2^{32} is preferred as it produces the least round-off error, and hence, the precession is set to 32 bits.

3.3 Data Type to Represent the Signal

VLSI and DSP platforms provide various methods to represent a signal-like fixed point, floating point, CSD, Q-Notation, signed integer, binary, and more. Every representation has its own limitations and strength of use. Effective method can be chosen

Table 1 Impact of scaling factor of tap 11 FIR Filter coefficient using Hamming Window (fs = 48,000 Hz, fc1 = 300 Hz, fc2 = 3400 Hz)

Coefficient Value	*Scaling Factor*	*Scaled coefficient*	*Round off*	*Round off error*	*Retained coefficient after scaling & round off*
0.000017135000 58 Minimum Value among coefficients	2^8	4.3865E-3	4E-3	0.38E-3	0.000015672
	2^10	0.01754624	0.018	0.45E-3	0.000017578
	2^12	0.07018496	0.070	-0.18E-3	0.000017089
	2^14	0.28073984	0.2	-0.0807	0.000012207
			0.3	0.0192	0.00001831
	2^16	1.12295939	1	-0.1229	0.000015258
			2	0.8770	0.000030517
	2^18	4.49183759	4	-0.4918	0.000015258
			5	0.5081	0.000019073
	2^20	17.9673503	17	-0.9673	0.000016212
			18	0.0326	0.000017166
	2^24	287.4776	288	0.5224	0.000017166
	2^28	4599.64	4600	-0.36	0.000017136
	2^32	**73594.26**	**73594**	**0.26**	**0.000017134**
0.128956710 47709 Maximum Value among coefficients	2^8	33.0129	33	0.0129	0.12890625
	2^10	132.05	132	0.05	0.12890625
	2^12	528.20	528	0.20	0.12890625
	2^14	2112.82	2113	-0.18	0.12896728
	2^16	8451.30	8451	0.30	0.128952026
	2^18	33805.22	33805	0.22	0.128955841
	2^20	135220.91	135221	-0.09	0.128956794
	2^24	2163534.58	2163535	-0.42	0.128956735
	2^28	34616553.36	34616553	0.36	0.128956709
	2^32	**553864853.8**	**55386484**	**-0.2**	**0.128956710**

based on precession and dynamic range of data to be represented. In the present work, binary notation is used whereas the signs associated with coefficient can be resolved by using an add/sub block as a need to fulfill the functionality of FIR filter.

3.4 Sample Rate

It is dependent on the type of signal being processed and signal reconstruction requirements; sampling rate for the case study is set to 48,000 Hz.

3.5 Filter Coefficients

Selection of filter coefficients is a very important task as the filter coefficients determine how effectively the filter can segregate the wanted signal from noise. Equiripple, least square, windowing, and more methods are available for the selection of filter coefficients.

4 Results and Discussions

The present work gives detailed analysis on the various architectures for FIR filter and its bottlenecks are as quoted above. The work is simulated on Xilinx 9.1i ISE and synthesized on Virtex device xcv600-5-hq240. Specification of filter designed is mentioned in early section, and to generate FIR filter coefficients, FDA/FILTER_DESIGNER tool on MATLAB is used.

4.1 Simulation Results

Digital FIR filter processes discrete samples as input and produces filtered version input samples. To observe this behavior and verify its functionality, a common set of input $x(n)$ samples are applied for various architectures of FIR filter, and it is observed that the same set of results are obtained as shown in Figs. 8, 9, 10, and 11 for tap 3 and tap 11 FIR filter

Fig. 7 Simulation result of tap 3 FIR filter (DF, TF, CSD, and pipeline architecture)

Fig. 8 Simulation result of tap 3 FIR filter for parallel architecture

Fig. 9 Simulation result of tap 11 FIR filter (DF, TF, CSD, and pipeline architecture)

Fig. 10 Simulation result of tap 11 FIR filter for parallel architecture

Fig. 11 RTL schematic of tap 3 FIR filter with all coefficients in DF representation

4.2 RTL Implementation

It gives the implementation details on FPGA hardware selected after the process of synthesis. On observing the top-level RTL schematic, the following deductions are made for the FIR filter architectures present in the work.

The RTL implementation of DF uses all the three coefficients of FIR filter, 2 adders and 3 multipliers are used with shift registers interconnected as seen in Fig. 11. The implementation resembles the DF structure (Fig. 12).

Using the symmetric coefficients of filter and revising the RTL code, the implementation details of DF FIR filter is changed to 2 adders and 2 multiplies. There is a reduction in number of multipliers required to implement the structure and increase in computational speed is observed.

Figure 13 gives the RTL implementation of TF structure of FIR filter; it uses 2 adders and 3 multipliers due to the TF arrangement, the computation is fast as its critical path is $\mathbf{T_M + T_A}$.

Fig. 12 RTL Schematic of tap 3 FIR filter optimized with symmetric coefficients in DF representation

Fig. 13 RTL schematic of tap 3 FIR filter with all coefficients in TF representation

In TF RTL implementation, the input gets multiplied with the coefficients at a time, as $b_0 = b_2$, the required multiplier is reduced by one as shown in Fig. 14. It uses 1 adder and 2 multipliers.

With the use of CSD to implement the constant multiplications observed in TF architecture, the RTL is implemented by a series of shift and add units, thereby reducing the area consumed to a greater extent as seen in Fig. 15.

Fig. 14 RTL schematic of tap 3 FIR filter optimized with symmetric coefficients in TF representation

Fig. 15 RTL schematic of tap 3 FIR filter optimized with CSD for multiplier-less architecture

Fig. 16 RTL schematic of tap 3 FIR filter with pipeline architecture for DF representation

In the RTL implementation of FIR filter using pipelined architecture, we use feed forward cut and inserts pipeline registers to reduce the critical path of the FIR filter and to improve the performance in terms of speed. Pipelined register can be seen in the RTL implementation of Fig. 16.

The RTL implementation of FIR Filter using parallel architecture with block size of 3 is given in Fig. 17, it can be verified that there are 3 identical blocks of DF structure as we have used threefold the parallel structure generating 3 outputs. It uses adders and multipliers three times the DF structure producing with the highest computational speed. Fast FIR algorithms can be used to enhance efficiency and area utilization [7].

4.3 Comparative Analysis

In the present work, the comparison is made for tap 3 and tap11 FIR filters using direct form, transposed form, multiplier-less (CSD), pipeline and parallel architectures for our study and its performance is as shown in Tables 2 and 3, respectively.

Area metric in FPGA is number of LUT's used and intern the FPGA slices consumed. Delay metrice on FPGA is minimum period which is further subdivided

Fig. 17 RTL schematic of tap 3 FIR filter with parallel architecture for DF representation

Table 2 Comparative analysis of tap 3 FIR filter architectures

Parameters		**Tap 3 FIR Filter architectures synthesis on device xcv600-5-hq240 Virtex**						
FIR Architectures Vs ***FPGA Status***		DF	DF with SC	TF	TF with SC	Multiplier less TF (CSD)	Pipeline DF 2 stage	Parallel DF L= 3
Multiplier	**32*32**	3	1	3	2	0	3	9
	33*32		1			0		
Adder	**32 bit**		1					
	64 bit	2					1	3
	65 bit		1	1	1	6	1	3
Subtractor	**64 bit**					16		
	65 bit							
Register	**32 bit**	3	3				4	5
	64 bit	1		3	2			
	65 bit		1	3	1	5	2	
No. of slices		**1002**	**737**	**1061**	**716**	**666**	**1031**	**3009**
Slice Flip Flop		183	161	309	183	314	246	224
4 input LUT		1926	1324	1903	1259	1128	1894	5618
Min period		**26. 292**	**24.858**	**8.218**	**8.169**	**8.078**	**24.310**	**4.759**
Logic delay		15.752	15.698	6.7381	6.6891	6.738	15.110	1.859
Route delay		10.540	9.160	480	480	1.340	9.200	2.900
Max Freq(M Hz)		**38.034**	**40.229**	**121.684**	**122.414**	**123.79**	**41.135**	**210.128**
Min I/P arrival time(ns)		2.676	2.676	22.622	21.872	36.802	2.676	2.676
Max O/P after clk(ns)		7.511	7.511	7.511	7.511	7.511	7.511	34.702
IOB		97	98	98	98	98	98	228

Table 3 Comparative analysis of tap 11 FIR filter architectures

Parameters		Tap 11 FIR Filter architectures synthesis on device xcv600-5-hq240 Virtex						
FIR architectures versus FPGA status		DF	DF with SC	TF	TF with SC	Multiplier less TF (CSD)	Pipeline DF 2 stage	Parallel DF L= 10
Multiplier	32*32	11	1	11	6	0	10	110
	33*32		5					
Adder	32 bit		5					
	64 bit	10	5			32	6	100
	65 bit			10	10		4	
Subtractor	64 bit					27		
	65 bit					10		
Register	32 bit	11	11				12	21
	64 bit	1	1	11	11	6	1	
	65 bit			11	11	12	2	
No. of slices			**1992**	**3632**	**2341**	**1922**	**3376**	**33867**
Slice Flip Flop			416	1302	1247	1123	528	1427
4 input LUT			3730	6271	3818	3619	6495	65141
Min period			**35.436**	**9.087**	**8.737**	**8.637**	**31.554**	**4.759**
Logic delay			20.566	6.787	6.787	6.787	18.284	1.859
Route delay			14.870	2.300	1.950	1.850	13.270	2.900
Max Freq (MHz)			**28.220**	**110.05**	**114.46**	**115.78**	**31.692**	**210.13**
Min I/P arrival time (ns)			2.676	30.815	24.473	38.935	2.676	2.676
Max O/P after clk (ns)			7.511	7.511	7.511	7.511	7.511	53.357
IOB			97	98	98	98	97	673

into logic and route delay whereas the multipliers, adders, subtractors, and registers on FPGA gives the logic blocks utilized.

Tap 3 FIR filter, the impact of architectures, and use of symmetric coefficient are:

- DF with symmetric coefficients has slices reduced by 26.44% and speed is increased by 5.7% compared to DF without symmetric coefficients.
- TF with symmetric coefficients has slices reduced by 32.5% and speed is increased by 0.59% compared to TF without symmetric coefficients.
- Multiplier-less TF has slices reduced by 6.9% and speed is increased by 1.73% compared to TF with symmetric coefficients. It has slices reduced by 33.5% and speed is increased by 225.5% compared to DF without symmetric coefficients.
- Pipeline has slices increased by 2.8% and speed is increased by 8.15% compared to DF without symmetric coefficients.
- Parallel has slices increased by 320.2% and speed is increased by 71.6% compared to TF with symmetric coefficients. It has slices increased by 200.2% and speed is increased by 452.7% compared to DF without symmetric coefficients.

Tap 11 FIR filter, the impact of architectures, and use of symmetric coefficient is:

- DF with symmetric coefficients has slices reduced by 41.27% and speed is increased by 21.79% compared to DF without symmetric coefficients.
- TF with symmetric coefficients has slices reduced by 35.5% and speed is increased by 4% compared to TF without symmetric coefficients.
- Multiplier-less TF has slices reduced by 17.89% and speed is increased by 1.15% compared to TF with symmetric coefficients. It has slices reduced by 43.33% and speed is increased by 399.6% compared to DF without symmetric coefficients.
- Pipeline has slices reduced by 0.4% and speed is increased by 36.77% compared to DF without symmetric coefficients.
- Parallel has slices increased by 1346.6% and speed is increased by 83.58% compared to TF with symmetric coefficients. It has slices increased by 898.43% and speed is increased by 806.85% compared to DF without symmetric coefficients.

5 Conclusion

The present work provides a detailed study on FIR filter architectures like DF, TF, multiplier-less pipeline, and parallel architecture with their bottlenecks and possible solutions to overcome it for a reconfigurable platform. Comparative analysis and implementation details justify the performance of filter designed while providing the architectural impact. Inherent parallelism available in algorithms is explored using reconfigurable platform, and its performance is measured. Selection of effective architecture for an application can be made using Tables 2 and 3 and deductions made in comparative statements.

Acknowledgements We would like to thank Head of Research Center, Department of E&CE, SDMCET, Dharwad for proving us the tools, resources and platform to carry out the research.

References

1. K.K. Parhi, *VLSI Digital Signal Processing System-Design and Implementation* (Wiley Students Edition 2013). ISBN: 978-81-265-1098-6
2. C. Vijaya, U. Kumar, *Digital Signal Processing*. Elite Printers & Publishers. ISBN 1234567156882
3. A.P. Vinod et al., An improved common subexpression elimination method for reducing logic operator in FIR filter implementation without increasing logic depth. Elsevier, Integr. VLSI J. **43**, 124–135. https://doi.org/10.1016/j.vlsi.2009.07.001
4. P. Patali, S.T. Kassim, High throughput and energy efficient FIR filter architectures using retiming and two level pipelining, CoCoNet'19. Elsevier, Procedia Comput. Sci. **171**, 617–626 (2020). https://doi.org/10.1016/j.procs2020.04.067
5. C. Cheng, K.K. Parrhi, Hardware efficient fast parallel FIR filter structures based on iterated short convolution, in *IEEE Proceeding ISCAS 200*4, pp. 361–364. https://doi.org/10.1109/TCSI.2004.832784

6. B.K. Mohanty, P.K. Meher, A high—performance FIR filter architecture for fixed and reconfigurable applications. IEEE Trans. VLSI Syst. **24**(2), 444–452 (2016). https://doi.org/10.1109/TVLSI.2015.2412556
7. Y.-C. Tsao, K. Choi, Area-efficient VLSI ımplementation for parallel linear-phase FIR digital filters of odd length based on fast FIR algorithm. IEEE Trans. Circ. Syst. II: Express Briefs **59**(6), 371–375 (2012). https://doi.org/10.1109/TCSII.2012.2195062
8. M. Pristach, V. Dvorak, L. Fujcik, Enhanced architecture of FIR filter using block memories. Elsevier IFAC-2015 **48**(4), 306–311. https://doi.org/10.1016/j.ifacol.2015.07.052
9. P.K. Meher, Y. Pan, MCM-based implementation of block FIR filters for high-speed and low-power applications, in *IEEE Proceeding IC-VLSI&SoC* (2011), pp 118–121. ISSN: 2324-8432. https://doi.org/10.1109/VLSISoC.2011.6081653
10. M. Kadam, K. Sawarkar, S. Mande, Investigation of suitable DSP architecture for efficient FPGA implementation of FIR filter, in *IEEE Proceeding ICCI&CT 2017*. ISBN: 978-1-4799-5522-0. https://doi.org/10.1109/ICCICT.2015.7045672

Copy-Move Image Forgery Detection Using Discrete Cosine Transforms

R. P. Vandana and P. S. Venugopala

Abstract Digital image forgeries have led us to a situation where no digital image obtained using computing devices like PCs, smartphones and laptops are trusted to be authentic. As there are plenty of tools available free and open source, digital image forgery is no more a sophisticated job. Here we have come up with a simple but an effective technique for copy-move forgery detection in digital images as this type of forgeries is hard to be identified visually. The proposed technique uses block-based forgery detection technique. The features are extracted for each overlapping block, and then, it is compared with the features of other blocks to identify if there is forgery existing in the image. Few different images were tried, and difference in the accuracies has been observed. Accuracy varies depending on the copied image.

Keywords Copy-move forgery · Block based forgery detection · Discrete cosine transforms

1 Introduction

In this digital era, we all have realized that images speak more than text. "A picture is worth thousand words" is found to be true. As the field of computer science grows, introduction of new tools and techniques for manipulating digital images also grows at rapid rate.

Digital image forgery can be explained as falsely and fraudulently altering the digital images. This malicious activity started may be a century ago. But until the advancement in image forgery tools, these types of activities were hardly used by common man. The free and open-source availability of image manipulation tools nowadays have made the task of image forgery a funny and easy task to perform.

R. P. Vandana (✉) · P. S. Venugopala
Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, Karnataka, India

P. S. Venugopala
e-mail: venugopalaps@nitte.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_27

Broadly classified, there are four types of image forgeries possible. They are copy-move forgery, image splicing, resampling and retouching [1]. In copy-move forgery, a part of the image itself is cut and pasted on to another part. This may be used to conceal some information in the image or to add more objects onto the image. In image splicing, two or more images are used. The copied portions of different images are combined to form a new false image. Researchers have been concentrating on copy-move and image splicing than the other two methods as these techniques are widely used for illegal purposes by cyber criminals. Researchers who work on image forgery detection have always been trying to device an ideal technique for forgery detection, but that is the most challenging job as criminals are always trying to find new techniques for performing image forgery.

This paper proposes a simple but effective technique for detecting copy-move forgery in digital images. Copy-move forgery is the one where one portion of the image is copied and pasted onto another portion of same image. This method is said to be an active technique because you need no original image for detecting the forgery. As the copied portion is from the same image, it is hard to be identified as the properties like noise, color and texture do not change.

The detection techniques for this type of forgeries are classified into three types.

- Brute force technique
- Block-based technique
- Key-point-based technique.

Brute force technique follows exhaustive search auto-correlation techniques, whereas key point method is based on algorithm that captures image features and detects forgeries using the key points obtained [2–4]. Block-based method is the third method for forgery detection. In this method, we first divide the image into overlapping blocks. Features are extracted for these blocks and not for the image as a whole. Algorithms such as discrete cosine transforms, principal component analysis, signature value decomposition and discrete wavelet transform are used to perform forgery detection [5–8].

Any block-based technique for forgery detection in digital images must follow five steps given in Fig. 1 [9, 10].

2 Proposed Technique

The technique that we are proposing here is a block-based technique. In this technique, the given image is divided into blocks before the feature extraction is performed. The list of steps that we are following here is well described in Fig. 2.

- Preprocessing: This is the first step in the detection process. Here we convert the color image into grayscale image. When an image is converted to grayscale, its complexity decreases. Less number of values are enough to define a pixel. It is a reduction process.

Fig. 1 Steps to follow in block-based digital image forgery detection

- Segmentation: The grayscale image obtained in the first step is divided into overlapping blocks of size 16 × 16 or 32 × 32 pixels depending upon the need. The further processing is done on individual blocks and not on the image as a whole.
- Discrete cosine transform (DCT): DCT usually expresses a finite sequence of data points in terms of a sum of cosine functions. DCT is applied to each and every block obtained in the segmentation process. It helps to separate the image into spectral sub-bands.
- Zigzag scan: Zigzag scan is applied onto the DCT values obtained in the previous step. Zigzag scan is preferred over raster scanning because zigzag scanning separates the DCT coefficients such that the ones with higher energy appear first and lower energy after that. Here we need not consider the lower energy regions. We can just neglect them. It is enough if higher-energy regions are considered. The vector hence obtained is converted to quantization matrix.
- Lexicographical sorting: The vector of DCT values obtained in the above step is lexicographically sorted. This type of sorting can be seen in dictionaries where words are sorted based on the letters constituting them.
- Euclidean method: This method is used for finding out if there are similarities between the blocks that are sorted lexicographically. This step is very crucial as it identifies the actual forgery in the image. If the similarity indices between two blocks are beyond the threshold value, then such blocks are said to be similar. A threshold value for distance and similarity must be decided by the designer before the coding is done. The threshold values have high impact on the accuracy of

Fig. 2 Steps followed in the proposed technique

detection. The direction of the vectors which are found similar using Euclidean are also noted. As the last step, greatest direction value is marked and all the vectors in this direction are marked.

Presentation of the detection results is another challenge. A prediction mask can be the output of detection. This is an image containing only white and black pixels. The area with white pixels in prediction mask show the forged region. Fine-tuning the borders of the forged region is a biggest challenge in any type of detection. The accuracy also depends on this.

3 Results

The aim of this kind of study is identifying the forged region in the copied image as accurately as possible. This can be achieved by identifying the duplicated regions where the similarity index between them is less than a certain threshold and the duplicated regions are non-overlapping. Therefore, for forgery detection, two conditions are imposed over the duplicated block detection procedure: (1) The blocks should be non-intersecting and non-overlapping, and (2) the similarity index does not exceed a threshold. To meet the first requirement, the shift distance criterion is used. For this, let us consider that (x_i, y_i) and (x_j, y_j) are the top-left corner coordinates of the two blocks that are represented by the feature vectors fv_i and fv_j, then

$$\forall\sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{1}$$

If the two feature vectors satisfy Eq. (1), then only they are considered for similarity index calculation in the next step where Euclidean distance method is used as given in Eq. (2). Here d_t is the minimum threshold distance.

$$d\left(fv_i, fv_j\right) = \sqrt{\sum_{k=1}^{10}\left(fv_{ik} - fv_{kj}\right)^2} < d_t \tag{2}$$

3.1 Performance Evaluation

Accuracy of the detection is calculated by comparing the real and prediction masks. Confusion matrix method is a well-known method for visualization of performance of an algorithm. Here we make use of same method for calculating the following performance evaluation factors.

$$\text{Recall} = \text{True Positive} \,/(\text{True Positive} + \text{False Negative})$$

$$\text{Precision} = \text{True Positive (True Positive} + \text{False Positive)}$$

$$\text{Accuracy} = (2 * \text{Precision} * \text{Recall})(\text{Precision} + \text{Recall})$$

Precision denotes the probability which shows that a detected forgery is indeed a forgery, while recall denotes the probability where actually a forged image is detected [11]. Precision and recall are used to show the accuracy of a detection technique at image level and can be computed using the formulae given above. Images shown in Figs. 3 and 4 along with their real masks showing the forged regions are tested. The results obtained are individually shown in Figs. 5, 6, 7, 8, 9 and 10.

Observing Figs. 5, 6, 7, 8, 9 and 10, we can understand that there are differences in the real mask and prediction mask. The real mask shows which part of the image has been copied and pasted onto other part of the image. Prediction mask is prepared by the code that is detecting the forgery. Greater the similarity between the prediction and real masks, greater is the accuracy.

By looking at Table 1, we can make out that the value of accuracy can vary to a great extent when different forged images are used for testing. Figure 11 shows a chart that compares the different results. We can observe great variation in accuracy in the chart. Obtaining a good accuracy for any sort of image is a biggest challenge for researchers in this area. The chart given below compares the recall, precision and accuracy for the different tested images.

Fig. 3 **a** Copied Image1 and corresponding real mask, **b** Copied Image2 and corresponding real mask and **c** Copied Image and corresponding real mask

Fig. 4 **a** Copied Image4 and corresponding real mask, **b** Copied Image5 and corresponding real mask and **c** Copied Image6 and corresponding real mask

Fig. 5 Results obtained for Image1 producing an accuracy 93.99

Fig. 6 Result obtained for Image2 producing accuracy of 74.78

Fig. 7 Result obtained for Image3 producing accuracy of 51.69

4 Conclusion

Copy-move forgery in the field of digital images has become one of the most troubling activities. It is mainly used for illegal purposes which is creating a lot of problems related to cyber-crime. The researchers are taking this topic for their study as there is always need for a better technique as criminals keep coming with new methods for performing forging activities. The technique proposed here works well in identifying

Fig. 8 Result obtained for Image4 producing accuracy of 91.88

Fig. 9 Result obtained for Image5 producing accuracy of 45.87

the copied region in the image. But accuracy of the technique varies with the input images. There is scope for further studies in the same area for getting better and stable accuracy for all type of forged images. As we have artificial intelligence applied to all fields of image processing, the same can be tried with forgery detection also. Deep learning techniques are well suited for solving such kind of problems [12]. Convolution networks are used in the paper [12], and good results are obtained. When a neural network is trained for set of images, it can identify the forgery in the new input image. There are plenty of datasets available online for training and testing such models. So application of deep learning techniques for forgery detection can

Fig.10 Result obtained for Image6 producing accuracy of 60.39

Table 1 Results for different images

Image	Recall	Precision	Accuracy
Image1	0.0905	0.9954	0.9399
Image2	0.6046	0.9801	0.7471
Image3	0.3941	0.7510	0.5169
Image4	0.9934	0.8546	0.9188
Image5	0.2976	0.9990	0.4587
Image6	0.7602	0.5009	0.6039

Fig. 11 Comparing the results obtained for various images

be a very good area for research. A technique alone cannot achieve the best result. Instead, combination of available techniques must be tried to obtain the best results.

References

1. K.B. Meena, V. Tyagi, *Image Forgery Detection: Survey and Future Direction* (2019)
2. N.K. Gill, R. Garg, A. Doegar, *A Review Paper on Digital Image Forgery Detection Techniques* (2017)
3. Y. Li, J. Zhou, *Fast and Effective Image Copy-Move Forgery Detection via Hierarchical Feature Point Matching* (2018)
4. H. Chen, X. Yang, Y. Lyu, *Copy-Move Forgery Detection Based on Key Point Clustering and Similar Neighborhood Search Algorithm* (2020)
5. R. Dixit, R. Naskar, *Review, Analysis and Parameterization of Techniques for Copy-Move Forgery Detection in Digital Images* (2017)
6. S. Alagu, K. Bhoopathy Bagan, *Copy-Move and Splicing Image Forgery Detection Using DCT and Local Binary Pattern* (2019)
7. B. Soni, D. Biswas, *Image Forensic Using Block-Based Copy-Move Forgery Detection* (2018)
8. M.A. Elaskily et al., *Comparative Study of Copy-Move Forgery Detection Techniques* (2017)
9. A. Kashyap, R.S. Parmar, M. Agarwal, H. Gupta, *An Evaluation of Digital Image Forgery Detection Approaches* (2017)
10. Y. Sun, R. Ni, Y. Zhao, *Nonoverlapping Blocks Based Copy-Move Forgery Detection* (2018)
11. V. Christlein, C. Riess, J. Jordan, C. Riess, E. Angelopoulou, *An Evaluation of Popular Copy-Move Forgery Detection Approaches* (2012)
12. Y. Liu, Q. Guan, X. Zhao, *Copy-Move Forgery Detection Based on Convolutional Kernel Network*

Sentiment Classification and Data Analysis

A Detail Analysis and Implementation of Haar Cascade Classifier

Gaurav Ghosh and K. S. Swarnalatha

Abstract Haar cascade is a one of the popular machine learning algorithm used for object detection. The Haar algorithm identifies objects in image as well as video. The Haar algorithm was initially used to identify the body parts; later, it was used to for identifying any kind of object. The Haar identify the objects based on the features provided. The Haar algorithm is divided into four-part/stages: Haar feature selection, creating integral images, AdaBoost training, and cascading classifiers; the detailed analysis and implementation of all the four stages along with statistical data have been carried out in this paper, and also the process/step to be followed to create the Haar cascade file has been discussed in this paper. We can conclude that the Haar algorithm is more effective to recognize the object based on the features.

Keywords Haar · Cascading · Features · Images

1 Introduction and Methodology

(i) Haar Feature and Feature selection: CNN can be used for identifying and selecting Haar features, consider an image in Fig. 1, and it is a color image and every color image is made up of three colors that are red, green, and blue. To identify similar kind of object, for example, car, you have to use the following image as a positive sample of your dataset, but having color images for training might not be a very wise choice, as:

- It increases the size of the overall sample/training set.
- It provides unwanted distractions for the model like: saturation, color value, and hue which would decrease the efficiency of the classifier [1].

So, to remove these problems, you must convert your image to gray scale image. Gray scale is the image which has its pixel value ranges from 0 to 255 ([0, 255]). This would not only reduce the size of every image but also remove the unwanted factors.

G. Ghosh · K. S. Swarnalatha (✉)
Department of Information Science and Engineering, NMIT, Bangalore 560064, India
e-mail: swarnalatha.ks@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_28

Fig. 1 Sample color image

Fig. 2 Gray scale image

After converting the image to a gray scale image, it would look like as mentioned in Fig. 2

As humans can see the image like this, but computers see it in the form of a matrix of numerous pixels. As every image is made of some pixels, the computer reads the value of those pixels to understand the image. As this is a gray scale image, the pixel matrix would have multiple values between 0 and 255 [9]. It is impossible to show the pixel matrix of the whole image (as it will become unreadable), and therefore, Fig. 3 is an image of a sample pixel matrix of a part of an image.

In actual implementation, these matrices are enormously large, that is they can go to a size like 1900 × 2500 and still that is a small image. In the above image, you can see that every value inside the image is between 0 and 255, showing that it is a gray scale image. In the pixel matrix, we would even be able to see distinguish edges in it. Need to find the pixel values where you can see the contrast of two values by seeing the pixel values.

In Fig. 4 image within the two orange boxes, the two-pixel values have completely opposite values, that is one is dark and one is white. We can represent these boxes as separate colors like as shown in Fig. 5. The below image, i.e., Fig. 5, is called a Haar feature. Just like the above image, many other Haar features are created by different type of pixel values arranged in different ways. Some of them are shown in Fig. 6 [1].

Fig. 3 Sample pixel matrix

255	0	137	137	137	137	0
0	128	255	128	137	255	137
128	0	0	64	128	64	64
128	128	0	255	137	255	0
0	255	128	137	137	137	0
128	137	137	137	0	255	64
255	128	128	128	128	64	64

Fig. 4 Contrast of image pixels

255	0	137	137	137	137	0
0	128	255	128	137	255	137
128	0	0	64	128	64	64
128	128	0	255	137	255	0
0	255	128	137	137	137	0
128	137	137	137	0	255	64
255	128	128	128	128	64	64

Fig. 5 Haar features

Fig. 6 Haar features with pixel values

Fig. 7 Gray scale image

The Haar features are created internally, using inbuilt formula, in real-world scenarios; the grayscale image values are generally limited to a range of 0 and 1 [5]. Due to ease of calculation, the calculated values do not become very huge. This can be done by taking a normal grayscale image and normalizing it, so the values range between 0 and 1 as shown in Fig. 7.

We can also invert these values, that is 0 would mean white and 1 would mean black as shown in Fig. 8. In the future examples, we are going to use the inverted values, and the reason for using this is out of the scope, so we are not going to focus that.

If we put these values into the image pixel matrix, then our pixel matrix might look like as shown in Fig. 9.

To aim to detect any possible edges present, we consider any pixel with values less than 0.5 to be light or to be within the white group and any pixel values greater than 0.5 to be dark or within the black group.

According to Viola–Jones algorithm, the formula for detecting a Haar feature is:

$$\Delta = \text{dark} - \text{white} = \frac{1}{n}\sum_{\text{dark}}^{n} I(X) - \frac{1}{n}\sum_{\text{white}}^{n} I(X) \tag{1}$$

Fig. 8 Inverted image

Fig. 9 Image pixel matrix

0	0	1	1
0	0	1	1
0	0	1	1
0	0	1	1

0.1	0.2	[illegible]	0.8
0.2	0.3	0.8	[illegible]
0.2	0.1	[illegible]	0.8
0.2	0.1	0.8	0.9

The above formula says that we have to subtract the average of all the white pixels from the average of all the dark pixels. Here, $\mathbf{I}(\mathbf{x})$ is the intensity of the dark and the white pixels which are summed upon and divided by the total number of those pixels (or simply the average).

In the above pixel matrix, you can see the left one which describes an ideal case and the right one is depicted as a real-world scenario case. In the ideal case, the value of delta is always 1. According to the Viola–Jones algorithm, the real case scenario must be as close as possible to the ideal case to detect a Haar feature

Calculate the values of the right pixel matrix showing a real scenario:

$$\Delta = 0.74 - 0.18 = 0.56 \tag{2}$$

since the delta value is greater than 0.5, you can consider it as a Haar feature although not as a strong oneif this is to be used on a sample gray scale image as shown in Fig. 10.

The algorithm can find the features as shown in Fig. 11

(ii) Create Integral Images: consider Fig. 10 image. Run the algorithm; it would search for features by iterating through every pixel by using a particular sized block. You can visualize it as shown in Fig. 12.

Fig. 10 Gray scale image with haar feature

Fig. 11 Features finding

Fig. 12 Integral image

The image is being scanned by a block of some size. The blocks have been drawn very far away from each other, but in reality, the blocks are just a few pixels apart and in some cases, a new block is drawn by adjusting the box by just one pixel. Inside every single box, we can find thousands of pixels to be summed up and computed to compute a specific region for several number of times, which is very time expensive and this operation is executed in Big O of n^2 [O(n^2)]. To address this problem, we use integral images concept which is far more efficient than the one discussed. It provides a constant time complexity of O(1). These integral images are created as mentioned in Fig. 13, below you can see a pixel box for normal image and beside it another pixel box for the integral image.

Structure wise it looks the same but, if you have a closer look, you would see that the values are different from the original image and if you try a bit, you would be able to understand that every pixel value in the integral box is the sum of all the values to its left and all the values above it. So, let us see the following example shown in Fig. 14.

In Fig. 14 you can see that $1.2 = 0.1 + 0.1 + 0.2 + 0.3 + 0.1 + 0.4$ that is all the pixels to the left of 1.2 and above 1.2.

Fig. 13 Pixel box for normal image and integral image

Fig. 14 Transformed pixel box

(iii) AdaBoost

AdaBoost is one of the ensembles learning method (also known as "meta-learning") which was initially created to increase the efficiency of binary classifiers. It uses an iterative approach to influence the output of the weak learners, and by correcting the errors in the previous weak learners in our case study, we are using decision trees as the weak learners in the ensemble method. Decision trees are the ones which do not have a determined depth of its own, but if you restrict its depth only to 1, then every such tree would be having only 1 root and 2 leaves which are called stumps. AdaBoost technique has only decision trees in the form of stumps or better said as an ensemble of stumps. Since stumps have only one root node, so it can have only one variable to decide on and cannot have the advantage of having different weightage of different variables. Due to this reason, stumps are generally called weak learners.

Now, you might think that what would be the reason behind using something which is considered weak when we want to build something very strong and accurate.

So, to understand this, you must know that when you use Random Forest in an algorithm, it consists of a variety of trees which has the same weightage as the others

Fig. 15 Stumps in Adaboost

Chest Pain	Blocked Arteries	Patient Weight	Heart Disease	Sample Weight
Yes	Yes	205	Yes	1/8
No	Yes	180	Yes	1/8
Yes	No	210	Yes	1/8
Yes	Yes	167	Yes	1/8
No	Yes	156	No	1/8
No	Yes	125	No	1/8
Yes	No	168	No	1/8
Yes	Yes	172	No	1/8

but when you use stumps in AdaBoost you will note that each stump has a different weightage than the other depending upon the number of correct predictions. Learning from the mistakes of one stump, the other stump also influences its predictions accordingly and ultimately the accuracy level rises. The weightage of a stump is called "Say," due to which the order of these stumps also become important.

Therefore, the three ideas behind AdaBoost are:

- AdaBoost combines a lot of weak learners (mostly stumps) to make classifications.
- Some stumps have more weightage to their say in the classification, than others.
- Each stump is made by taking the previous stumps mistakes into account.

Let us take an example on how stumps are used in AdaBoost to make predictions. Let us assume you are given the following database as shown in Fig. 15.

Here you can see that the data has three columns and one output column. Make sure that the column "Sample Weight" is not a data column, but it is a column assigned because initially.

To give every row with same weightage

$$\text{That is: } \frac{1}{\text{total number of samples}} = \frac{1}{8} \tag{3}$$

To create the stumps for each of the column, since the weight of every row is the same initially, so, at this stage, it does not matter which stump is created first.

Here you can see that three stumps are created due to three different columns present in the dataset as shown Fig. 16. Now, the Gini index of each stump is calculated to determine which one does a good job in classification. After calculating the Gini index,

Fig. 16 After creating stumps

Chest Pain—0.47
Blocked Arteries—0.5
Weight—0.2.

The stump with column weight does the best job in classification as its Gini index is the least, and thus, this will be the first stump for our classification. Need to calculate the amount of say which stump will be having. To calculate the say, a formula is specified which is:

$$\text{Amount of Say} = \frac{1}{2}\log\left(\frac{1 - \text{Total Error}}{\text{Total Error}}\right) \tag{4}$$

Before calculating the amount of say of the stump, you must see the graph between amount of say and the error on the stump as shown in Fig. 17.

In the above graph, the say of a stump is high when it has a low error in it (that is close to 0) and the say of the stump will be low if the error is found to be high. Calculate the say of our stump. To calculate the say, you first need to calculate the error of the stump. Our stump incorrectly classifies one example out of total eight examples, and thus, the error becomes 1/8. Now let us put this error value in the formula for the amount of say. On calculating the values, you will find,

$$\text{Amount of say} = \frac{1}{2}\log(7) = 0.97 \tag{5}$$

By looking at the value, you can say that the stump does a pretty good job classifying and this value sits very near to 1 in the above graph. Now you need to update the weights of the rows, so that the next stump created learns from the mistake made

Fig. 17 Error on stumps

by this stump and predicts a better output. To update the weights, you need to make sure that the weight of the incorrectly classified row is increased, so that the next stump gives more importance to it.

To increase the sample weight, the formula is as mentioned below i.e., Eq. 6:

$$\text{New sample weight} = \text{sample weight} \times e^{\text{Amout of say}} \tag{6}$$

By the amount of say value in the formula with the previous sample weight, we get the new increased sample weight. Thus, on putting 0.97 in the amount of say and 1/8 in sample weight, we get new sample weight = 0.33 which is larger than 1/8 (0.125). Now let us decrease the weights of other samples, and for doing, so we have a very similar formula, i.e., Eq. 7:

$$\text{New sample weight} = \text{sample weight} \times e^{-\text{Amount of say}} \tag{7}$$

The only difference between this formula and the previous one is a negative sign to the power of e. This would decrease the value exponentially as we increase the amount of say. On putting the amount of say as 0.97 and sample weight as 1/8 as previously used, you get the new sample weight = 0.05 which is less than 1/8 (0.125). Now your updated table with updated weights looks as in Fig. 18 [6].

The total sum of the new weight $\neq 1$, whereas previously we had the total sum of previous sample weights = 1. So, to make the sum of new sample weights = 1, you need to normalize the value by dividing each value by the current sum which is equal to 0.68. After dividing each value, you can see the sum = 1. Now we can

Fig. 18 Updated weights

Chest Pain	Blocked Arteries	Patient Weight	Heart Disease	Sample Weight	New Weight
Yes	Yes	205	Yes	1/8	0.05
No	Yes	180	Yes	1/8	0.05
Yes	No	210	Yes	1/8	0.05
Yes	Yes	167	Yes	1/8	0.33
No	Yes	156	No	1/8	0.05
No	Yes	125	No	1/8	0.05
Yes	No	168	No	1/8	0.05
Yes	Yes	172	No	1/8	0.05

replace the previous sample weights with the new one which changes the table as follows in Fig. 19.

As you started the first iteration by choosing any of the columns previously, you cannot do the same now because the weight of each row is not the same now. There are few different ways to start, but here you would be learning the method of making a new collection of samples that contains duplicate copies of the samples with the largest sample weights.

Fig. 19 Replaced sample weights

Chest Pain	Blocked Arteries	Patient Weight	Heart Disease	Sample Weight
Yes	Yes	205	Yes	0.07
No	Yes	180	Yes	0.07
Yes	No	210	Yes	0.07
Yes	Yes	167	Yes	0.49
No	Yes	156	No	0.07
No	Yes	125	No	0.07
Yes	No	168	No	0.07
Yes	Yes	172	No	0.07

So we start by making a new, but empty, dataset that is the same size as the original one. Then, we pick a random number between 0 and 1, and we see where the number falls when we use the sample weights like a distribution. If the chosen number is between 0 and 0.07, then we put the first row into the new collection of samples. If the number is between 0.07 and 0.14 ($0.07 + 0.07 = 0.14$), then we put the second row into the new collection of samples. If the number is between 0.14 and 0.21 ($0.14 + 0.07 = 0.21$), then we put the third row into the new collection of samples, and if the number is between 0.21 and 0.70 ($0.21 + 0.49 = 0.70$), then we put the fourth row into the new collection of samples [6].

Now, for example, the first number picked is 0.72, then the fifth row would be kept into the new collection of samples. If the second number picked is 0.42, then the fourth row would be kept into the new collection of samples. We continue to pick random numbers and add samples to the new collection until the size of the collection is the same size as the original. After completion, you would notice that the row with higher sample weight would be added to the new collection multiple times reflecting its larger sample weight.

Now this table, i.e., Fig. 19 is used further for second iteration and now each of the row, is again given the same sample weight as earlier (1/8); however, that does not mean that the next stump will not emphasize the need to correctly classify these samples. As four samples are the same, therefore they would be treated as a block, creating a large penalty for being misclassified.

The forest of stumps created by AdaBoost makes classification. Imagine four stumps are saying for a patient as having heart disease with their respective says as: 0.97, 0.32, 0.78, and 0.63. Alternatively imagine that two stumps are saying the patient does not have heart disease with their respective says as: 0.41 and 0.82. Now if we sum up the say values of positive stumps, it comes out to be 2.7 and summing the negative stumps comes out to be 1.23. Now, we can see that the sum value of positive stump is greater than the sum value of negative stump, and thus, the patient is ultimately classified as having a heart disease. This is how the errors that the first stump makes influence how the second stump is made and so on.

2 Cascading Classifiers

The cascade classifier have multiple stages as shown in Fig. 20, and within each stage, the boosting process is been performed, that is AdaBoost with multiple weak learners (ensemble method). The weak learners are the decision stumps, and boosting helps us to have highly accurate classifiers by taking the weighted average of the say value produced by the weak learners. You will be able to understand the workflow and the process of each segment while making an XML file by focusing on Fig. 20 [1, 4].

The classifier hovers over a specific region as per the current location of the sliding window on the image pixel matrix. The stage one of the classifier tries to detect any features within that window, and if any feature is found then that stage is said to be

Fig. 20 Larger sample weights

Chest Pain	Blocked Arteries	Patient Weight	Heart Disease
No	Yes	156	No
Yes	Yes	167	Yes
No	Yes	125	No
Yes	Yes	167	Yes
Yes	Yes	167	Yes
Yes	Yes	172	No
Yes	Yes	205	Yes
Yes	Yes	167	Yes

positive and the next stage in the same window starts searching for features. This continues until the last stage is reached, and if the last stage also returns positive then it is confirmed that a feature is present within the window and then the window slides to the next location. To the contrary, if any of the stages returns negative, then the window slides to the next location assuming that it has not encountered any promising feature for the detection. As mentioned above, the stages can neglect the negative features very quickly. Here, time complexity comes in to picture, where a big majority of the features are negative, whereas the object we want to detect is rarely found and must need time to get fully verified [5].

- A true positive is found when a positive feature is correctly classified.
- A false positive is found when a negative feature is mistakenly classified as positive.
- A false negative is found when a positive feature is mistakenly classified as negative.

Now to get an efficient result, the false negative rate of each stage should be as low as possible. This is because if a positive feature is classified as negative then, the mistake cannot be corrected in the subsequent stages and thus in such a case the classification stops. However, if it classifies a negative feature as a positive one, then it can be changed in the later stages, and this is the reason why a classifier can have a high false positive rate. Thus, if we keep on adding multiple stages, then the overall false positive rate decreases, but to the contrary, the true positive rate also decreases as the detected feature has to go through several levels of verification [7].

3 To Create the Haar Cascade Files

There are many ways to create a Haar cascade file, but the method that we focus here is the easiest and fastest way to create it with a high level of accuracy, as it offers a different UI to the user. Please do visit this link (http://amin-ahmadi.com/cascade-trainer-gui/) to download the software. This software is named Cascade Trainer GUI which provides you with a very user-friendly interface to work with. This is developed by Amin Ahmadi (Tazehkandi) an Iranian author, developer, and computer vision expert [2, 3].

After downloading the software, it will look something like this as shown in Fig. 21.

Here, you can see that there are four partitions made:

- Input (used to provide the path for training data)
- Common (used for the settings of classifier)
- Cascade (settings used while cascading)
- Boost (settings used while boosting).

Before starting with the stages, we should first create a folder in any of your desired location and name (in my case let us name it cascade), and inside it, you must create two separate folders named 'p' (for positive image samples) and 'n' (for negative image samples). The positive image samples in the folder 'p' should strictly contain only the image that you want the algorithm to detect (e.g., if you want to detect faces, then the images in p folder should contain images which has only faces and no other objects in it). It would be even more helpful for your algorithm to train if you convert all the colored images in this folder to gray scale, so that the algorithm does not find any distractions in the image (in the form of hue, saturation, and brightness).

This technique is even more special because this software does not need to be fed a huge amount of samples for better accuracy. You can get decent accuracy even with a positive sample size of 50. Remember to provide at least double the negative number of samples than the positive ones. Similar to the positive samples, the negative samples can have anything related to the detecting object but should not contain any part of the object to be detected.

The Input Block:

Fig. 21 Process of XML file

This stage contains only four parameters listed below.

(i) *Sample Folder*: You will be providing the folder directory/or the path of the folder in which you have kept the positive and negative samples.
(ii) *Positive Image Usage*: This is the percentage of positive images used for training in every stage of cascading as mentioned in page no. 16.
(iii) *Negative Image Count*: This is the number of negative images used for the training in every stage of cascading. Make sure not to make this as a high number, as this might result in memory shortage error.
(iv) *Force Positive Sample Count*: This is where you can create multiple positive images with a single positive image. This tab shows you the number of positives you want to generate from each positive sample (try to stick to 1 unless really required).

The common settings: common setting page as shown in Fig. 22.

Let us start with the **Common Settings Block**:

(i) *Number of stages*: This is the number of stages you want in every block of detection in the image. Generally, these stages are between 15 and 20 in number.
(ii) *Pre-calculated Buffer Size* (Mb): You can see two options by this name in your screen, and both of them are related to the speed of your training. It is on you how much memory you want to assign to these tabs, but always keep in mind that you should not assign more memory combining both, than you have in your system in total. To the contrary, assigning too less memory is also not good as it may not satisfy your algorithm's memory needs. Let us assume, if

Fig. 22 Cascade Trainer GUI

you have 8 GB ram in your system, then your combined allocated memory should not be more than that, or else you might encounter an algorithm crash.

(iii) *Number of Threads*: This is the highest number of threads you can use during your training. Make a note that the number of threads available may differ according to your machine. If you have TBB installed, then you can see that the total available threads are being displayed automatically, and which becomes a default option. It is recommended for you to have this software if you want to make changes to this parameter.
TBB is known as threading building blocks. It is a library in C++ which was introduced by Intel for purposes like parallel programming on systems having more than 1 core. It is used to break a problem into multiple sub-problems and compute each one simultaneously. Inside this library, you get functions which tell the threads to work in an organized way. To use TBB, you first need to install it in your system, and you can download it through this link https://github.com/oneapi-src/oneTBB (recommended not to use it unless you have a stronghold on it).

(iv) *Acceptance Ratio Break Value*: This is a parameter which shows how precise your algorithm must be. Generally this value is set to 10e−5, to ensure that the model gets perfectly trained. This value also shows that if the model crosses this threshold, then you might have a chance of getting a model which is over-trained. Now, if you want to disable this feature, then you can simply change the value to −1. This is a kind of threshold preciseness value that your algorithm must satisfy.

(v) *Base Format Save*: This parameter is related to the Haar features that you have learnt earlier. If you click the checkbox then you are actually telling the software to use the ancient version of the classifier because you do not possess the required specifications in your system to go forward with your training. Thus, do not put much importance on this option unless you have a really old configuration system.

The **Cascade Block**: the cascade block as shown in Fig. 23

(i) *Sample Width*: This is the width of the image that you are providing for training (in pixels).

(ii) *Sample Height*: This is the height of the image that you are providing for training (in pixels).
While setting the size of the image, it is recommended to keep the aspect of 24, and obviously do not put the size of the image very large or else it might affect the speed of your training. Regarding the aspect let us say you have an image of size 500 × 300, and on calculating its aspect, you get as 1.66:1. Now multiply both sides with 24 to keep the same aspect. On doing so, we finally get 40 × 24 for your image's width and height, respectively.

(iii) *Feature Type*: Type of features: Haar—Haar-like features, LBP—local binary patterns. HOG—histogram of oriented gradients.

Fig. 23 Common settings in Cascade Trainer GUI

We use HOG only if we have OpenCV version of 3.1 or later. Since in this paper we are focusing on the details of Haar and its features, so please do keep this parameter value restricted to Haar.

(iv) *Haar Feature Type*: This parameter is defined if you want to use the basic features of Haar or all of the features. Basic includes only the vertical and the horizontal features described in page no. 4, whereas all include the features which are diagonal or even at an angle.

The Boost Block: the boost block is shown in Figs. 24 and 25

Fig. 24 Cascade block

Fig. 25 Boost block

(i) *Boost Type*: You can see these options in the dropdown list {DAB, RAB, LB, GAB(default)}. These are the four different types of AdaBoost techniques that you can apply here. The different names of these techniques are: DAB (Discrete AdaBoost), RAB (Real AdaBoost), LB (LogitBoost), and GAB (Gentle AdaBoost). I recommend to use the GAB (default value), the one explained in the above sections.

(ii) *Minimal Hit Rate*: Hit rate (HR) is the minimum threshold that any algorithm must pass through in each stage of the classifier. It can be defined as the number of positive objects that are correctly identified within the given stage. The default value is 0.995, and you can change this if you want your model to be lenient or strict accordingly. The total HR can be calculated using the formula as shown in Eq. 8

$$\text{minimum hit rate}^{\text{number of stages}} \tag{8}$$

(iii) *Maximal False Alarm Rate*: False Alarm (FA) is the maximum threshold that any algorithm must have in each stage of the classifier. It can be defined as the number of negative images that are incorrectly identified within the given stage. The default value is 0.5, and you can change this if you want your model to be strict or lenient accordingly. The total FA can be calculated using the formula as shown in Eq. 9

$$\text{maximum false alarm rate}^{\text{number of stges}} \tag{9}$$

(iv) *Weight Trim Rate*: This parameter can have value between 0 and 1. It is used to save computational time. Samples with summary weight $\leq 1 -$

Weight Trim Rate do not participate in the next iteration of training. You can set the value to 0, to turn off this functionality. The default value to this parameter is 0.95.

(v) *Maximal Depth Weak Tree*: This is the maximum depth of the weak tree you want. Default value is 1 that is a stump as discussed earlier on page 10. You can even change to any value, but you must know the insights of the process when you use maximum depth more than 1.

(vi) *Maximal Weak Trees*: This is the maximum number of weak learners/stumps you want to have in each stage of the classifier.

4 Results and Conclusion

The detailed analysis and implementations proved that the Haar cascade algorithm can be effectively used to identify the features of an object, and the mathematical analysis is shown for all the stages of the Haar cascade algorithm, i.e., Haar feature selection, creating integral images, AdaBoost training, and cascading classifiers. Finally, an overview and step-by-step procedure has been shown to create a Haar cascade file.

References

1. http://www.willberger.org/cascade-haar-explained/
2. T.R. Phase, S.S. Patil, Building custom HAAR-Cascade classifier for face detection. Int. J. Eng. Res. Technol. (IJERT) **08**(12) (2019). ISSN (Online): 2278-0181
3. http://amin-ahmadi.com/cascade-trainer-gui/
4. https://docs.opencv.org/master/dc/d88/tutorial_traincascade.html
5. R. Yustiawati et al., Analyzing of different features using Haar cascade classifier, in *2018 International Conference on Electrical Engineering and Computer Science (ICECOS), Pangkal Pinang* (2018), pp. 129–134. https://doi.org/10.1109/ICECOS.2018.8605266
6. D.K. Ulfa, D.H. Widyantoro, Implementation of Haar cascade classifier for motorcycle detection, in *2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Phuket* (2017), pp. 39–44. https://doi.org/10.1109/CYBERNETICSCOM.2017.8311712
7. É.K. Shimomoto, A. Kimura, R. Belém, A faster face detection method combining Bayesian and Haar cascade classifiers, in *2015 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Santiago* (2015), pp. 7–12. https://doi.org/10.1109/Chilecon.2015.7400344
8. I. Paliy, Face detection using Haar-like features cascade and convolutional neural network, in *2008 International Conference on "Modern Problems of Radio Engineering, Telecommunications and Computer Science" (TCSET), Lviv-Slavsko* (2008), pp. 375–377.
9. Q. Li et al., Multi-view face detector using a single cascade classifier, in *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), Chengdu* (2016), pp. 464–468. https://doi.org/10.1109/SKIMA.2016.7916267
10. http://uu.diva-portal.org/smash/get/diva2:601707/FULLTEXT01.pdf

Sentiment Analysis of Twitter Posts in English, Kannada and Hindi languages

Saritha Shetty, Sarika Hegde, Savitha Shetty, Deepthi Shetty, M. R. Sowmya, Rahul Shetty, Sourabh Rao, and Yashas Shetty

Abstract The World Wide Web has always been a huge repository for information. With this, an era of social networking came into existence that allowed people across the globe to connect and communicate with each other. The data present here is of various types like text, images, videos, audios, etc. The major part of the internet includes textual data which is written in different human-understandable languages. Recently, there has been tremendous increase in this data written in different languages over social media. This paper analyzes the semantic nature of a text, classifying them as either positive or negative for English, Hindi and Kannada languages. The following classification task is resolved with a CNN-LSTM-based neural network.

Keywords Twitter post · Semantic analysis · Machine learning · Natural language processing

1 Introduction

Social networking has evolved to become one of the major sources of information. Microblogging websites are one among these social networking websites. Twitter

S. Shetty (✉)
Department of MCA, NMAM Institute of Technology, Karkala, India
e-mail: shettysaritha1@nitte.edu.in

S. Hegde · S. Shetty · R. Shetty · S. Rao · Y. Shetty
Department of CSE, NMAM Institute of Technology, Karkala, India
e-mail: sarika.hegde@nitte.edu.in

S. Shetty
e-mail: shettysavi1@nitte.edu.in

D. Shetty · M. R. Sowmya
Department of CSE, NMIT, Bangalore, India
e-mail: deepthi.shetty@nmit.ac.in

M. R. Sowmya
e-mail: sowmya.mr@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_29

being one of the very popular microblogging websites where the users have freedom to create and share short messages, also called as "tweets." Sometimes, these tweets are used to express the opinions as well as comments of the users on different topics that may be political, social, etc. Large number of businesses spend a fortune in these microblogging sites to engage with their customers and understand their opinions about different products and services. Sentiment analysis plays a major role in determining the emotions and attitudes of the user.

In this paper, we have applied sentiment analysis on Twitter to indicate if a tweet is of negative or positive sentiment. The text used for analysis can range from big document to small messages or tweets. A supervised learning algorithm is preferred for training a classifier that needs a hand-labeled suitable trained data. As the amount of data and the number of languages increases, it becomes very difficult to manually collect all data to do training for sentiment classifier in the case of tweets. Along with the already existing data sets, we added more training data to analyze the posts based on the ongoing trends and fashion. This helps the consumers to research about the products or services before making a purchase. The semantic analysis is used by organizations and businesses to get customer reviews and feedback about problems in newly released products. We present the experiment results on three different languages and also discuss on how to improve these results.

2 Literature Survey

2.1 Related Work

As we all know, there has been a vast amount of research in the field of sentiment analysis. This is true when we need to analyze product or movie reviews. Sentiment analysis of the tweets is much harder than the conventional documents we find. It is not only true because of the informal languages used but also because of the excessive use of short-length slang words. Pang and Lee [1] provide an up to date survey of the preceding work on sentiment analysis. Sentiment analysis started from the document-level classification which was then handled at the sentence level and now much recently is being carried out at the phrase level. Bruns et al. [2] discussed regarding communication patterns in Twitter tweets. Go et al. [3] explored making use of different n-gram features in conjunction with the parts of speech (POS) tags while training the supervised classifiers like naïve Bayes (NB), maximum entropy (MaxEnt) and support vector machines (SVMs). They found that MaxEnt trained to form a combination of POS tags unigrams by almost 3%. Also, a different finding was reported in Go et al. [3] that adding POS tag features in the case of n-grams gives better accuracy for sentiment classification on tweets.

Agarwal et al. [4] explained the POS features, lexicon feature along with microblogging features. After combining the different features, they also designed tree representation of the tweets to combine the many categories of features into

one concise representation. Another significant word on this classification task is by Feng and Barbosa [5]. They used polarity prediction from three different websites and labeled these as noisy labels to provide training for a model and further they manually labeled 1000 tweets for further tuning and another 1000 for testing. They did not mention how they have collected their test data. They indicate the use of syntax features of tweet like retweet, hashtags, link, punctuation and exclamation marks along with features like prior polarity of words and parts of speech of words.

Instead of directly integrating the microblogging features into semantic classifier training. Speriosu et al. [6] constructed a graph which has all some of the microblogging features like hashtags and emoticons together with users, tweets, word unigrams and bigrams as its nodes which are connected based on the link that existed among them (e.g., Users are connected to the tweets they created and the tweets are connected to the word unigrams that they contain).

Existing work concentrates on the use of three main types of features: lexicon, POS features and microblogging features in case of sentiment analysis. Some emphasized the use of microblogging features, whereas some others argued about the importance of the POS tags with or without word priority involved.

In the paper written by Berend et al. [7], they used a supervised learning approach and text preprocessing and feature engineering to classify a Twitter post as negative, neutral or positive. Our approach to problem is a similar technique, but the feature engineering is embedded in the neural network that we use.

Shetty et al. [9] discussed about transliteration of input text from Kannada to Braille. Shetty et al. [10] discussed about the significance of anusvara and visarga in Kannada to English transliterations.

3 Proposed System

The proposed system classifies if a given text sample is of positive or negative sentiment. For this, we need to have the text ready for classification which has to be done by applying various preprocessing and transformations to make it suitable for our training model.

3.1 Model Training

We define three different models for English, Hindi and Kannada languages. The training inputs to these models are a preprocessed dataset. The preprocessing involves removal of stop words, unnecessary characters, lemmatization in the case of English language and finally tokenizing the sentence. A bag of words model is created which contains a list of all words from the dataset. A unique integer mapping is created for each token and we use this to map our sentence to a new integer vector. This vector is

given to each model for training along with their corresponding positive or negative labels.

3.2 Classification

After the training process, we save the model weights and the bag of words model. The new test sample is preprocessed with the same steps and then given to our model. The model makes a prediction on a probability scale, saying if a given text is positive or negative.

4 Methodology

4.1 Dataset

In order to train our three different models, we needed datasets in English, Kannada and Hindi languages. It was quite easy to find the English dataset which was already classified as positive or negative in the sentiment 140 dataset. Here, they tried to gather data based on searching certain keywords. The work about this is mentioned in the paper written by Richa Bhayani, Lei Huang and Alec Go. This constituted about 1600 k samples of posts with already classified sentiment as neutral, negative or positive. For this paper, we took the positive and negative samples of English from this dataset. For the Hindi language, the dataset was taken from a website source where they used similar crawler to capture posts from Twitter that were written in Hindi [8]. Additional data were collected from Hindi news articles and were manually tagged as positive or negative. About 2 k positive and negative samples were made for Hindi. Table 1 shows the dataset used (Fig. 1).

Kannada text samples are quite less in numbers when compared to the remaining two languages. We found few samples by going through some article websites, and for the rest of the data, we used the Google translator API to convert these English and Hindi texts to Kannada. A better translation was occurring when English was converted to Kannada than Hindi. So initially, we took some 2 k pre-classified positive and negative samples from English dataset and converted them to Kannada.

Table 1 Dataset used

Dataset	Total samples
Sentiment 140 for English	1,600,000
Hindi dataset	2500
Kannada dataset	6300

Fig. 1 Structure of the proposed system

Since the Twitter posts had some unnecessary or false tokens in English, the translator would straightaway keep them as it is. This issue was resolved in the preprocessing stage where the relevant tokens were only preserved. Later, we added in some 4 k extra samples by applying the same conversion process to increase the training accuracy and to increase the word count in the bag of word model.

4.2 Preprocessing

The dataset which was collected from Twitter posts will have unnecessary symbols like emoticons, foreign characters, hashtags or website links. The objective of this section is to remove all the noise from the data and prepare a normalized form that is suitable for our neural network. The following preprocessing steps were taken for all the three languages to normalize the data.

4.2.1 Removing Mentions

Mentions are common features for Twitter. They are used to specify or tag a particular individual person. All the mentions in the post always begin with '@' symbol. Since these mentions do not actually contribute to our sentimental classification, it is an unnecessary token. A regex string matching token r'@[a-zA-Z0-9]+' is used to identify a general set of mentions that occur in the post and are replaced with an empty character. For example, the text "happy @SparksCharity" is transformed to "happy."

4.2.2 Processing Hashtags

Hashtags are used to pin a Twitter post into a particular category and they always begin with "#" symbol. In most cases, people write a post and describe their emotions by writing in these hashtags. There the contents that come after the hash symbols are retained for further preprocessing. Consider, for example, the text "Feeling good #happy." This is transformed to "Feeling good happy" which retains the hashtag content.

4.2.3 Removing Links

People tend to post links with their post that points to a picture, video, gif or to some other website. These links do not contribute to the sentimental processing and are therefore removed. The links on the internet begin with a "http" or "https" and are followed by sequence of alphabets, digits and some of the punctuation symbols like "_" or "&". A regex expression is used to identify a general set of these from the Twitter post and is then replaced with null characters.

4.2.4 Processing Texts

For the English language, we have lower and upper case alphabets. To keep everything normalized, we convert any alphabet to its lower case form so that text tokenization and vectorizing become easier. Hindi and Kannada languages do not have this concept of lower and upper case letters. So no such transforms are applied in this case. Further unnecessary symbols, punctuation marks, numbers and foreign characters were removed by checking them against the corresponding alphabet sets for each language. As mentioned before, the Kannada dataset retained some English words, so these foreign characters are removed by the same procedure. Trailing whitespaces were removed from the text which were unnecessary for tokenizing.

4.2.5 Tagging Unknown Symbols

There are cases when some of the symbols do not match the alphabet sets or do not contribute to the other unknown symbols. For example, some of the Unicode characters contain a mixture of these alphabets and punctuations. We identified such symbols by the process of elimination and then marked them as "<UKN>" tag which specified that the token identified there is unknown.

4.2.6 Lemmatization

It is the process of converting a word or token to its actual form by removing its inflectional endings. The output of this is the word that conveys the meaning of the original entity. In any language, the words are in different forms. Just by considering the tenses, we can form it in three tenses like past, present and past participle.

In English, the word "studies," "studied" is related to "study." Here, "study" is the actual lemma transform of those two words. This transform was applied over English processing text using the WordNetLemmatizer provided by the natural language toolkit (NLTK) library. For the other language, we kept them as such and continued with the next stage.

4.2.7 Removing Stop Words

Stop words are the commonly used words in a sentence that do not contribute to its meaning or sentiment. For English, the words like "the, a, an, and, is" are some of the stop words. They are only used to specify a particular action or talk about an entity. Nltk library provides an English set of stop words which were used to eliminate the stop words in English dataset. As of Hindi dataset, a website source [10] provided with some set of stop words like " इस", " इसका", " इसकक" which was used to remove the stop words from Hindi dataset.

4.3 Bag of Words Model

The final model can only take in sequence of integer-based input for training and classification. For this, we need to have a mapping between each word to a unique number. We make use of bag of words model to collect and store all the words available in the dataset.

The first step for this is to tokenize our text into individual words. After applying all the preprocessing steps, we do this stage and collect all the words. Then we use a set data structure to store these words so that only unique words are collected. This model now represents a word to integer mapping where each integer specifies the location of that word in that bag (Fig. 2).

4.4 Dataset to Vector Transform

Using the bag of words, model and our tokenized set of texts, we create a new vector which is of some fixed length. This vector contains a sequence of integers where each integers maps on to a specific word. Since the neural network needs to take in a fixed-length input for training and classification, we limit the number of words in

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 70, 128)	37282304
conv1d_4 (Conv1D)	(None, 70, 32)	12320
max_pooling1d_3 (MaxPooling1	(None, 35, 32)	0
conv1d_5 (Conv1D)	(None, 35, 64)	6208
max_pooling1d_4 (MaxPooling1	(None, 17, 64)	0
lstm_2 (LSTM)	(None, 128)	98816
dense_1 (Dense)	(None, 1)	129

Total params: 37,399,777
Trainable params: 37,399,777
Non-trainable params: 0

Fig. 2 Structure of the neural network along with details about the number of parameters and output shape at each layer of the network

a text to 70 words. As of this paper, the tweet length is from 140 to 280 characters and based on this, we assume that a maximum length of words that might appear to be around 70. So a sentence is supposed to be less than or equal to 70 words.

In the case of texts with less word count, we pad in extra 0's to the list to make its length equal to 70 words and for larger length texts, the words after 70 are ignored. The procedure is used to convert English, Hindi and Kannada datasets to a suitable vector. For example, if our bag of word model has the following words: [a, is, the, apple, fruit], then the sentence "Apple is a fruit" is converted to [3, 1, 0, 4, 0…0].

4.4.1 Splitting the Dataset

The training process for our neural network needs two types of datasets. One for training and another for validation. The entire dataset is divided into 80–20% where the 80% contributes to the training steps and the remaining 20% for the validation. In most cases, the dataset tends to group positive or negative classified dataset in sequence of rows. Hence, shuffling was done before splitting in order to have the dataset normally distributed with both samples.

4.4.2 Training the Neural Network

The neural network model uses embedding layer as its first hidden layer where it takes in the size of vocabulary which in our case is the size of the bag of words and plus and also the length of each text sample which is 70 words per vector. Vocabulary size defines the number of unique words along with unknowns that are there in the entire dataset.

Following this, we introduce convolutional neural network (CNN) architecture layers which are found to produce promising results for classification problems. 32 filters are used at the first CNN layer and the outputs of these are downsampled by passing them through a maxpooling layer. The outputs of this layer are given as inputs to another CNN layer with 64 filters and are downsampled again.

A rectified linear activation function is used at CNN layers for triggering the neurons based on a particular threshold value. The outputs of this is given to LSTM and are mapped onto a single output layer where the result is normalized by using a sigmoid activation function. The normalized output lies within the range of 0–1. Here, the output specifies the sentiment of the given text vector. If the value is close to 0, then the sentiment is negative and if it is close to 1, then the sentiment of the text is positive. The binary cross-entropy is used as its loss function since our problem involves classifying results into two set of classes. Adam optimizer function is used to keep track of the model accuracy and adapt our model to classify the data samples into two categories.

The same architecture was used for all the three languages but trained with their corresponding dataset. Training was done for about 20 passes with giving in 128–256 data samples at a time. Validation dataset was used to compute the validation accuracy for the unknown samples.

5 Results and Discussion

5.1 Results for English Dataset

The model when trained for English dataset which had about 1600 k samples provided about 95% accuracy for 20 pass through the dataset. Initial model was created with lesser parameters to train and when it was trained for 5 epochs, it provided with 75% accuracy. Later it was redesigned and modified to give about 89% accuracy.

The current model was trained to 95% accuracy with 20 epochs. The graph in Fig. 3 shows how the model learned to improve its understanding about the dataset being positive or negative as it passed through the samples again and again. Then the model was tested against custom inputs as shown in Fig. 4. Here, for the input "this is bad," clearly, it depicts a negative context and due to which the model was able to identify that the sentiment is negative. Here, the numerical besides the results shows in the actual result of the model. The model generates a value of 0.15 which is quite close to 0, and it classifies the text as negative. As in general, we assumed that the values lesser than 0.5 are negative and higher or equal values are of positive sentiment.

The first example "this is a nice product" shows a positive sentiment as the keyword "nice" represents a good context. The model generates an output of 0.99 which is strongly close to 1 and shows that the text is of positive sentiment.

Fig. 3 **a** Graph of accuracy versus time for the model trained over English dataset. **b** Graph of loss versus time for the same model

```
-----------------------------------
this is a nice product
Result: Positive [0.9990637]
-----------------------------------
this is bad
Result: Negative [0.15898627]
-----------------------------------
good show
Result: Positive [0.9611314]
-----------------------------------
hi how are you mate?
Result: Positive [0.95421857]
```

Fig. 4 Results for English samples

5.2 *Results for Hindi Dataset*

The model was about to achieve 99% accuracy with the Hindi dataset. Figure 5 represents how the model started off with 0.7 loss and as it trained over the data again and again, it was able to reach a loss close to 0.

Fig. 5 **a** Graph of accuracy versus time for the model trained over Hindi dataset. **b** Graph of loss versus time for the same model

In Fig. 6, actual result represents the output from the model and expected result represents the real output.

Clearly, the expected and actual results are close to each other. The same network which was used for English dataset was used but with new dataset for training. Here,

Fig. 6 Results for Hindi samples

the last two test samples were chosen out of the dataset and the results of them being positive is shown by getting the probability near to 1 and for negative case, we get a probability 0.

5.3 Results for Kannada Dataset

Figure 7 shows how the model trained over the Kannada dataset and improved its accuracy over time. The model reached about 99% accuracy with 6 k dataset just by training it for 20 epochs.

To test out the model, we took some sample cases from the dataset and some were manually typed. The expected and actual result accuracy matches with a small difference (Fig. 8).

Demonstration is done on the value of using semantics to classify a Twitter tweet as positive or negative. The model is built using neural networks consisting of different

Fig. 7 **a** Graph of accuracy versus time for the model trained over Kannada dataset **b** Graph of loss versus time for the same model

Fig. 8 Results for Kannada samples

```
ಅವನು ಒಳ್ಳೆಯ ಮನುಷ್ಯ
Actual Result: Negative [0.1837786]
Expected Result: 1
----------------------------------------
ವ್ಯಾಯಾಮ ಆರೋಗ್ಯಕ್ಕೆ ಒಳ್ಳೆಯದು
Actual Result: Positive [0.78110516]
Expected Result: 1
----------------------------------------
ಅವನು ಕೆಟ್ಟವನು
Actual Result: Negative [0.21596132]
Expected Result: 0
----------------------------------------
ವೇಗವಾಗಿ ಚಾಲನೆ ಮಾಡುವುದು ಅಪಾಯಕಾರಿ
Actual Result: Positive [0.7166374]
Expected Result: 0
----------------------------------------
ಧ್ಯಾಮ್! ನಾನು 11:11 ತಪ್ಪಿಸಿಕೊಂಡಿದ್ದೇನೆ.

Actual Result: Negative [0.00055438]
Expected Result: 0
----------------------------------------
```

layers of CNN and LSTM, where the given input is mapped down to a single output of classification being positive or negative. The results discussed before show how well the English model behaves even though with 95% accuracy when compared to other two languages with 99%. Here, the accuracy mainly specifies how well the model behaves for the unknown validation dataset. Since the English dataset had 1600 k samples, the bag of word model it created consisted of huge words when compared to that of Kannada and Hindi language. So the English dataset trained model can take in a large set of values from the domain and produce a better result. The words appearing in the test run which are not in the bag of words would be replaced with an unknown or a padding symbol. This causes a problem where the entire words in the text become unknown and at that time, the model produces bias result.

To resolve this problem for English language, the NLTK library provides with WordNet which is a graph of how different words are connected. Words with similar meaning are close to each other. By using this and analyzing the text, we can replace these unknown words with the ones that match close. Again, there is a problem where the words never appear on the WordNet graph are used in the sentence. Usually, in Twitter posts, there could be cases where people use words that depict emotion like "Haha" or "huh." These words do not correspond to any means as they just represent an emotion. For these cases, our model produces bad results when the sentences have no other contexts. Solution for this includes creating or adding our custom words to WordNet which depicts these emotions. There are many models that still make use of n-gram feature extraction for classification. Here, they create a list of these n-grams which are words of length n. So a text "hello" can be reduced down to their bi-grams ($n = 2$) as "he, el, ll, lo" and use these to create the vector. This vector can be later used to classify the sentences. But the problem seen here is that the words

Table 2 Comparison of accuracies for various datasets

Dataset	Samples	Accuracy (%)
Hindi dataset	2500	99
Sentiment 140 for English	1600 k	95
Kannada dataset	6300	99

with prefix or suffix which changes the meaning of the word make the model produce ambiguous result. Considering the word "unhappy," the prefix "un" means that the word in actual is a negative meaning but the model needs to see the word as a whole than just its grams. Table 2 indicates accuracies for various samples used.

6 Conclusion and Future Work

We indicated how neural networks are used to classify a Twitter post as positive or negative sentiment for English, Hindi and Kannada languages. With the bag of words model, the network produces good results if the words from this bag are present in the sentence. More research is needed for techniques where parts of speech tagging is used to identify the relevant words and then perform other preprocessing operations and for identifying similar words. Using Google translator to produce the dataset for other languages helped us create more samples and produce a better result. Adding in the CNN and LSTM layers in the neural network helped in increasing the accuracy of the model in a significant way. When the model was evaluated against sarcastic posts where the intended meaning of the sentence is different than the post, here the model produced the wrong result due to the negative contexts being used. Even adding more samples to the dataset did not improve the output of these sentences. For handling such posts, we might need to make use of different networks to classify if the sentence is sarcastic or not and then apply sentimental analysis.

The future work for this involves adding in a word similarity check with WordNet model. Our current model only classifies the posts into negative or positive classes. We can add in more classes like neutral, anger, sad, happy, etc. and classify a text based on these. A post can also belong to multiple categories but not just two classes. Hence, work is needed on this to improve the sentimental analysis process.

References

1. B. Pang, L. Lee, Opinion mining and sentiment analysis. Comput. Linguist. **35**(2), 311–312 (2009)
2. A. Bruns, S. Stieglitz, Quantitative approaches to comparing communication patterns on Twitter. J. Technol. Hum. Serv. **30**(3–4), 160–185 (2012)
3. A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision. CS224N project report. Stanford **1**(12) (2009)

4. A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R.J. Passonneau, Sentiment analysis of twitter data, in *Proceedings of the Workshop on Language in Social Media (LSM 2011)* (2011), pp. 30–38
5. L. Barbosa, J. Feng, Robust sentiment detection on twitter from biased and noisy data, in *Coling 2010*: *Posters* (2010), pp. 36–44
6. M. Speriosu, N. Sudan, S. Upadhyay, J. Baldridge, Twitter polarity classification with label propagation over lexical links and the follower graph, in *Proceedings of the First Workshop on Unsupervised Learning in NLP* (2011), pp. 53–63
7. V. Hangya, G. Berend, R. Farkas, Szte-nlp: Sentiment detection on twitter messages, in *Association for Computational Linguistics* (2013)
8. Github Sentiment-Analysis-of-Hindi-Tweets [Online]. Available https://github.com/Negibabu/Sentiment-Analysis-of-Hindi-Tweets
9. S. Shetty, S. Shetty, S. Hegde, K. Pandit, Transliteration of text input from Kannada to Braille and vice versa, in *2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* (IEEE, 2019), pp. 1–4
10. S. Shetty, S. Shetty, S. Hegde, K. Pandit, English transliteration of Kannada words with Anusvara and Visarga, in *Advances in Artificial Intelligence and Data Engineering* (Springer, Singapore), pp. 349–361

An Efficient Algorithm for Fruit Ripeness Detection

Sharath Kumar and Ramyashree

Abstract The assessment and ranking are now carried out on the basis of observations and practice. A major computing concern is the identification of various fruit images. This thesis describes the grading approach applied to the fruit. The identification of good and bad fruits in this analysis illustrates the approaches used by MATLAB. The created approach starts the procedure by taking the image of the fruit using a normal digital camera. The image is then sent to the processing level, where MATLAB is used to isolate, identify, and rate the elements

Keywords Ripeness · Preprocessing · MATLAB · Pixel

1 Introduction

The primary industrial field is agriculture, and it plays a major role in India's economic growth. In India, 70% of the population relies on agriculture, but there is no robotic form of automation in agriculture for Indian industries. Service workers are particularly oriented on the processes of manual labeling and ranking that are used to differentiate between different fruit varieties [2]. As these methods are driven by human activity, they are vulnerable to some kind of error. Because humans are exposed to fatigue and lack of energy, it is necessary to adopt an automated system to remove labor, and the automated system also helps to minimize the time wasted on manual techniques [9].

University researchers are designing several new agricultural automation technologies, posing concerns about the efficiency and effectiveness of current agricultural operations. Fruits are delicate materials, but they can be checked by non-destructive processes. Classification is important for the evaluation of an agricultural

S. Kumar · Ramyashree (✉)
Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi, Karnataka, India
e-mail: ramyashree.cs@sode-edu.in

S. Kumar
e-mail: sharathkumar.cs@sode-edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_30

Fig. 1 Set of fruits

commodity [1]. The most significant physical property is fruit scale, while color resembles a visual property. It is therefore important to distinguish fruits in the assessment of agricultural product, to meet quality requirements and to improve market demand. In preparation, packing, transportation, and marketing activities, it is also useful. The process can be too cumbersome and sometimes error-prone if classification and rating are achieved by way of manual procedures. We may use various algorithms to classify the fruit and discover the quality of the fruit by using MATLAB software as an image processing tool [4]. Finally, after receiving a lot of information, we identify whether the fruit is good or poor (Fig. 1).

2 Literature Survey

Kavdir et al. [1] have built up a framework which grades apples by utilizing shading and size. Fluffy rationale framework is intended for this reason. Reviewing results acquired from fluffy rationale demonstrated 89% general concurrence with the outcomes from the human master, giving great adaptability in mirroring the master's desires and evaluating principles into the outcomes. This utilization of apple evaluating can be completely computerized by estimating the necessary highlights by methods for innovative sensors or machine vision and settling on the reviewing choice utilizing fluffy rationale.

Feng and Qixin [2] proposed a system in which color image preparation-based smart organic product arrangement structure has been investigated in which the shading proportion of the natural product, which was calculated with HSI shading room, was chosen as an arrangement. The organic commodity structure was recognized by an outstanding Bayes classifier, the limits of which were obtained by an inspection panel.

Mustafa et al. [3] developed an automatic grading system that was suggested to classify four fruit and vegetable types, namely apples, bananas, oranges, mangoes, and carrots. Qualities, for example, field, fundamental pivot, little hub, and edge, have been utilized to distinguish the examples. In this article, methods, for example,

the assist vector with machining, are utilized to distinguish the examples and to rate the reviewed tests with fluffy rationale.

Chang et al. [4] have built up a dream-based framework for arranging the organic products. It depends on proportions of fluffiness and level of coordinating. Fluffy methodologies were utilized to decide ideal thresholding estimations of natural product's pictures, and fluffy level of coordinating was applied to order the shading and size of organic products.

Nagganaur and Sannanki [5] introduced the arranging and evaluating of organic products utilizing picture handling methods. The framework begins the cycle by catching the organic product's picture. At that point, the picture is communicated to the MATLAB for highlight extraction, arrangement, and evaluating. Both grouping and evaluating are acknowledged by fluffy rationale approach.

The shading and shape attributes of the organic product are definitive for visual examination. An incredible self-ruling natural product arranging framework must have the option to effectively order all boundaries. Natural product structure can precisely be delivered from an advanced picture utilizing customary strategies for picture preparing. There are a few techniques, for example, fuzzy rationale, neural network, in light of color histogram, genetic algorithm, etc. [6].

Wajid et al. [7] gave an overview to address the difficulties in natural products. Farming has a significant function in the financial improvement of our nation. Beneficial development what is more, high return creation of natural products is fundamental and required for the business. Use of picture handling has helped farming to improve yield assessment, infection discovery, natural product arranging, water system, and development evaluating. Picture preparing strategies can be utilized to lessen the time utilization and has made it cost effective. In this paper, they gave an overview to address these difficulties utilizing picture preparing methods.

Vaviya et al. [8], proposed the framework getting the picture of organic product under the test and contrast it and the highlights of normally aged leafy foods aged foods grown from the ground the yield with the likelihood. This method uses the cell phone that runs the Android application and the convolutionary neural organization to identify the natural product that is misleadingly aged.

Pramod and Devalatkar et al. [9] provide the techniques for preservation of fresh food, with the gains made in preserving fresh food items such as fruits and vegetables. The advanced visualization techniques for fruit and vegetable protection are now important parameters. Based on its different features, it involves the implementation of a fixed rule for individual items as a safe keeping algorithm. This research paper provides an overview of the tomato age factor focused on one of the color-based visual features of the tomato.

Iqbal et al. [10] explained about the classification of fruit diseases using ANN algorithm. Efficient growth and improved yield are so critical and important. As the agricultural industry is in high demand, it is important to use successful smart farming techniques in order to minimize more time and substitute manual sorting and does not offer significant results. The neural network are built for the classification of diseases and pattern matching and the network (ANN) algorithm are used. This methodology decreases human effort and can provide reliable results of 90%. The

work to be done in the future is to suggest proper work treatments accordingly until the condition is compromised.

3 Methodology

The architectural depiction is concerned with setting up a basic layout of a system. System design mainly deals with recognizing major components of the system being developed and how the interaction takes place among these components. The entire process includes various steps. The flowchart of this process is given as follows (Fig. 2).

Step 1: Image acquisition
Step 2: Preprocessing
Step 3: Pull out the characteristics of fruit selected
Step 4: Feature training
Step 5: Feature matching
Step 6: Displaying the results.

Image acquisition

Image acquisition deals with getting the image of the fruit of the interest, and it will be the first step in the entire process [3]. The success rate of the entire process purely depends upon the quality of the image being captured, and in order to achieve more accuracy, digital camera with high resolution is preferred. The captured image will be stuffed into the MATLAB with the help of built-in function imread, and with appropriate parameter, entire image is represented in the form of two-dimensional matrix which consists of rows and columns [13].

Preprocessing

As mentioned before, the quality of image is of prime concern. The image being captured will be preprocessed, so as to eliminate the noise components present or elimination of unwanted background or to enhance the resolution of image that is

Fig. 2 System design

Fig. 3 Preprocessing

being captured by using camera [5]. Preprocessing mechanisms involves following steps

- RGB image: Entire image is represented by using three color components: red, blue, and green. In this format, the image is represented in the form of matrix, and the values within the matrix represent each pixel intensity value.
- Background subtraction: Area of interest is separated by removing the background objects with help of background diminution process. Background diminution is a process extricating of foreground objects [11]. One more advantage of removing unwanted background will reduce data processing overhead
- Gray image: The gray-level images will have only one color component, which has an intensity of gray color in abundant ranges in between.
- Binary image: In binary image, each pixel is represented as bit which will be in the form of 0 or 1 (Fig. 3).

Feature extraction:

Assembling of input data entities into group of characteristics is called as feature extraction. The process of feature extraction begins with the extricated boundary of the sample [6]. The function employed to discover the features is "regionprops." The key characteristics extracted are major axis, area, and minor axis. In this yellow, green, and red colors are used for categorization. As we know that, fruit's skin color will be different. Hence, these colors are useful for arranging the fruits. The green and red constituents are computed by summing up the pixel values which belong to the green and red colors, and yellow constituent is computed by first transforming the original image into CMY [12].

Feature training:

Feature training is a database which includes plenty of trained characteristic values for bad and good fruit [8]. As the number of trained characteristics increases, more accuracy can be obtained. These trained values are compared with that of the original fruit image to get the results.

Feature matching:

Feature matching mechanisms basically consist of determining the characteristics in image which can be matched with characteristics in supplementary images from which a transformation prototype can be determined [7]. Feature matching is a key task in the area of image processing. With extracted characteristics, individual values are correlated with other and based on that decision can be taken about good or bad fruit.

3.1 Detailed Methodology of Fruit Ripeness Detection

The mechanism comprises of the following steps:

1. From the menu provided choose the fruit name and the necessary image from the group of images exhibited.
2. Rescale the image selected and project it.
3. Ascertain the different color components red, blue, and green in the selected image.
4. Represent the selected image in the form of matrix.
5. From the matrix determine the greatest pixel value.
6. Differentiate all the color components (r, g, b) with greatest value selected.
7. Based on the color component, color of the fruit and ripeness parameter is decided; if the red constituent is more in the image being tested, then fruit may be red, orange, or yellow, and we can conclude saying that fruit is ripe.
8. If blue constituent is more in the image being tested, then we can conclude saying that fruit is ripe, and it is blue in color.
9. Fruit will be of color green if green constituent is more and fruit is not ripe.
10. A dialog box that shows the output will be displayed (Fig. 4).

Fig. 4 Architecture of fruit ripeness detection

3.2 Detailed Methodology of Fruit Defect Detection

The mechanism comprises of the following steps:

11. From the menu provided, choose the fruit name and the necessary image from the group of images exhibited.
12. In window provided exhibit the input image.
13. Select the Analyze button from the menu provided.
14. Metamorphose the input image selected into HSV format and extricate all three components separately (H, S, and V).
15. Masking of the chosen image is done by comparing each color bands to the corresponding threshold value.
16. Once all three color components are masked, all the three components will be combined into a single image, and it will be displayed as a thresholded image
17. Thresholded image will be further processed and that will be represented in other forms such as grayscale and binary image and that will be displayed on the window provided.
18. The deficiency on the fruit is investigated and decision is taken; if the ripens quotient is more than 25, then rejected stamp will be displayed, otherwise approved stamp will be displayed (Fig. 5).

Thresholding

Thresholding includes various steps; out of these steps, the dominant one is background subtraction. Background subtraction is a mechanism to bring out foreground objects in a specific locus of an image [10]. A foreground object is interpreted as an object of scrutiny which assists in decreasing the quantity of data to be handled.

Fig. 5 Architecture of fruit defect detection

4 Results and Discussion

In main menu, there will be three functions: check ripeness, defect detection, and exit. Check ripeness function will check fruit is ripe or unripe; defect detection will analyze fruit and provide defected proportion in terms of percentage; and exit function is used to come out of the application (Fig. 6).

Once check ripeness function is selected, type of fruit selected from dropdown menu and fruit image must be uploaded for further processing (Fig. 7).

Once defect detection function is selected, select image from the drive and analyze the image. Based on image quality, approved or rejected stamp will be applied (Figs. 8 and 9).

Results on dataset:

As there is no relevant dataset available, we have gathered 76 images from few of the standard dataset and other Internet resources and analysis is done

Truly detected (TD)—The defective fruit which was concluded as defective one or healthy fruit which was declared as healthy fruit.

Falsely detected (FD)—Wrong decision taken by the algorithm either healthy fruit was detected as defective or defective fruit was concluded as health fruit.

Detection rate (DR)—TD/no. of images.

False detection rate (FDR)—FD/no. of images (Table 1).

Fig. 6 Main menu

Fig. 7 Selecting an input image from the dataset and displaying the corresponding output

Fig. 8 Final output showing fruit is approved

Fig. 9 Final output showing fruit is rejected

Table 1 Efficiency of proposed system

No. of images	DR	TD	FD	FDR
76	90.78	69	7	9.21

5 Conclusion and Future Work

This work confers mechanisms for arranging and grading of fruits based on quality. This process begins with recording the fruit's image using normal digital camera; the shape, size, and color of the fruit are used to categorize and estimate fruit's edibility. This automated approach can solve the problems of manual operating procedure, and we can make use of this in Agriculture Produce Corporation and Marketing, etc.

In this method, it is difficult to judge the quality of fruits that will not alter color whether they are unripe and ripe, while classifying one extra feature size can be included to get more accurate results.

References

1. I. Kavdir, D.E. Guyer, Apple grading using fuzzy logic. Turk. J. Agric. **2003**, 375–382 (2003)
2. G. Feng, C. Qixin, Study on color image processing based intelligent fruit sorting system, in *Proceedings of the 5 World Congress on Intelligent Control and Automation, Hangzhou, P.R. China*, June 15–19 (2004), pp. 4802–4805
3. N.B.A. Mustafa, S.K. Ahmed, Z. Ali, W.B. Yit, A.A.Z. Abidin, Z.A.M. Sharrif, Agricultural produce sorting and grading using support vector machines and fuzzy logic, in *2009 IEEE International Conference on Signal and Image Processing Applications* (2009), pp. 391–396
4. W.-H. Chang, S.C.S.-C. Lin, P.-Y. Huan, Y.-Y. Chen, Vision based fruit sorting system using measures of fuzziness and degree of matching. Department of Agricultural Machinery Engineering & Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (1994), pp. 2601–2604
5. Naganur and Sannakki, Fruits sorting and grading using Fuzzy Logic. Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET) 1(6) (2012)
6. Sagare, S.N.K., Fruits sorting and grading based on color and size. Int. J. Emerg. Technol. Comput. Appl. Sci. (IJETCAS), 12–333
7. A. Wajid, N.K. Singh, P. Junjun, M.A. Mughal, Recognition of ripe, unripe and scaled condition of orange citrus based on decision tree classification, in *2018 International Conference on Computing, Mathematics and Engineering Technologies, Invent, Innovate and Integrate for Socioeconomic Development, iCoMET 2018*, vol. 2018, January, no. April (2018), pp. 1–4
8. H. Vaviya, A. Yadav, V. Vishwakarma, N. Shah, Identification of artificially ripened fruits using machine learning, in *2nd International Conference on Advances in Science & Technology (ICAST)* 2019 on 8th, 9th April 2019 by K. J. Somaiya Institute of Engineering & Information Technology (Mumbai, India, 2019)
9. P.G. Devalatkar, S.R. Koli, Identification of age factor of fruit (tomato) using MATLAB-image processing. Int. J. Recent Trends Eng. Res. (IJRTER) 02(07) July-2016. ISSN: 2455-1457
10. S. Iqbal, A. Gopal, P.E. Sankaranarayanan, A.B. Nair, Classification of selected citrus fruits based on color using machine vision system. Int. J. Food Prop. **19**(2), 272–288 (2016)
11. S.R. Dubey, A.S. Jalal, Adapted approach for fruit disease identification using images. Int. J. Comput. Vis. Image Process. (IJCVIP) **2**(3), 44–58 (2012)
12. J.D. Pujari, R. Yakkundimath, A.S. Byadgi, Grading and classification of anthracnose fungal disease of fruits based on statistic texture features. Int. J. Adv. Sci. Technol. 52 (2013)
13. S. Maheswaran, Identification of artificially ripened fruits using smart phones, in *International Conference on Intelligent Computing and Control (I2C2)* (IEEE, 2017)

Kannada Document Classification Using Unicode Term Encoding Over Vector Space

R. Kasturi Rangan and B. S. Harish

Abstract Today, there is a great demand for extracting useful information and ability to take actionable insights from heaps of raw textual data. The processing of regional language texts, notably with low resources is challenging. Especially for Indian Regional Languages (IRL), language text processing is in huge requirement with regard to the applications of natural language processing (NLP) and text analytics. As a part of regional language processing and text analytics text documents classification for IRL are yet to be explored. Kannada is one of the official Indian regional languages. In this paper, the new benchmark Kannada document's dataset is created and analyzed using machine learning algorithms. This paper proposes an explicit Unicode term encoding based Kannada document classification, using the vector space model. Both term frequency (TF) and term frequency-inverse document frequency (TF-IDF), statistical measures are used for classification of Kannada documents using K-NN and SVM classifiers. SVM (linear) classifier performs better than the K-NN classifier in classifying the Kannada documents with 98.67% mean accuracy over K-Fold experiments.

Keywords Language processing · Text classification · Unicode encoding · Vector machine

1 Introduction

We are in the age of information, where the proliferations of data in the form of online documents, e-books, digital libraries, journal articles, etc., on the web, on corporate intranets, on news wire and elsewhere is overwhelming. However, ability to absorb

R. Kasturi Rangan (✉)
Department of Information Science & Engineering, Vidyavardhaka College of Engineering, Mysuru, India
e-mail: rkrangan@vvce.ac.in

B. S. Harish
Department of Information Science & Engineering, JSS Science and Technology University, Mysuru, India
e-mail: bsharish@jssstuniv.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_31

and process this large amount of data remains static. Text mining tries to solve this information explosion using various techniques like machine learning, natural language processing, information retrieval, and knowledge management. The textual information can be in any natural language. UNICODE for various natural language characters made easier in analyzing and handling multi-lingual data. India is a multi-lingual nation, where English is not widely understood. People prefer their regional language over English in many contexts. Hence computability of Indian regional languages overpasses many societal gaps in education, economy, and healthcare sectors.

Automatic text categorization is one among many tasks in natural language text analysis. It is the process of approximating an unknown category assignment function $F{:}D \times C \rightarrow \{0, 1\}$, where D represents the set of documents and C is the set of predefined categories. The value of $F(D, C)$ is 1 if the document D belongs to the category C and 0 otherwise [14]. Text analysis can be made at documents level, paragraphs level or at sentences level. The automatic text categorization is enforced in applications where there is a need of categorizing the dynamic flow of text information.

Kannada is one of the 22 official languages of Indian constitution. According to 2011 census, there are around 56 million Kannada speakers in India. In Kannada Document Classification (KDC), Kannada text documents are categorized into different domains based on the similarity. As there is lack of standard Kannada document's dataset, new dataset is created and presented in this paper. Section 2 of this paper consists of review of related work. In Sect. 3, preprocessing of data, explicit Unicode term encoding, feature extraction and selection and vector space model is explained. Dataset description, K-fold experimentation with K-NN and SVM classifiers are described with tabulation of results and graphs in Sect. 4. Concluding remarks are presented in Sect. 5.

2 Related Work

The abundant research works had carried out on text classification task in English and other non-Indian regional languages. Spam detection, documents categorization, named entity recognition, tagging contents or products in platforms like E-commerce, blogs, news agencies and so on are few applications of text categorization. Indian regional languages are less resourced, morphologically rich and agglutinative in nature [6]. Hence, it is quite difficult to develop intelligent systems for real time applications when compared to English natural language.

For Tamil which is one of the Indian regional languages, Rajan et al. in 2009 [11] presented automatic Tamil document classification system using most established and well-known vector space model (VSM) and also experimented using artificial neural network. The authors in Rajan et al. [11] created their own Tamil documents dataset. They experimented on both CIIL (Central Institute of Indian Languages) developed Tamil corpus and their own dataset. Rajan et al. [11] achieved around

90.33% and 93.33% accuracy, respectively, for VSM and ANN models. Similarly in VSM, Dhar et al. [1, 2, 5, 8] extensively experimented on Bangla text document categorization. The author in Dhar et al. [2] used TF-IDF features and distance metrics like Euclidean and Cosine similarity to categorize the dataset of 1000 Bangla documents. The results achieved are 95.20% and 95.80%, respectively. In [1], additional to term frequency (TF) and inverse document frequency (IDF) they added inverse class frequency (ICF) values and obtained good results for the newly created dataset of 4000 Bangla documents. In their experimentation, they incorporated K-fold ($K = 5$) cross validation. The proposed approach achieved a 98.87% of mean accuracy using Naïve Bayes Multinomial classifier. Further [5] presents the categorization task using multilayer perceptron (MLP) classifier with same TF-IDF-ICF frequency selection technique. Accuracy of 98.03% is achieved. Authors in Dhar et al. [8] use rule based classification algorithm called PART classifier for classifying 8000 Bangla documents into 8 categories. Term association and term aggregation are used in feature extraction process. Further K-fold cross validation is applied ($K = 5$) and obtained a mean accuracy of 96.83%. The experiments concluded that the PART classifier performs better when compared to other rule based classifiers (like RIPPER, Decision table, OneR) and Bayesian classifiers (like Naïve Bayes Multinomial, Naïve Bayes). Puri and Singh [13] experimented on Hindi documents using support vector machines (SVM). But the documents taken for experimentation are only four. Other than statistical approach, based on availability of Wordnet lexical database, few experiments are conducted on text classification. Mohanty et al. [3] presented Sanskrit language text classification based on semantic WordNets. This method constructs lexical chain of significant words from source document using semantic network of WordNet.

Text classification can be performed at document level, paragraph level or at sentence level. Jayashree et al. [12] focused on sentence level text classification in Kannada language. Both CIIL (Central Institute of Indian Languages) developed corpus and custom built dataset are used for experimentation. They found SVM classifier performed better than Naïve Bayes, and results are concluded by opining that the text classification can be extended to paragraph or document level. Similarly, Tummalapalli and Mamidi [10] presented sentence level classification based on sequence of syllables for English, Hindi, and Telugu languages. Further, based on sentence level classification, tasks like sentiment classification and question classification is experimented on 5 datasets. Following to this work, Tummalapalli et al. [4] applied neural network methods for sentence classification. Characters or syllables play a major role in understanding the morphologically rich languages. Hence in this work, the sentences are considered as input in both representations, i.e., sequence of syllables (or Characters N-gram) and sequence of words representations. Tummalapalli et al. [4] proposed MultiInput-CNN model that performs better with the representation of sequence of syllables over word vectors, in sentiment and questions classification tasks for Hindi and Telugu language datasets.

The clustering techniques are also used for document classification of Indian Regional Languages (IRL) by researchers. In 2017, Narhari et al. [7] worked on Marathi documents categorization using modified label induction grouping (LINGO)

clustering algorithm. They created their own dataset of 200 Marathi documents. They claimed that their results (Precision = 88.89%) were better than the former LINGO algorithmic experimental results. As there is a lot more research experiments yet to be carried out on Indian regional languages in various domains, this work motivates to do experiments more on Kannada document level text classification task using machine learning algorithms.

3 Proposed Method

As there is a lack of publicly available standard Kannada document's dataset, Kannada document classification is performed on newly created Kannada document's dataset. From the literature survey, document classification in Kannada language found undiscovered and researchers suggested experimentations at paragraph and document level classifications [12]. In newly created, Kannada document dataset there are total 300 documents, spread across five categories. Details of the dataset are discussed in Sect. 4. The Kannada document classification is the task of assigning the category (c_i) for the given document (d_j), where $C = \{c_1, c_2, c_3, ...c_m\}$ are set of predefined (m) categories and $D = \{d_1, d_2, d_3, ...d_n\}$ are set of (n) documents. The block diagram of proposed methodology is shown in Fig. 1. The block diagram depicts the steps followed for the Kannada Document Classification (KDC).

3.1 Preprocessing

At first, the raw Kannada text documents are preprocessed to remove unwanted texts and clean the data as per the computational requirement. The removal of unwanted information in the texts like punctuations and numbers from the dataset are carried out based on the regular expressions. Removal of these, makes data clean and avoids

Fig. 1 The proposed block diagram of Kannada Document Classification (KDC) model

the ambiguity during the categorization. Later, tokenization is applied to obtain the tokens. Tokens are the basic unit of the documents. Each term of the document forms a token because tokenization is done at Word/Term level. It can also be achieved at sentence, character, and at N-gram levels [15]. Further, the documents are analyzed by the terms/tokens identified during tokenization. Tokenization requires delimiters to identify the tokens in given document. Here, whitespace is taken as delimiter and identified each Kannada word separated by whitespace as a token in the document. Natural language tool kit (NLTK) library (RegexpTokenizer) is used for the tokenization [20].

3.2 Unicode Encoded Feature Extraction

3.2.1 Unicode Encoding

Unicode (Universal coded character set) is designed to overcome the incompatibility of ASCII encoding, for the non-English language characters. Encoding is the process of revamping the Unicode characters into the sequence of bytes. In contrast, decoding is the process of reversing a sequence of bytes into a set of Unicode characters. In Unicode encoding, unique integer called code-point is used to encode every character uniquely. Unicode is a standard encoding system and Unicode string consists of zero or series of code-points. Code-points are written in the form \uXXXX, where XXXX is the four-digit hexadecimal value and \u indicates that the value is Code-point. Hence, each character is assigned with a standard value called code-point in a string.

E.g.: Normal String (Kannada language): “ ಕನ್ನಡ”.

In programming, Unicode strings can be handled like normal strings. However, if Unicode strings are to be part of file operations like storing and retrieving or to display in terminal, it has to be encoded as a stream of bytes [19]. Hence to perform any file accessing operations (like Count-Vectorizing, Retrieving, etc.) on the raw text files containing natural languages should be encoded. Unicode transformation format (UTF) is a way to encode the code-points. Based on the size of the bits used to store the code-points, there are three variants of UTF and they are UTF-8 (8 bits for a Code-point), UTF-16 (16 bits for a Code-point), and UTF-32 (32 bits for a Code-point).

The machine transliteration to English language or Romanization can be achieved for the Kannada terms. Later, default ASCII encoding for these Romanized Kannada terms makes language processing tasks easy. But the Romanization process itself is computationally tedious, it requires both languages’ huge corpus and it has to maintain phonetics of source languages. Adding to this issue, if the normal text processing operations are performed directly on Kannada text then empirically we found that some agglutinative characters are faded in terms. This leads to the loss of valuable information of the features. Hence, we propose Unicode encoded number for each Kannada term which is not tedious, keeps term intact and not required additional language corpus. Encoded number can be decoded back to original term

Table 1 An example for Unicode string encoding and unique number generation

Characters	ಕ	ನ	್	ನ	ಡ
Unicode standard values	\u0c95	\u0ca8	\u0ccd	\u0ca8	\u0ca1
Encoding (UTF-16)	b'\xff\xfe\×95\x0c\xa8\x0c\xcd\x0c\xa8\x0c\xa1\x0c'				
Decimal representation	39085169431873023553583183 35				

later with same encoded byte address order either Big-endian or Little-endian byte order.

In experimentation, Kannada terms are Unicode encoded (UTF-16). Further encoded hexadecimal values are converted to decimal numbers to retain the Kannada term intact during various operations. These numbers are unique and represent unique terms in the document. From these decimal numbers, the original term can be retrieved back by decoding in same byte order. Table 1 shows how unique decimal number is generated for a Kannada term. Steps to attain unique number generation for terms are shown in Algorithm 1.

Algorithm 1: Unicode encoding and numerical representation of terms
Input: Kannada Text Documents $D = \{d_1, d_2, d_3, ...d_n\}$.

Data: $k =$ total number of tokens in a document, $x =$ total number of characters in a term, $n =$ total number of documents, $h =$ encoded sequence of hexadecimal code points of a term.

Output: Unicode encoded numerical representations of Kannada terms in documents.

STEP1: **for** $i \leftarrow$ to n by n do (each document in total documents).
STEP2: Tokenization and Preprocessing.
STEP3: **for** $j \leftarrow$ to k by k do (each term/token in a document).
STEP4: **for** $l \leftarrow$ to x by x do (each character).
STEP5: $(U + XXXX)^{l}_{i,j}$ Unicode character encoding based on system byteorder (UTF-16).
STEP6: **END** for each character.
STEP7: $z = 2 * len(h)$.
STEP8: $DecimalValue = \sum_{s=0}^{z} (h_s * 16^s)$.
STEP9: **END** for each term.
STEP10: Every term's numerical value is appended in a document with a space.
STEP11: **END** for each documents.

3.2.2 Feature Extraction

In text documents, the terms or words of the document are considered as features. These features represent the respective documents. Each term can be considered but

it will lead to high-computation time, excess memory and more over it may create more ambiguity for machine to make decisions. All documents will be consisting of discriminative terms or features and even it could be a combination of features which will be discriminative factor for a document. Hence judicious selection of terms is preferred instead of all terms [16]. If given document is (d_j), then $T = \{t_1, t_2, t_3, ...t_k\}$ where (T) represents the (k) unique terms representing the document (d_j) so $T \in d_j$.

After the feature extraction, stop words are removed from the documents. Removing the stop words reduces the feature size and enhances the efficiency of the learning algorithm. Stop words are removed in two ways either to have a dictionary of the stop words or remove based on frequency of terms across the documents. In KDC, stop words are removed based on the frequency of terms across the documents. The lower term frequency threshold is kept at 2 and higher term frequency threshold is kept at 1150, where these thresholds are considered based on empirical evaluations.

3.2.3 Vector Space Model

Before applying any machine learning classifier, the digital text should be represented by its features in a more compact and computationally appropriate form. From the survey, vector space model (VSM) is most established and well-known method for document-term weighting. In VSM, sequential order information is not maintained. Each term or feature is weighted by its count across the documents is called as term frequency (TF). Another statistical measure TF-IDF (Term Frequency-Inverse Document Frequency) is also used for term weighting [17]. TF-IDF weight represents the relevance of the term to a document and specifically IDF (Inverse Document Frequency) is a good index of a term in classification [18]. In the document (d_j), set of terms $T \in d_j$ and $T = \{t_1, t_2, t_3, ...t_k\}$ contains (k) unique terms. These unique terms represent the respective text document. Each (t_i) has weight (wt_i) and calculated using TF(d_j, t_i) or TF-IDF(d_j, t_i) statistical measure. TF-IDF consists of TF(t_i) and IDF(t_i).

$$\mathrm{IDF}(t_i) = \log\left(\frac{N}{d_f(t_i)}\right) \tag{1}$$

$$wt(t_i, d_j) = \mathrm{TF} * \mathrm{IDF} \tag{2}$$

In Eq. (1), N is the total number of documents and $d_f(t_i)$ is the frequency of documents containing term (t_i). In Eq. (2), $wt(t_i, d_j)$ represents the TF-IDF weight for each term. Later document (d_j) can be represented as a k-dimensional vector as shown in Eq. (3).

$$\overrightarrow{d_j} = (wt_1, wt_2, ...wt_k) \tag{3}$$

These k-dimensional vectors of all documents form the document-term matrix. Here, rows are documents and terms are represented in columns. Weights are the values assigned to the terms of the documents in the matrix. This document-term matrix is used further by machine learning algorithms for the classification task.

3.3 Classification

3.3.1 K-Nearest Neighbor (K-NN) Classifier

K-Nearest Neighbor (K-NN) classifier is a non-parametric method. This algorithm is an instance-based learning algorithm or lazy learning algorithm. It searches entire dataset for the 'K' most similar instances (documents) for the prediction of unseen data instance. Later among 'K' most similar instances, majority instance's label will be considered for the prediction. In Kannada Document Classification (KDC), the similarity between the instances is measured by one of the most widely used statistical measures called Euclidean distance. Equation (4) gives the Euclidean distance measure between instances.

$$\left|\overrightarrow{X} - \overrightarrow{Y}\right| = \sqrt{\sum_{i=1}^{k} (X_i - Y_i)^2} \tag{4}$$

In Eq. (4), $\overrightarrow{X}$ and $\overrightarrow{Y}$ are the document vectors and (k) represents the unique terms in the respective document vectors.

3.3.2 Support Vector Machine (SVM) Classifier

Support vector machine (SVM) is an efficient supervised algorithm for high-dimensional space classification task [9]. It is also used for regression tasks. This algorithm aims to find a hyper plane in an N-dimensional space which distinguishes the data points into its respective categories. Further, this hyper plane must have the maximum margin, i.e., maximum distance between the hyper plane and its closest data points of both classes. The data points nearer to the hyper plane are called as support vectors and they influence the orientation and position of the hyper plane. Maximum margin of hyper plane depends on these support vectors.

In linear SVM, distinguishing samples to their classes is done based on the output of linear function (Eq. 5). This linear function draws the plane between the classes. The optimization of positioning this plane in the sample space for better classification is the aim of the SVM classifier. As formerly discussed, margins helps in positioning this plane as far as possible from support vectors of classes. If there are two class and labeled as -1 and $+1$, then output of the linear function greater than '1' ($y \geq 1$) will

be identified with one class and if output is '−1' or lower ($y \leq -1$), then it belongs to another class. In Eqs. (6) and (7), this condition is represented in terms of vectors. Hence [−1, 1] pair acts like a margin between classes.

$$y = f\left(\overrightarrow{d_j}\right) = w_0 + \sum_{i=1}^{k} w_i t_i \tag{5}$$

From Eq. (5), (w_i) is the hyper plane parameter which is adjusted over the experiments in order to find the maximum margin between the support vectors. (w_0) is the bias parameter.

$$w^T x + b \geq +1 \text{ for class} + 1 \tag{6}$$

$$w^T x + b \leq -1 \text{ for class} - 1 \tag{7}$$

In Eqs. (6) and (7) w^T is the transpose of weight vector, x is the input feature vector of samples and b represent the bias. Equations (6) and (7) can be combined as $y\left(w^T x + b\right) \geq 1$ and hence the hyper planar equation becomes as shown in Eq. (8). Further to maintain maximum margin and less misclassification, i.e., for better generalization, weights are balanced with minimum error as shown in Eq. (9). From Eq. (9), C is the regularization parameter and ζ is the slack variable which stores the deviation of samples from margin.

$$y\left(w^T x + b\right) \geq 0 \tag{8}$$

$$\min \frac{\|w\|^2}{2} + C \sum_{n} \zeta^n \text{ with respect to } y\left(w^T x + b\right) \geq 1, \forall n \tag{9}$$

For the new test document (d_j) with (k) unique terms, if $f\left(\overrightarrow{d_j}\right) > 0$ then it belongs to a class else it belongs to other class in two class problem. Further, it will be same logic implied for the multi-class classifications also. If samples are not able to classify linearly then various Kernel tricks (e.g., polynomial, radial basis function, sigmoid) can be used, which intern transform the data into new space and eases the classification task.

Table 2 Details of Kannada documents dataset

Categories	Total number of documents
Space	40
Politics	95
Crime	55
Sports	60
Economics	50
Total	300

4 Experimental Results and Discussion

4.1 Kannada Documents Dataset

The Indian regional languages are less resourced [5]. Hence, there is a lack of publicly available standard datasets (in particularly for Kannada) for the various language processing tasks. For the Kannada document classification task, we have created our own Kannada documents dataset. This dataset consists of 300 Kannada documents; each document is comprised of two or three news articles. There are 5 categories like sports, politics, economics, space, and crime. Table 2 presents the details of the available documents at each category.

The articles in the documents are collected from the online publicly available news. Each document consists of more than 500 words and labeled with respect to its category. The source links are (a) https://www.kannadaprabha.com, (b) https://kannada.oneindia.com, and (c) https://kannada.asianetnews.com/india-news.

4.2 K-*Fold Cross Validation*

The cross validation techniques assess how the results of statistical analysis will generalize to an independent dataset. Cross validation address the problem of selection bias for training and testing division of samples and gives insight of model's ability of prediction for unknown data. In K-fold cross validation, given dataset samples are randomly divided into 'K' equal divisions. Among these divisions, $K - 1$ divisions will be taken for training and remaining will be used for testing the model. This process continues for 'K' number of experiments. In each experiment among all divisions one untested division will be part of testing. Finally, it generates result of each experiment and average accuracy of all experiments.

In KDC (Kannada Document Classification), the dataset is of multi-class labels hence the Stratified K-Fold cross validation is performed. In stratified K-Fold, the samples are divided into 'K' divisions unlike the normal K-Fold each division will have minimum representative samples from all labels. There will be 'K'

number of experimental results from which average accuracy is calculated. In this experimentation, we have considered 3, 5, and 10 values for '*K*'.

4.3 Experimentation

The KDC experiments are carried out on 300 Kannada text documents. This dataset is comprised of 5 different categories and documents are unevenly distributed among these categories. The preprocessing, tokenization and Unicode term encoding are performed as formerly mentioned. Hence, each unique term is represented with unique number. Each Unicode encoded term is given weights with respect to the term frequency (TF) and term frequency-inverse document frequency (TF-IDF) statistical measures. This forms the document-term matrix with term weights. Experiments are carried out on both weighting measures. Later encoded terms with the highest and the lowest frequencies (stop words) over the documents are removed. Further, stratified *K*-Fold train-test division is applied, where we consider 3, 5, and 10 folds for experimentation. For $K = 3$, among 300 documents 3 stratified splits of 100 documents are created and each split of 100 documents undergo testing in any of the 3 experiments. Thus, at each experiment, the train and test division ratio will be 70:30, where 200 documents for training and 100 for testing. Similarly for $K = 5$ it is 80:20 train-test division ratio and for $K = 10$ it is 90:10 train-test division ratio. K-NN and SVM classifiers are used for classification task. In case of K-NN classifier, the nearest neighbors 3 and 5 are considered for classifying the Kannada test document. In SVM, the linear kernel is used to classify the Kannada test documents. The *K*-Fold mean accuracies obtained for KDC experiments for both weighting measures are presented in Figs. 2 and 3.

Evaluation of the classification results of the classifiers is calculated by the accuracy metric. From Figs. 2 and 3, classification using SVM classifier does better in all the K-Fold experimentation. It attained mean accuracy of 98.67% for 3-Fold experiment ($K = 3$) with TF-IDF as term weighting. It comparatively performs

Fig. 2 *K*-Fold mean accuracies for TF based classification

Fig. 3 *K*-Fold mean accuracies for TF-IDF based classification

better in TF-IDF term weighting measure and attains 93.67% and 79.66% accuracies in 5-Fold and 10-Fold experimentations, respectively. Table 3 presents the SVM classifier's performance in *K*-Fold experiments.

Table 4 shows the performances of K-NN and SVM classifiers at the 3-Fold cross validation experiments. At 3-Fold experimentation, classifiers do better when compared to all other K-Fold experimentations. SVM attains 98.67%, K-NN ($n = 3$) attains 97.00% and K-NN ($n = 5$) attains 96.67% for TF-IDF Unicode encoded term weighting measure.

The combination of Unicode Encoded term's TF-IDF weighting measure and SVM classifier, in 3-Fold cross validation experiments performs better in KDC (Kannada Document Classification) for the newly presented dataset.

Table 3 Performance of SVM classifier

K-Fold cross validation	TF (accuracy) (%)	TF-IDF (accuracy) (%)
$K = 3$	98.30	**98.67**
$K = 5$	91.33	93.67
$K = 10$	75.33	79.66

Table 4 Performance of classifiers in 3-Fold cross validation

Machine learning classifiers	TF (accuracy) (%)	TF-IDF (accuracy) (%)
K-NN ($K = 3$)	82.67	97.00
K-NN ($K = 5$)	83.00	96.67
SVM (linear Kernel)	98.30	**98.67**

5 Conclusion and Future Work

Kannada is one of the Indian regional languages. As it is a low-resource language, it lacks in publicly available standard corpus for document text categorization. Hence, we have created a benchmark Kannada documents dataset consisting of 300 documents collected from online Kannada news articles. The document texts are in Kannada scripts. During experimentation, the Kannada terms are represented in Unicode encoded number because it aids in efficient, unambiguous and easier text processing task. These terms or features are represented in vector space model after the removal of term frequency based stop words. The conventional K-NN and SVM machine learning classifiers are used for document classification in K-Fold experiments. The SVM (linear) performs better when compared to K-NN classifier and gives 98.67% mean accuracy for the 3-Fold experiments, where term weighting is based on TF-IDF measure.

Here, the Kannada document classification is accomplished using supervised knowledge. Hence, still more research experiments are yet to be carried out under unsupervised condition. In using the similarity metrics, other than the Euclidean distance various measures like cosine similarity measures can be explored. In future, the corpus size should be increased to build complex and standard dataset. Analysis to be done for cohesive, complex and bigger datasets using better kernels, which handle the classification tasks efficiently. The various machine learning and neural classifiers can be explored using N-gram based features for large Kannada documents dataset. Deep learning networks handle complex and huge data efficiently. They concern over semantic issues, required for language understanding and are yet to be experimented on large Kannada dataset. In recent, pre-trained networks are developed and are used for arduous language tasks, which are well-trained with large corpus. These pre-trained models are to be explored on Kannada language corpuses for various language tasks.

Acknowledgements This work is supported by Vision Group on Science and Technology (VGST), Department of IT, BT and Science and Technology, Government of Karnataka, India [File No.: VGST/2019-20/GRD No.:850/397].

References

1. A. Dhar, N.S. Dash, K. Roy, Classification of bangla text documents based on inverse class frequency, *in 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)* (IEEE, 2018), pp 1–6
2. A. Dhar, N. Dash, K. Roy, Classification of text documents through distance measurement: An experiment with multi-domain bangla text documents, in *2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA)(Fall)* (IEEE, 2017), pp 1–6

3. S. Mohanty, P. Santi, R. Mishra, R. Mohapatra, S. Swain, Semantic based text classification using wordnets: Indian language perspective, in *Proceedings of the 3th International Global WordNet Conference, South Jeju Island, Korea*, (Citeseer, 2006), pp. 321–324
4. M. Tummalapalli, M. Chinnakotla, R. Mamidi, Towards better sentence classification for morphologically rich languages, in *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing*
5. A. Dhar, N.S. Dash, K. Roy, Categorization of bangla web text documents based on tf-idf-icf text analysis scheme, in *Annual Convention of the Computer Society of India* (Springer, 2018), pp. 477–484
6. P.K. Panigrahi, N. Bele, A review of recent advances in text mining of Indian languages. Int. J. Bus. Inf. Syst. **23**(2), 175–193 (2016)
7. S.A. Narhari, R. Shedge, Text categorization of Marathi documents using modified lingo, in *2017 International Conference on Advances in Computing, Communication and Control (ICAC3)* (IEEE, 2017), pp. 1–5
8. A. Dhar, N.S. Dash, K. Roy, An innovative method of feature extraction for text classification using part classifier, in *International Conference on Information, Communication and Computing Technology* (Springer, 2018), pp. 131–138
9. L. Wang, *Support vector machines: Theory and applications*, vol. 177. (Springer Science & Business Media, 2005)
10. M. Tummalapalli, R. Mamidi, Syllables for sentence classification in morphologically rich languages, in *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation* (2018)
11. K. Rajan, V. Ramalingam, M. Ganesan, S. Palanivel, B. Palaniappan, Automatic classification of Tamil documents using vector space model and artificial neural network. Expert Syst. Appl. **36**(8), 10914–10918 (2009)
12. R. Jayashree, K. Srikantamurthy, B.S. Anami, Sentence level text classification in the Kannada language—A classifier's perspective. Int. J. Comput. Vis. Robot. **5**(3), 254–270 (2015)
13. Puri, S., Singh, S.P., An efficient Hindi text classification model using SVM, in *Computing and Network Sustainability* (Springer, 2019), pp. 227–237
14. B.S. Harish, D.S. Guru, S. Manjunath, Representation and classification of text documents: A brief review. IJCA, Spec. Issue RTIPPR **2**, 110–119 (2010)
15. J.J. Webster, C. Kit, Tokenization as the initial phase in nlp, in *COLING 1992 Volume 4: The 15th International Conference on Computational Linguistics* (1992)
16. M. Revanasiddappa, B. Harish, A new feature selection method based on intuitionistic fuzzy entropy to categorize text documents. IJIMAI **5**(3), 106–117 (2018)
17. G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval. Inf. Process. Manage. **24**(5), 513–523 (1988)
18. T. Tokunaga, I. Makoto, Text categorization based on weighted inverse document frequency, in *Special Interest Groups and Information Process Society of Japan (SIG-IPSJ, Citeseer* (1994)
19. S. Bird, E. Klein, E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit (O'Reilly Media, Inc., 2009)
20. Project N (2020) https://www.nltk.org/_modules/nltk/tokenize/regexp.html. Last updated on 13 Apr 2020

Determining Stock Market Anomalies by Using Optimized *z*-Score Technique on Clusters Obtained from *K*-Means

Bibek Kumar Sardar, S. Pavithra, H. A. Sanjay, and Prasanta Gogoi

Abstract *K*-means has always been the most efficient technique to detect anomalies on any kind of dataset. It would be interesting to explore whether the algorithm could do marvels when used on a stock market dataset. Given the state-of-the-art methodologies, stock market data is prevailingly the most challenging data to work on, since the data values increase at a fast pace. Additionally, data analysis performed on time series data, taken from stock markets, has gained lot of popularity in recent past. Identification of any kind of anomaly in such dataset could be compelling; since this information can pave the way of growth for companies and investors hoping for higher returns and higher profits at lower risk. The manuscript aims to facilitate detection of such volatility by ascertaining outliers in the stock market data, without any prior knowledge of possible abnormalities. Though, *z*-score has been used extensively for determining deviations associated with data values for a given distribution, we strive to formulate a similar scoring formula, dev-score, that computes deviation for two-dimensional data (can be extended for more than 2D data as well), after generating clusters using *K*-means. The manuscript plots clusters for stock market data and identifies those stocks that deviate from their normal value on a particular trading day. It is important to note that the deviations are computed only for specific features of stock market data (volume and fluctuations), and this model can be easily extended on large number of features.

Keywords Machine learning · Anomaly detection · *K*-means clustering algorithm · *z*-score

1 Introduction

Dissemination of scientific knowledge, across the researcher and practitioners plays a key role in technology advancement in both industry and academia. Machine learning has become an important subject of exploration and an unavoidable tool to solve problems related to classification, clustering, and regression. Machine learning

B. K. Sardar (✉) · S. Pavithra · H. A. Sanjay · P. Gogoi
Nitte Meenakshi Institute of Technology, Yelahanka, Bangalore, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_32

domain basically is a subset of a larger domain—Artificial Intelligence that provides machines an impeccable ability to train, evaluate, and later improve its experience with the help of explicit programming. Around the globe, there have been an infinite number of articles which have explored powerful machine learning tools and their usage in solving problems of different domains. Essentially, machine learning domain solves problems of three broad categories that are designated under the umbrella of supervised, unsupervised, and reinforcement learning algorithms. Supervised algorithms make use of labeled dataset, observe the under lying patterns in these datasets, and predict a new sample by making use of the discovered patterns. Alternatively, unsupervised machine learning algorithms does not need any class-labels from training samples. These algorithms identify a mathematical relationship across the features and solve problem associated with regression and generation of clusters. On the other hand, reinforcement algorithms interact with their environment, predict future actions to solve the problem in hand, and calculate errors at every step and simultaneously, assigning reward points to these actions.

Anomaly, also known as abnormality, irregularity, or inconsistency, is a value that has deviated from the normal trend. An anomalous data can be a real threat leading to inconsistent results and could possibly raise multiple challenges in domains like statistics, data mining, signal processing, finance, and networking. The major reasons behind the existence of anomalies on dataset are measurement errors, data entry errors, data processing errors, experimental errors, and sampling errors. Alternatively, anomalies, also known as outliers, are detected by a method popularly known as outlier detection. Outlier detection is the method of discovering fluctuation in a given set of data. The method has a large field of application in various domains; some examples are fault detection, fraud detection and event system, health monitoring detection systems used in computer networks, sensor networks, and so on. The importance of anomaly detection is to enable us to find out outliers without knowing characteristics of possible abnormalities. In machine learning, it includes many classical methods of anomaly detection; some examples are isolation forest, support vector machine, and so on.

A very popular unsupervised algorithm, K-means is often used for detecting outliers from noisy datasets. Internally, it generate group of data points in such a way that points that are similar fall in the same group. Alternatively, points that are different are assigned to different group or cluster. In order to compute similarity between data points, different ways are suggested by different scientists over the period of time. The most popular choice of computing cluster similarity is by measuring Euclidean distance. Other methods suggested and used by researchers to generate clusters using K-means are cosine similarity, Manhattan distance, Mahalanobis distance, etc.

In this manuscript, we aim to build a mechanism that brings out anomalous data points in an intractable stock market data. Volatility of stock markets is always a cause of concern for many investors and traders who have invested a huge amount of capital in stock market and a single mistake may incur heavy monetary loss to them. Our focus is to identify and work on parameters which could not only indicate disparity in current stock market data but also predict information about divergences

that may happen in future. TO achieve this, we took a synthetic dataset on stock market, performed K-means to recognize the center mass of the cluster (centroids) and computed dev-score (enhanced z-score) on every data point to identify how far the point is lying from the center mass. If the dev-score of a point is above a threshold, the point is declared as an outlier.

The manuscript is organized as follows. We begin by exploring the previous research conducted in analyzing the effectivity of K-means in anomaly detection. The section (section II) also examines the functional ability of z-score in computing the divergence associated with any data distribution in hand. Section III explains the methodology followed by us in determining anomalies for stock market data. It explores the working of K-means and set down a platform where enhanced z-score formula (termed as dev-score) helps in determining deviating values in stock market fluctuations during trading hours. Next section (Section IV) unveils the complete experimentation done on synthetic stock market data and demonstrates the effectivity of the methodology used by us. We end our manuscript with conclusion section.

2 Literature Review

A lot of research has been conducted for detecting anomalies in variety of dataset. Most of these methods are found to be complex, robust, and capable of providing results on smaller datasets. Mohammad Saeid Mahdavinejad [1], presented in their research how machine learning has helped in providing solutions for the challenges thrown by smart IoT applications. Thudumu et al. [2] have worked on anomaly detection which, according to the authors, is a process of identification of rare items, events, and observation seen in the dataset. Kotu and Deshpande [3] submitted paper work on outlier detection method of finding abnormal behavior in a given record. According to the authors, anomalies are the fluctuations among the normal behavior which may not align with the desired output from the task associated with a given dataset. These anomalies could create undesirable results and may lead to difficulties in obtaining the expected output. Authors further explored that anomaly detection inherently deals with machine learning domain, statistical modeling, k-means clustering, automation, cloud, big data, regression, and clustering. Lei [4] wrote an article on anomaly detection method that identifies outliers by virtue of the neighborhood information to understand the techniques of unsupervised learning. A recent research conducted by Duan et al. [5] indicates that work done by using K-means by incorporating cluster-based, distance-based, and statistical-based mechanism not only generates efficient clusters but also eliminates anomalies from the noisy dataset.

Furthermore, exploring on the same direction to evaluate the efficacy of clustering based methods for detecting outliers, Knapp and Langil [6] suggested a policy-based detection while providing rule-less methods to find abnormalities. At industry level, strong defense acquired by accurate behavior can be calculated, and abnormal behavior can be detected. The authors laid stress on the working principle of the

industrial network system and suggested that network behavior is predictable by applying effective detection techniques, and outliers can be determined.

Cardenas and Safavi-Naini [7] showed in their research that anomaly detection, based on network security community is being used since decades. Although they have ability to detect zero-day attack, novel attack, or previously unknown attack, there is always a probability of generating from false alarm (ability to detect false abnormalities). Similarly, Jyothsna and Prasad [8] submitted a survey on anomaly detection and conveyed that, with the invention of anomaly based detection, various techniques were developed in past to track zero-day attack as well as previously unknown attacks (novel attacks). They claimed that maximum detection rate and minimum alarm rate up to 1% can be obtained by applying these methods. Instead of using anomaly-based detection methods, they followed signature-based detection for mainstream implementation of outliers' detection. It was difficult to compare the variety of detection methods purely on the basis of their strengths and weaknesses. Additionally, industries do not approve of using anomaly-based detection by validating the efficiency of all these techniques. To find an optimal solution, the authors survey and reviewed the present state of the trial in the field of outlier-based detection. The research paper presented a detailed analysis of drawbacks, limitation, and future scope of all the recent anomaly detection methods.

Denning et al. [9] submitted an article which is based on a real-time detection expert system model that has the ability to detect penetrations, break-ins, and many forms of computer misuse. The algorithm is based on supposition which breaks the network security can be found out by supervising the dataset for abnormal behavior that system usages. The model acquires knowledge about the abnormal behavior of the dataset with respect to features value in terms of mathematical calculation such as standard deviation, z-score, Euclidean distance, statistical model, or mathematical model to audit records and detect fluctuations on normal behavior. The algorithm is independent for different frameworks, environment, application, and system. In this paper, the general purpose of finding abnormal behavior in given audit records detect fluctuation on normal behavior.

Siddharth Misra et al. [10], wrote an article on detecting outliers for geophysical survey and well logs. Geophysical survey is a collection of geophysical data that has many applications in geology, archaeology, mineral, energy exploration, and oceanography. Well logs is a collection of formation of particle records while drilling wells, and basically it generates an alarm of any irregular activities observed during the drilling of wells. Alongside, it is difficult process, consumes lot of time, requires lot of human labor, and dedicated human expertise is required to get important dataset for well logs. So, to get the expected dataset, authors used ab unsupervised machine learning outlier detection. Oghenekaro Osogbo et al. [10] proposed an outlier detection which is used to find and collect the unusual dataset like geology, archaeology, mineral, and energy exploration. Outlier detection can be found using simple methods such as box plot and z-score. A box plot represents the distribution of samples corresponding to various features using boxes and whiskers, Mark Powers et al. [10] suggested usage of z-score for outlier detection that computes distance between data samples and its mean (μ). It can be interpreted as: if z-score value is

greater than or less than + 3 or − 3, respectively, the data point can be marked as outlier. z-score is expressed as,

$$z - \text{score} = \frac{x - \mu}{\sigma} \tag{1}$$

In the present paper, we perform anomaly detection for network industries and where he has discussed major problems in detecting irregular activities or issues mainly in network industries. Usman Habib, G. Zucker, M. Blochle, F. Judex, and J. Haase et al. [11] explored on the anomalous data and found that a huge amount of data is recorded by checking various building structures using sensors. These sensors will note down some physical properties of real-life phenomena, for example, temperature, weight, and stream rate. Sensors sometimes give wrong dataset due to weather fault or fault measurement due to defect sensors and hence that leads to various mistakes/errors in the dataset. Hence, before considering the data to obtain some important information, it is important to approve it. So the detection of outlines helps in removing the data that is normal or expected behavior of the machine. S. Katipamula et al. [12] proposed paper is to create a method that can automatically find anomalies in data but checking the data is labor intensive and not achievable. This paper talks about the means required for detecting outliers in the information got from retention chillers utilizing these on/off state information. It explains methods for automatic detection of on/off and/or missing data status of chillers. The initial step of anomaly detection on/off and missing information state is finished by utilizing a two-layer K-mean clustering algorithm calculation. The missing information state will speak to when there is no information recorded for any sensor of the chiller. After discovery of the on/off express, the cycle based z-score standardization is utilized with a clustering calculation to discover the anomalies in the information. In the present paper, we find anomaly detection for a huge database. K-mean clustering algorithm and z-score is used to find anomaly detection.

Shashank Singh Yadav et al. [13] wrote an article on one of the major problems of today's life which is heavy traffic that is caused by an expansion in the quantity of vehicles and substantial traffic needs consistent observation for guaranteeing security of individuals. The very normal reason for traffic is people. Some of the traffic violations are use of mobile phones while driving, signal jumping, drunken driving, not wearing seat belts and helmets, etc. So, video-based traffic observation is the solution considered in this paper. Hence, the aim of this paper is to find anomalies using traffic video surveillance. Anomalies include use of mobile phones while driving, signal jumping, drunken driving, not wearing seat belts and helmets, wrong U-turn, etc. So that traffic can be controlled, and if any accidents take place, ambulances are sent to the accident place quickly. Unsupervised machine learning is used for detecting traffic anomalies. Video recordings are taken as input data, and these recordings are changed into frames from each frame into highlights and from these extracted highlights, a single centroid point is taken with reference to the centroid point from this category of traffic validation taken. Ashok Kumar et al. [14], proposed K-mean clustering which is used to find irregular activities and define them as anomalies.

3 Proposed Methodology for Anomaly Detection

Inspired from the work done by different researchers across various domains for detecting anomalous behavior, we intent to utilize the core idea behind K-means and z-score and introduce a more precise and expressive formula with the name, dev-score. Dev-score, or more formally, deviation score, aims to compute divergence or deflection of data point from its mean in a given distribution of data. The basic idea used in the manuscript involves execution of K-means for identification of mean (also called centroids). Once the centroids are identified, we utilize dev-score to figure out all such points that deviate from centroid beyond a certain threshold, λ. These points are referred as anomalies. Interestingly, computation of dev-score normalizes the distance of every data points from its mean. Dev-score is inspired from the popular method, z-score (extreme value analysis) which is essentially a parametric method for computing outliers. Formally, z-score is represented as:

$$\mathrm{z-score} = (\mathrm{X} - \mu)/(\mathrm{std}) \tag{2}$$

Here, μ (mean) and std (standard deviation) are computed on a single feature value using standard formula. The essence of z-score is it normalizes value of entire dataset into one specific range, eliminating the effect of outliers in the data, if any. Based on the same formula, the extended version of z-score is termed as dev-score. Dev-score works well for two or more that two-dimensional data. It is given as:

$$\mathrm{dev-score} = \frac{D}{\mathrm{std}_C} \tag{3}$$

where

$$D = \sqrt{(x1 - c1)^2 + (x2 - c2)^2}$$

Here, D is the Euclidean formula for distance computation, that computes distance between X and C in two-dimensional space. Dev-score further computes deviation score of every X, by dividing distance (D) with standard deviation (std_C) for cluster, C. It is important to mention here that dev-score is effectively doing the same thing as z-score, with the difference that it generates a normalized score for data points in higher-dimensional space comprising of two or more than two features.

***K*-means clustering algorithm**—is an unsupervised machine learning method that utilizes the inherent information observed in the dataset to generate meaningful clusters. The algorithm formally divides 'n' data points into k-cluster that follows three essential steps of initialization, assignment, and update. In initialization, k initial centroids (means) are picked-up randomly from the dataset. Next step assigns every data point to the nearest cluster by computing the Euclidean distance of the point and centroids. Once all the points are assigned to nearest cluster, the last step

updates centroids of the every cluster. This process continues till algorithm reaches convergence and centroids do not change any further.

For computing distances between sample point and the centroid of the cluster, Euclidean distance is used that is represented as:

$$\text{Euclidean distance } d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{4}$$

where x_1, x_1…..are the two coordinates of data point X, in two-dimensional space.

3.1 K-*mean Clustering Algorithm*

Figure 1 depicts formation of four clusters from the available data points in two-dimensional space. The algorithm runs iteratively over each point and assigns it to the nearest cluster. The three basic steps are as follows:

Step 1: Initialization
K initial "means" (centroids) are randomly generated as first step.

Step 2: Assignment
Each sample is assigned to one cluster depending upon its closeness to the centroid. This closeness is evaluated by computing Euclidean distance.

Step 3: Update
The mean of the centroid is calculated by taking mean of all the points in that cluster.

Step 2 and Step 3 are repeated iteratively until convergence is observed. By convergence, we mean that the centroid obtained from previous iteration does not change. After convergence of the algorithm, the points are assigned to each cluster and the sum of squared error is minimized between points and their respective centroids. The

Fig. 1 Illustration of clusters seen on the scatter plot of data points

best clustering solution for the given data points is the one picked after repeating the above three steps for ten times.

More detailed explanation of *K*-mean clustering method: Here are some finer details and aspects of the algorithm. These aspects are discussed on the basis of three basic steps:

1. While choosing the initial centroid values—Once, the decision is made on number of groups/clusters we want to have, there are few options for picking initial centroid values. Select k random points from dataset and run *K*-means on it. Repeat the process n times and choose the iteration that has minimum least square error. This process is far more reasonable and accurate. We have used this model during executions. Specifically, the number of random initializations to perform *K*-means clustering model is achieved using sci-kit learn package from Python library.
2. While assigning data points to a cluster—Once the selection of initial centroids is done, the next step is to assign each data point to a cluster. Mathematically, this can be represented as (for *i*th iteration):

$$S_i^{(t+1)} = \left\{ x_p : \| x_p - m_i^{(t+1)} \square^2 \leq \| x_p - m_j^{(t+1)} \square^2 \ \forall j, 1 \leq j \leq k \right\} \quad (5)$$

   where $S^{(t)}$ indicates *t*th cluster, $m^{(t)}$ is the mean, and x_p is the data point. Here, if the distance from the data point (x_p) to the mean (m_i) is smaller than distance to the mean of all other centroids (m_j), then x_p is assigned to $S^{(t)}$.
3. While reassigning data points to new clusters—As soon as all points are allocated to a cluster, there is a need to calculate new centroid values and to reassign data points until the process converges on a cluster segmentation that stops changing. The mean is computed using the formula below

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \in S_i^{(t)}} x_j \quad (6)$$

4 Experimental Result and Discussion

This section explores the experimentation done using *K*-means and also brings forth the admirable results through this method. In order to evaluate the efficacy of the algorithm for bringing out anomalies, we have used synthetic stock market dataset. Working of *K*-means is explored in the previous section, now we present a brief explanation of the stock exchange data and also put across the working process.

A stock market is a place where equities of company are bought or sold, by different buyers and sellers, in order to maintain ownership on companies or businesses. Buying and selling of shares is done at stock exchanges, which internally maintains list of companies who intent to participate in trading of equities. Buying

and selling are formally called as trading which involves transfer of stocks (also known as shares and equities) from seller to buyer after both parties officially agrees upon a price. This is reflected as stock price of a company. The stock exchange not only facilitates trading of securities across companies (or individual investors) but also maintains the entire trading data that gets generated as a result of everyday transactions. It involves the names of companies that are listed for trading, their symbols, stock price, trading volume, opening price, closing price, and highest and the lowest price in a day (or any other time frame). Exchanges also create and maintain charts that indicate movement of stock price for various companies. Generation of charts and the after-market analysis of trading data helps prospective investors in making decisions related to buying (or selling) of equities and thus play a creditable role in making future profits. Let us take a closer look at the stock market dataset used in the model.

Figure 2 shows the input data with distinct features and values with respect to date of transaction of equities. Open, close, high, low, volume, adj_close, and fluctuation are the feature values associated with each listed company stocks. There are 5990 stock values used as dataset for experimental reasons.

- Symbol—The symbol column represents the company name and the symbol of stock assigned.
- Date—The date on which the company's stock prices are observed.
- High—The high column represents the highest price on that date.

1	symbol	date	open	high	low	close	volume	adj_close	Fluctuation	f2	f3
2	A	11/30/201	45.76	45.82	45.38	45.56	1215100	45.56	0.438975	0.439999	0.19999
3	AA	11/30/201	28.97	29.378	28.01	28.08	2651500	28.08	3.169512	1.368	0.88999
4	AAAP	11/30/201	26.73	26.92	26.6	26.76	42700	26.76	0.112108	0.32	-0.0
5	AAC	11/30/201	6.88	7.508	6.83	7.24	199100	7.24	4.972376	0.678	-0.3
6	AAL	11/30/201	47.42	47.66	46.47	46.69	4494100	46.69	1.563502	1.189999	0.72999
7	AANC	11/30/201	49.1	53.5	45.62	53.5	17900	53.5	8.224303	7.880001	-4.
8	AANE	11/30/201	3.95	4.2	3.95	4.1	50400	4.1	3.658537	0.25	-0.1
9	AAN	11/30/201	32.07	32.29	31.52	31.99	431100	31.99	0.250078	0.770001	0.0
10	AAOI	11/30/201	24.25	24.25	23.21	23.44	197100	23.44	3.455627	1.040001	0.80999
11	AAON	11/30/201	33.75	33.75	32.9	33.05	69800	33.05	2.118006	0.849998	0.70000
12	AAP	11/30/201	171.32	172	168.6	169.12	489300	169.12	1.300859	3.399994	2.20001
13	AAPC	11/30/201	9.84	9.84	9.84	9.84	0	9.84	0	0	
14	AAPL	11/30/201	116.65	117.2	115.43	115.82	30253100	115.82	0.716631	1.769997	0.83000
15	AAT	11/30/201	42.41	43.34	42.41	43.08	231300	43.08	1.555251	0.93	-0.6
16	AAU	11/30/201	1.08	1.08	0.97	0.97	452400	0.97	11.34021	0.11	0.1
17	AAV	11/30/201	6.7	6.85	6.7	6.75	66900	6.75	0.740741	0.15	-0.0
18	AAWW	11/30/201	52.3	52.85	51.3	52.15	140500	52.15	0.287626	1.549999	0.14999
19	AAXI	11/30/201	55.28	55.33	54.84	54.93	860300	54.93	0.637173	0.490002	0.34999
20	AB	11/30/201	23.8	23.85	23.42	23.45	218300	23.45	1.492529	0.43	0.34999

Fig. 2 Stock market dataset

- Low—The low column is lowest price on that day.
- Volume—The volume column shows the number of shares traded on that particular date.
- Open—The open column represents the price at which the stock was opened for trading.
- Close—This column indicates the close price at which the stock stopped trading.
- Fluctuations—This column tells about the prices fluctuation due to the difference between profit and loss.
- Dev-score: dev-score column represents the mathematical calculation of deviation from its centroids with respect to standard deviation.
- Distance: Euclidean distance, distance from centroids to a point.
- Cluster0 and cluster1: The column cluster0 and cluster1 represent the two different clusters formed due to K-means for given dataset.
- Volume: This column represents the volume of shares stock trading on particular date.

Working: With all the feature values listed and shown above, we took two critical parameters to highlight anomalous deviations in stock trading, volume, and fluctuations. For a particular company's stock, the deviation seen in the trading volume decides stability or volatility of market and the company. It is an indicator of how large the stock price may vary around its mean thus gauging the risk associated with the equity. This implies that if large fluctuations are observed, the stock price is considered more volatile and risk associated with it is high. In order to conduct the proposed work to detect anomalous variation in the trading volume and fluctuation on a specific day, following steps are followed-

Step 1: Clusters are generated using K-means for the entire dataset, using two parameters, volume, and fluctuations. Figure 3 shows the formation of 2 clusters with their respective centroids marked with star.

Step 2: Store the centroids of the two clusters for computing distance from every point to its centroid.

Step 3: Compute the distance of data point (X) from the two centroids. Let $C1$ and $C2$ be the two centroids obtained after the convergence of K-means on the stock market data. To show calculations, assume $\times$ 1 and $\times$ 2 to be the two-dimensional representation of X, and c1 and c2 are the two-dimensional representation of C1 (with assumption that points X belongs to cluster C1), then

$$D = \sqrt{(x1 - c1)^2 + (x2 - c2)^2}$$

Step 4: Divide distance D with standard deviation of the cluster. This process generates dev-score of every data point present in that cluster. This is a reflection of the extent to which a point is deviated from the center of the cluster.

$$dev - score = \frac{D}{std_{C1}}$$

Fig. 3 Flowchart of the methodology used

where

dev-score	is the deviation score for point X.
D	is the Euclidean distance of X from C1.
std_{C1}	is the standard deviation computed for C1.

Step 5: Compare the dev-score with the pre-computed threshold λ, if the dev-score is greater than λ, the data point is declared as anomaly.

The essence of the method used by authors is to find out all points that are lying farthest from the center of the two clusters. A threshold λ, is computed by identifying all those points which are lying 2-standard deviation away from centroids. It is statistically shown that the data points that are seen with 2 or more standard deviation away from mean are considered isolated from the rest of the data. Once these points are identified, mean of their *z*-scores is computed and used as threshold. The λ was observed to be 2.0 for the given dataset. Implementation of *K*-means and dev-score is done using Python 3.6 installed on Intel Core-i3 8145U CPU with 8 GB RAM. Figure 4 depicts the clusters and centroids emerged from the stock market dataset (Table 1).

Fig. 4 Cluster outliers

Table 1 Output of *K*-means

	Centroid position	
	Volume	fluctuation
Cluster 0	455,940.01	1.36
Cluster1	10,746,971.05	2.35

4.1 Analysis of the Output Dataset

As indicated, we have used 5990 points for stock market dataset. The following figure (Fig. 4) shows formation of two clusters as a result of running *K*-means. In this figure, volume and fluctuation are features used in the algorithm for generating two clusters indicated as 0 and 1. The blue and yellow dots indicate data points that belong to two different clusters. The star sign indicates their centroids. Clusters which are farthest from the centroids are detected as outliers. Once clusters appear, dev-score is computed for all the points present in two clusters. If this score is larger than the threshold set for each cluster, the point is labeled as anomaly. After running these steps on 5990 data points, we observed 5866 points in cluster 0 (shown in blue dots) and 114 points in cluster 1 (yellow points). After the computation of dev-score on each of these points, we could see 97 outliers for cluster 0 and 2 outliers for cluster 1.

In Fig. 5, we presented a screenshot of the output and observe two clusters identified as '0' and '1' for various stocks mentioned in 'SYMBOL', We also observe the value of dev-score for every company, for the companies named 'ABX' and 'FIT', the dev-score value of larger than 2.0. Hence these are declared as outliers for cluster 0 and for cluster 1.

SYMBOL	volume	Fluctuation	cluster	dev-score	distance
AAPL	2293100	0.164068909	0	1.1432529	1609223.9
ABT	128300	1.859678783	0	0.394702	555576.05
ABX	3579700	0.541147602	0	2.0573017	2895823.9
AKX	679600	0.227268182	0	0.0030379	4276.0535
AUY	9800	0.81300813	0	0.4788888	674076.05
BBRY	409800	8.690818457	0	0.1947139	274076.05
C	446200	0.71942446	0	0.168854	237676.05
CEMP	28509500	0.301333615	1	0.4117203	5764081.7
COG	16284700	0.197238659	1	0.4614801	6460718.3
CSCO	18578900	2.379493739	1	0.2976086	4166518.3
F	37662200	0.778210117	1	1.0654849	14916782
FB	15295900	1.138519924	1	0.5321087	7449518.3
FCX	13185800	1.314965519	1	0.6828302	9559618.3
FIT	55097900	0.087489064	1	2.3108926	32352482

Fig. 5 Intermediate results after computing dev-score

5 Conclusion

In this manuscript, we reveal a refined method to detect anomaly by performing *K*-means on the stock market dataset and later using the centroids generated from K-means for computing a score to detect outliers. We termed the score as dev-score. Additionally, multiple research papers from various domains of machine learning, big data, cloud, automation, data analytics, and data analytics with python were explored thoroughly to evaluate the essence of using standardization techniques like *z*-score for computing outliers. It is observed that *z*-score calculates the deviation from its respective centroids, and there is a need to use Euclidean distance and standard deviation in order to appropriately use *z*-score. Hence, dev-score becomes convenient. A *K*-mean clustering algorithm is presented in the manuscript in detail. Results obtained from this method are found to be appealing and easily extensible on any kind of time-series data.

References

1. M.S. Mahdavinejad, M. Rezvan, M. Barekatain, *Machine Learning for Internet of Things Data Analysis: A Survey*, vol. 4 (University of Isfahan, Iran Kno.e.sis—Wright State University, USA)
2. S. Thudumu, P. Branch, J. Jin et al., A comprehensive survey of anomaly detection techniques for high dimensional big data. J Big Data **7**, 42 (2020). https://doi.org/10.1186/s40537-020-00320-x
3. V. Kotu, B. Deshpande, *Data Science: Concepts and Practice Published in Cambridge* (Morgan Kaufmann Publishers, MA, 2019)
4. P. Lei, A framework for anomaly detection in maritime trajectory behavior. Knowl Inf Syst **47**, 189–214 (2016). https://doi.org/10.1007/s10115-015-0845-4
5. L. Duan, L. Xu, Y. Liu, J. Lee, Cluster-based outlier detection. Annals Oper. Res. **168**, 151–168 (2009). https://doi.org/10.1007/s10479-008-0371-9
6. E.D. Knapp, J.T. Langill, *Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, Scada, and Other Industrial Control Systems* (Waltham, MA: Elsevier, 2015)
7. A.A. Cárdenas, R. Safavi-Naini, *Handbook on Securing Cyber-Physical Critical Infrastructure* (2012). https://doi.org/10.1016/B978-0-12-415815-3.00025-X
8. V. Jyothsna, R.V.V. Prasad, A review of anomaly based intrusion detection systems. Int. J. Comput. Appl. **28**(7), 26–35 (August 2011)
9. D.E. Denning, An intrusion-detection model, in *IEEE Transactions on Software Engineering* (vol. SE-13, issue no. 2, pp. 222–232) (1987). https://doi.org/10.1109/TSE.1987.232894
10. S. Misra, O. Osogba, M. Powers, Unsupervised outlier detection techniques for well logs and geophysical data, in *Machine learning for surface characterization* (Texas A&M University, College Station, TX, United States, 2020). https://doi.org/10.1016/B978-0-12-817736-5.00001-6
11. U. Habib, G. Zucker, M. Blochle, F. Judex, J. Haase, Outliers detection method using clustering in buildings data, in *IECON 2015—41st Annual Conference of the IEEE Industrial Electronics Society*, Yokohama (2015, pp. 000694-000700). https://doi.org/10.1109/IECON.2015.7392181
12. S. Katipamula, M.R. Brambley, Review article: methods for fault detection diagnostics and prognostics for building systems—a review Part I. HVACR Res. **11**(1), 3–25 (2005)
13. S. S. Yadav, V. Vijayakumar, J. Athanesious, Detection of anomalies in traffic scene surveillance, *in 2018 Tenth International Conference on Advanced Computing (ICoAC)* (Chennai, India, 2018, pp. 286–291). https://doi.org/10.1109/ICoAC44903.2018.8939111
14. P.M. Ashok Kumar, V. Sathya, V. Vaidehi, Traffic rule violation detection in traffic video surveillance. Int. J. Comput. Sci. Electron. Eng. (IJCSEE) **3**(4) (2015)

Data-Driven Strategies Recommendation for Creating MOOCs for Effective Online Learning Experience

Tanay Pratap and Sanjay Singh

Abstract This paper presents the strategies for creating an online course from the learner's perspective. The paper examines the learners' requirements for signing up and completing an online course. Two hundred eleven online learners provided their responses to a survey request. It had an open-ended verbatim question. The responses were categorized and then re-categorized to get the most common expectation area. The survey found out that the learners care most about having real-life projects, the instructor's delivery style, beginner friendliness of the course, and understanding its purpose in the overall context. In each of these areas, there are suggestions and expectations with high-frequency data to work. This survey was done to an open group of people from across the world with an open-ended question. There was no control on the demographics of the learners. Hence, the next stage of the research should be on controlled demographics as the online courses generally are for a niche area. The findings are useful for online instructors and MOOC platforms alike. With data and computing resources becoming ubiquitous, many more teachers will create courses for online consumption. This paper can thus serve as a checklist for instructors creating such new courses.

Keywords Online learning · Lifelong learning · Pedagogical issues · Social media · Teaching/learning strategies

1 Introduction

While traditional classroom education has been ongoing for quite a few centuries, massive open online course (MOOCs) has just arrived on that timeline. It is far-reaching in the aspect of the opportunities it presents, and everyone from teachers,

T. Pratap
Microsoft India, Bengaluru, India
e-mail: tapratap@microsoft.com

S. Singh (✉)
Department of Information and Communication Technology, Manipal Institute of Technology, MAHE, Manipal 576104, India
e-mail: sanjay.singh@manipal.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_33

students to big-time universities recognize it. As per recent numbers from classcentral.com [19], which lists all the available MOOCs across the internet, 900 universities have a MOOC offering 13.5 K courses enrolling close to 110 M students.

Even though the scale of MOOCs is massive, the completion rate is still alarmingly low. There are multiple reasons for it, but one thing is sure that the key to delivering the perfect online course has not been cracked yet. MOOCs are trying to put traditional classroom on the internet, which could be annoying to the students as even though they get the teachers, they are not getting the peers and support system of a classroom.

This paper aims to relook at MOOCs' best practices from the one that matters the most: the students. We conducted an online survey with a relatively emotional and open-ended question about online courses. Over a decade of MOOCs, most online learners have figured out what works for them and what not. This paper tries to learn from the students and put those best practices for future MOOC creators.

2 Related Work

Work on Online learning, pre-MOOC era: Before MOOCs came into the picture, training and learning were done online. There is some good work in the area of online learning and teaching from the pre-MOOCs era.

Gilly Salmon's [18] work talks about creating an online tool first, then creating an online cohort, and moving on to information exchange. Only in the fourth phase does actual knowledge construction happen to lead to final development. If appropriately applied, these ideas would help in student retention in MOOCs; even though the tools are as old as email, ideas are quite robust.

Terry Anderson [3] focuses on the role of a teacher. This study stresses that the teacher should add his personality in online learning material and "personal reflections, anecdotes, and discussions of teacher's struggles and success." Students have found this inspirational and motivating. It then delves deep into discourse facilitation, assessment, and ideas to build direct instruction provision. This study is beneficial for a teacher looking to expand online and moderate a course. The study concludes with the qualities of e-teacher, pointing that "Internet efficacy" must be effective.

Work related to MOOCs: Oakley and Sejnowski [16] discuss the learnings from the world's most popular MOOC. It stresses heavily on using green screen and video editing techniques for creating content. It goes on to describe the teleprompter and studio lights and everything in the production, of course. Moreover, it is also covering what plays essential roles. It says that instructor presence and putting content on the green screen are a multitude better than using Powerpoint bulleted slides with the instructor on the side frame. It respects the signaling principle. It provides details on how to prepare a script for the course and how to add humor. It makes a bold statement that MOOCs' future is integrating video-making and video-gaming features. Thus, the next generation of "videos" on a MOOC may involve interactive video game-like components.

Based on the interview during the study, Zheng et al. [24] suggest that the completion or success of a MOOC depends on students' motivation. They concluded that few learners are there for certificates and others for learning. Thus, they advise two modules and present the course, which suits both categories' needs. A learning-driven module can remove problems such as lack of time confidence, falling behind the schedule. On the other hand, structured content in a timely fashion can be presented to the students preparing for some job as they need the certificate.

Recent work on MOOCs by Hew et al. [13, 14] study the learning from 3 and 10 most highly rated courses. Emphasis on problem-solving by instructors, their passion for the subject, and availability via various means whether email, forums, or live interactions, is most appreciated by learners. According to these studies, peer interaction and forums can be the most significant asset if appropriately done or the biggest deterrent in experience if left unchecked. Thus, both studies discuss ways to involve course staff in the moderation of content and encouragement of content by posting new threads in forums. Another problem which the studies discuss is superficial reviews in the case of peer to peer review-based assignments. Rating calibration and mandatory minimum (say 200) words comment are a few strategies that the studies suggest. Short and inflexible deadlines are highlighted as a reason for student dissatisfaction. Quality of assignments, unclear wording, or instructors in course work is a few more reasons for dissatisfaction in these studies. Similar findings around flexibility, feedback and peer interaction, and ease of access-to-knowledge being the motivators are presented in [21].

Research around student retention and better learning outcomes point toward one more important aspect: the social or forum presence. Chen et al. [6] show that while instructor to student interaction (on forums) helps in student persistence to learn further in MOOCs, and student interactions with peers are not a significant contributor. In a similar light, another study measures social presence and scores in MOOCs. It argues that perceived social presence is different depending on the size of the posting cohort and the duration of the course. It further stresses that even smaller sized (by the number of participants) courses evaluated the establishment of familiarity, emotional connection, and sense of trust as low. These studies collectively question the need for forums and their effectiveness.

In this paper, we worked with an open survey question presented to thousands of learners. While each of these studies has taken one or more areas/categories and surveyed or researched the data in the same area, our study tried a different approach to invite ideas from all directions possible without any preconceived classification of what makes an incredible MOOC.

Our study thus has holistic ideas for MOOC creation listed entirely by the consumers of these courses. It covers ideas from content design to content delivery and assessment. It provides varied and uncovered ideas around better navigation to getting help in reading the documentation. It has ideas even beyond course completion to encourage continued learning in the area of the MOOC by following thought leaders on social media. These are the problems that learners face while learning or in the real world. After completing the course, they would not have got a forum or checkbox in their work environment to express those. Going through the action

items list presented in the study would help any MOOC instructor, designer, or platform creators get a glimpse into learners' psyche as to what they require for active learning.

3 Methodology

This section explains the various steps of our methodology.

3.1 Data Collection

For our study, we collected data through a survey. A set of open-ended questions was asked to all the participants asking for feedback on their expectations if someone created an online course. We asked the following open-ended questions in our survey request:

- What is the one thing you love in an online course?
- What is one thing which you hate or think can be improved?

We asked the same set of questions across many Facebook groups. The groups selected were mostly around web development technologies as these groups mostly learn things online. Courses in universities do not cover web development despite massive demand in the industry, and thus, there are many online courses to fill that gap. These groups have members ranging from 10 K–190 K. Therefore, much traffic was seen on these posts cumulatively. Since this was open to all, people from all geographies and ethnicity participated. Also, volunteer participation meant that only interested learners submitted their responses.

Participants responded to the survey with comments. These comments ranged from a single line to paragraphs of 10 lines describing things in detail. The verbose input was incredibly helpful in understanding the psyche of an online learner.

3.2 Response Analysis

We collated all the comments received through the survey response. We have considered the suggestion given in a blog post by Cho [7] from SurveyMonkey.com to analyze the open-ended survey responses. The below mentioned four-step process is used to gather useful results out of all the collected responses.

1. **Most Frequent Terms**
   Terms which occurred more than once were selected and put in bins. Even though the question was open-ended, the responses had an overarching theme.

Terms like real-life projects, delivery style, video length, assignments appeared in a lot more responses.

2. **Broad Category Creation**
   Categories were then created around these terms, and similar responses were put in the same category. Created categories are: Pre-requisites

   - Real-Life Projects
   - Instructor and Video
   - Content
   - Testing and Assignments
   - Post course Engagement
   - Individual Ideas

3. **Individual Response Analysis and Scoring**
   Every response was analyzed and categorized accordingly. It ensured creating new broad categories, which were added to the above list. Similar ideas were assigned a higher score inside the category to pay more attention during course creation.
   To compute the score, we followed a simple process. There were 211 comments on the forum; we went through each comment and created categories based on groups of similar comments. Some comments were then found out to be echoing the same ideas repeatedly, so we added them together like 1 +1 +1 and then presented the final score on these ideas (which were repeating) in Fig. 1.

Fig. 1 Term frequency for course ideas

4. **Finalize Ideas**
   The ideas with a high score under each category were then finalized as useful. Future online courses can be designed considering these ideas. Other ideas can be taken as suggestions and can be used to create a wholesome learning experience.

4 Results

4.1 Overall Analysis

According to learners, the most important thing to learn is how to *do real-life projects*. There is an overgrowing sentiment that the course content does not relate to the industry's work. There are many ideas around this theme, which we discuss in the next section. The second thing is that the content should cover basic needs. The ideal *video length* and *delivery style* are something that came out quite clearly from this survey, which many would not consider an important factor while creating courses. Points such as using *simple English* in courses came from respondents of non-native speakers. Though the term is not the same, there is much stress on assignments, quizzes, after-course projects, which we discuss in the following sections.

4.2 Category Wise Ideas

As discussed in the methodology part, broad categories were created out of the survey results to get concrete ideas. These ideas would help in making upcoming online courses effective. The ideas in these categories are ranked as per their score in the survey.

1. **Pre-requisites**
   - Considering that the course will end up doing a project, a link to that project's working sample would help students decide their commitment to the course.
   - Explain the reasoning behind choosing a set of tools, i.e., the programming language, the framework, and the hosting.
   - course should have established pre-requisites. If possible, videos of the courses required before doing the course should be provided.
2. **Real-Life Projects**
   - It should work on a project example related to actual software being used by people around the world.
   - It should focus on relating to job interviews and interview questions that relate to the course work.

- Code written should be near to production code, i.e., the code written in industry. Formatting, file naming, and folder structure are what learners expect to get from the course.
- Setting up the development environment, keyboard shortcuts that programmers use, and better documentation navigation are other ideas. These could be an excellent addition to make a course near to real-life scenarios.

3. **Instructor and Video**
   Video length should be ideally between 5 and 10 minutes. It means content should be broken into smaller capsules. Most respondents commented that it gets sleepy and boring in one long video.

   - The instructor should explain with simple words and keeping beginners in mind. Many MOOCs are information and abbreviation heavy. However, it could be a good practice to have an introductory video explaining the terminologies used.
   - Respondents stressed the delivery style of the instructors as well with "should not be monotonous," "at least try to be funny," "fun, energetic, and engaging." High energy instructors with a dash of humor are appreciated everywhere, even outside the MOOC world. However, in MOOCs, the learner interacts only with one personality, i.e., the instructor. Thus, it is essential to have a session high on energy and engaging.
   - As a learning medium, students prefer an "interactive live coding" platform, which would present the students a chance to try out what is being taught themselves.
   - While most MOOCs have a Professor on screen, there are few online course sites like egghead [10], which has the only computer screen recording. Respondents to the survey expressed a strong opinion against slide or code editor with just voice over. Preference is to have an instructor on screen.

4. **Content**

   - Course should start from the very basics. If the basics are not covered, then links to learning should be given. The assumption that a learner would know something before-hand is not advisable.
   - Adequate time should be spent on the usefulness of the content. Answer to, "Why are we learning this language/framework/software library/concept?" It helps the learner relate to the content better.
   - When learning a new language or a framework, it is better to build small projects along. This practice builds muscle memory as well as gives a sense of accomplishment after doing it.
   - Lessons should be in small videos and sessions. It would enable learners to save time by skipping sections that they already know. Also, it makes learning more systematic and more comfortable to resume.
   - Many online courses are screencasts with the "DO-AS-I-DO" approach. In this case, learners see and type the code and pick up concepts from it. However, survey respondents voted against this approach. According to

the responses, an ideal course will have these three ingredients apart from screencasts:

  - Conceptualization process: thinking a project from the concept and covering the entire thought process
  - Logic building
  - Problem-solving approaches

- Interview Readiness: Learners do most of these online courses to get better jobs or employment purposes. Therefore, respondents have a big ask that the course covers some interview preparation tips related to course material. Some programming languages/frameworks have age-old interview questions/ patterns. It is then advisable to cover those as part of the course.

5. **Testing and Assignments**

- Actual exercises and challenges after each section/ lesson to test the concepts covered.
- Respondents were vocal about the need for lots of exercises, mini-projects, and major projects. It is because once a learner grasps a new language or framework, and he/she wants to try it out. However, it is unsure as to what all things can be created using the concepts. A given problem/project would scope the learning to it as the instructor has a much broader idea.
- The question inquired things which learners dislike; one thing that came out quite strong is being tested on something not covered in the course. It should be taken into consideration while creating assignments and quizzes.
- Capstone projects are the idea in online courses where everything taught over many lessons are brought together to solve one big problem and develop a practical application. Survey respondents were quite optimistic about such assignments at the end of a vast course.

6. **Post course Engagement**

- Course completion should be done with project ideas to practice upon and build expertise.
- One thing that respondents dislike about online courses is that some do not provide a book or some material to follow through. A supplement reading material would help in cementing the concepts further.
- In a technology spaced world, it is easier to follow thought leaders in any industry to know more about an area and stay on top of the trends. Respondents demand that courses should list out people or blogs to follow for a wholesome learning experience.

7. **Individual Ideas**

- Some of the respondents said that teaching via gaming is an effective technique. This technique is used in small tutorials and is very famous for learning Cascading Style Sheet (CSS) [4, 11, 20]. It may be that this could be explored more for learning other disciplines as well.

- Few of the respondents pointed out that the online course should be accessible to the physically challenged crowd. It is an area that needs much thought. How can one make videos accessible more? Moreover, that too videos with coding content/text content?
- Some respondents pointed out that they do not mind paying for suitable courses, but the course should be at the price point of Udemy [23]. Udemy is an online marketplace for courses that are incredibly professional, short, and to the point. It provides a completion certificate, and since the platform is mostly on "sale," the courses' prices are way less than competing Coursera and Udacity courses. It needs further research to ascertain whether the universities offering online courses through the MOOC platforms are being viewed as value for money by the course takers.
- The survey question asked for things people love in online courses. Survey responses are freedom of time, place, and what to learn is much appreciated, and the reason why MOOCs are being consumed at this rate.

5 Discussion

The responses from the survey were pinned into sections depending on the number of occurrences. There is an entire section about instructors, and learners had given feedback on their presence, energy, tone, and sense of humor.

Adamopoulos [1] has analyzed student retention in MOOCs. He has suggested that professors/instructors are the most important in online course retention and have the largest positive effect on a course. This trend is more supported on sites like Udemy, where multiple courses are offering the same subject, but instructors with higher ratings keep getting more students. New courses from these instructors see multiple signups, and the students have tuned themselves to the teaching style.

The idea of democratized learning via online resources seems exciting because it opens many doors for the learner. One of these doors is a job, a recurrent theme in the survey is about learning to build logic and getting sections dedicated to interviews. In another study, by Christensen et al. [8] have concluded that 21.3% of the responders participate in MOOCs to "gain specific skills to get a new job" while the other 49% said that they do it to "gain specific skills to do any job better." The results from these surveys are a clear indicator that the course should be job-oriented. It makes sense seeing in computer science learning people are paying $200 for egghead.io [10] or $400 for Frontend Masters [12] subscription every year. These sites produce courses for job-specific requirements, and most of the time, the employers end up buying bulk licenses for employees. To summarize, adding job-related projects would help students to commit more time and money to the course, knowing that knowledge gained will be compensated later.

Respondents have asked for exercises, assignments, and small projects between courses and even after the course. Quizzes have proved an effective way of increasing retention of the topic covered in memory. Quizzes require retrieval and thus makes

a student's brain to retain the content better. According to Brown et al. [5], while quizzes serve this purpose, coding or work assignments help cement the learning even further. Koedinger et al. [15] support this as well, going as far as to claim that the students benefit as much as six times when doing things rather than watching only video content. While multiple-choice questions are easy to evaluate, coding, or work assignments pose a challenge in evaluation when the number of students is massive. To solve this problem, let us go back to one of the survey questions: teaching via games. That suggestion, if implemented properly, has two benefits:

(a) it would encourage a student to "do more" rather than just "watching" the video, thus increasing participation and retention, and
(b) it makes grading assignments automated, just in time, and thus easier to scale the course as well [22].

While the survey covers most of the points, there is one crucial point that respondents have not included. This miss could be attributed to the fact that the tool they did not talk about was the tool they were using, i.e., the online learning forums. Almost all MOOCs today provide a forum for discussions and Q&A. These groups are quite helpful and should not be missed if creating a new course today. Students need these forums to ask for help when facing difficulty in the absence of which they are most likely to cease participation in the course [9]. However, in today's age choosing between a forum, a Facebook group, a Twitter handle, and many more choices can be exhausting. The work of Alario-Hoyos et al. [2] is quite helpful as it points out that students' most preferred tool is a forum. It goes forward to state that too many choices can, in fact, "hinder the real learning process" since participants can get let lost in too many posts without the proper filtering of relevant material.

The human brain responds to motion better. It is the reason all user interfaces today use animation to draw user's attention. The same is true while teaching. Therefore, survey respondents pointed out that they prefer to live code compared to slides. Learners also prefer professor/instructor on screen rather than just a voice over with screencast. It follows the feedback from the edX MOOC platform, which pointed out that participants would prefer a hand, writing an equation or sentence on paper than stare at with writing already on it [17].

In this study, we came out with learners' points to make the learning experience better following some previous research. However, a few of the ideas around making courses more accessible are unexplored. Accessible in the sense of getting it to people with disabilities. Furthermore, a price point, which is sensitive to the learner's spending capabilities are both excellent points to work.

Having blogs and people to follow at the end of a course is something we have not seen in any online course or MOOC. However, it points to the apparent need for people to stay updated using social media tools.

5.1 Limitations of Study

This survey was conducted on an open platform, and volunteers responded out of enthusiasm. Therefore, it suffers from response bias and does not have many dislike points around MOOCs.

Out of 211 responses, most responses were from men, and almost all (barring one) are from English speakers. While this group is the primary consumer of MOOCs, it would be good to include people from marginalized representation and work toward their inclusion. The survey was shared in groups that had computer science or STEM learners only.

Since the survey questions were open-ended, there are areas not covered in responses but useful when creating a MOOC. There were not responses around learning groups. Although all viewed the answers, there is a possibility some did not participate because their ideas were already voiced. Running an open for all, visible to all survey has its advantages and disadvantages as well. The absence of anonymity on open social media groups might be detrimental to many prospective participants.

6 Conclusion

The survey was conducted to understand the learner's psyche and need to create better courses. It is enough to say that the outcomes help any new course creator and make common sense.

When creating a course, two things are universal: (1) either the instructor is obsessed with getting the material so right that they forget about the generic points, about presentation and course structure (2) or, the general idea is to take an already taught classroom course, record videos and put it online trying to create the same experience. Both overlook the fact that MOOC is different and needs to be handled differently. With the reach presented by current MOOCs (the massive word is there for a reason), even special care should be taken while creating one, especially when you have some university or other credentials backing it.

This paper's findings will serve as a checklist of things to do before and after preparing an online course. The discussion has points to go through before even start preparing the course to one after the content is entirely consumed, i.e., the capstone project and providing whom to follow to read the next list.

References

1. P. Adamopoulos, What makes a great mooc? An interdisciplinary analysis of student retention in online courses, in *ICIS* (Milano, Italy, 2013), pp. 1–21
2. C. Alario-Hoyos, M. P´erez-Sanagust´ın, C. Delgado-Kloos, M. Mun˜oz-Organero, A. Rodr´ıguez-de-las Heras, et al.: Analysing the impact of built-in and external social tools in a

mooc on educational technologies, in *European Conference on Technology Enhanced Learning* (Springer, Berlin, 2013), pp. 5–18
3. T. Anderson, Teaching in an online learning context. Theory Pract. Online Learning **273** (2004)
4. C. Battle, Css code-golfing. https://cssbattle.dev/ (2020) [Online]. Accessed 29 June 2020
5. P.C. Brown, H.L., Roediger III, M.A. McDaniel, *Make it Stick* (Harvard University Press, 2014)
6. H. Chen, C.W. Phang, C. Zhang, C., S. Cai, What kinds of forum activities are important for promoting learning continuance in moocs? in *PACIS* (2016), p. 51
7. S. Cho, How to analyze open-ended survey responses. Available: https://www.surveymonkey.com/curiosity/open-response-question-types/. Online; Accessed 10 Apr 2020
8. G. Christensen, A. Steinmetz, B. Alcorn, A. Bennett, D. Woods, E. Emanuel, The mooc phenomenon: who takes massive open online courses and why? Available at SSRN 2350964 (2013)
9. D. Coetzee, A. Fox, M.A., Hearst, B. Hartmann, Should your mooc forum use a reputation system? in Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (ACM, 2014), pp. 1176–1187
10. O. Veblen.: Learn to code—egghead.io, video tutorials for badass web developers. Available https://egghead.io/ Online. Accessed 30 Apr 2020
11. Froggy: Flexbox Froggy. https://flexboxfroggy.com/ (2020), [Online]. Accessed 29 June 2020
12. FrontendMasters: Master the important JavaScript and front-end development skills—FrontendMasters, web development course providers. Available https://frontendmasters.com/ [Online]. Accessed 15 May 2020
13. K.F. Hew, Promoting engagement in online courses: What strategies can we learn from three highly rated moocs. Br. J. Edu. Technol. **47**(2), 320–341 (2016)
14. K.F. Hew, Unpacking the strategies of ten highly rated moocs: implications for engaging students in large online courses. Teachers Coll. Rec. **120**(1), n1 (2018)
15. K.R. Koedinger, J. Kim, J.Z. Jia, E.A. McLaughlin, N.L. Bier, Learning is not a spectator sport: doing is better than watching for learning from a mooc, in *Proceedings of the second (2015) ACM conference on learning scale* (ACM, 2015), pp. 111–120
16. B.A. Oakley, T.J. Sejnowski, What we learned from creating one of the world's most popular moocs. NPJ Sci. Learn. **4**, 1–7 (2019). https://doi.org/10.1038/s41539-019-0046-0
17. L. Pappano, *The New York Times*, 2 November 2012. Available https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html [Online]. Accessed 26 Apr 2020
18. G. Salmon, *E-tivities: The Key to Active Online Learning* (Routledge, 2013)
19. D. Shah, By the numbers: MOOCs in 2019. *Class Central*, 11 December 2018. Available https://www.classcentral.com/report/mooc-stats-2019/ [Online]. Accessed 26 Mar 2020
20. Sid: Learn react hooks by building a game. https://gumroad.com/l/hooks-team (2020) [Online]. Accessed 29 June 2020
21. R. Sujatha, D. Kavitha, Learner retention in mooc environment: analyzing the role of motivation, self-efficacy and perceived effectiveness. Int. J. Educ. Dev. Using ICT **14**(2) (2018)
22. N. Tillmann, J. De Halleux, T. Xie, S. Gulwani, J. Bishop, Teaching and learning programming and software engineering via interactive gaming, in *Proceedings of the 2013 International Conference on Software Engineering* (IEEE Press, 2013), pp. 1117–1126
23. Wikipedia contributors: Udemy—Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Udemy&oldid=906447764 (2020) [Online]. Accessed 29 June 2020
24. S. Zheng, M.B. Rosson, P.C. Shih, J.M. Carroll, Understanding student motivation, behaviors and perceptions in moocs, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (ACM, 2015), pp. 1882–1895

A Neural Attention Model for Automatic Question Generation Using Dual Encoders

Archana Praveen Kumar, Gautam Sridhar, Ashlatha Nayak, and Manjula K Shenoy

Abstract In the field of education, framing right questions is an important way of measuring the knowledge of an individual, and automation of the generation of questions would reduce the strain on educators and thus increase efficiency in learning. Literature survey reveals that while there are several promising methods for testing comprehension generating questions, but none of them can be used reliably for generating meaningful questions all the time, and thus, further research is required. A novel method of generating questions is presented in this paper from input text, by encoding the answers into the model and using attention that learns the dependencies between both input text and answers and between input text and questions, thus generating questions which are relevant to the answer given. The dataset used is the SQuAD dataset, and the model produces fluent questions which have been evaluated with the required metrics.

Keywords Comprehension question generation · Long short term memory · Word embeddings

A. P. Kumar (✉) · A. Nayak
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
e-mail: archana.kumar@manipal.edu

A. Nayak
e-mail: asha.nayak@manipal.edu

G. Sridhar
Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
e-mail: gautam.s2@learner.manipal.edu

M. K. Shenoy
Department of Information Communication and Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
e-mail: manju.shenoy@manipal.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_34

1 Introduction

Education is continuously evolving, along with this, the way we go about educating is also tremendously changing. Questions are a fundamental tool in gauging the skills of a student, and the ways of generating questions need not only a degree of skill, but also good amount of time. One most important application of generating questions in the field of education is to generate comprehension questions for passages. Question generation deals with the problem of generating or "asking" question, given a sentence or a paragraph. Figure 1 as shown has three manually generated questions from a given corpus. Such reading comprehension question has variety of uses, as the ability to ask questions from a given statement, which is a defining characteristic of intelligence. The initial methods of question generation used specific heuristics, which while not completely flawed, could not be generalized to different domains. These approaches have also created fairly simple questions, which were almost completely from the input text. Most recent research in this direction has been done using deep learning. One of the biggest advantages of this study contributed to the creation of labelled datasets. One such labelled standardized dataset is Stanford Question Answering Dataset (SQuAD) which was created by Rajpurkar et al. [1] .

The literature review shows all the models which have been implemented in this direction wherein the initial works were done by Suskever et al. [2]. The research had proposed a sequence-to-sequence model which would work on any sentence making minimal assumptions on the structure of the sequence. This required mapping the input to a fixed dimension vector using a multilayered long short-term memory (LSTM) and using another LSTM which could decode the target sequence sentence from the input vector. Though the experimental system was used from translation of English language to French, it paved the right path for making advancements in

Fig. 1 Example of a sentence picked from the first paragraph of the article Delhi, along with the manual generated questions and answers

the current question generation which is what has been used in this paper as well. Several papers since the work of Sutskever have reused the researcher's work in various ways; however, the base has remained the same. The inputs to the encoder and decoder architecture, however, have varied from approach to approach.

Xingwu et al. [3] encoded the position of the answer and used a copy mechanism; however, this caused the words to move far away from the answer. Xiyao Ma et al. [4] encoded lexical features along with the answer position as the input to the encoder. Yanghoon Kim et al. [5] replaced the answers with a special token and added a second encoder in order to obtain the information from the answers. Vrindavan and Marilyn [6] also encoded lexical features, but used more than just the basic named entity recognizer (NER) and parts of speech (POS) features, e.g. case feature, answer position, etc. Qingyu Zhou et al. [7] built upon the base model, added a copy mechanism to it and used lexical features to generate questions, while Du et al. [8] used a simple sequence-to-sequence model with attention for the same task.

The previous works do an excellent job of generating questions; however, the fluency of the questions is not the best in all types of questions, and they struggle with paragraphs. Our model is simpler and produces almost equal or better results, generating fluent questions even from the paragraph input. Our model uses two encoders, one to encode the answer, and one to encode the input text. The model uses the attention score between these two encoders and the decoder, to model the probability distribution of the vocabulary, in order to generate the questions. The model thus outperforms other previous works and therefore generates fluent questions which pertain to the answer and produces diverse questions.

The rest of the article is subdivided as follows. The related works are presented in Sect. 2. Section 3 gives brief explanation of background knowledge required to solve the problem. Methodology is explained in Sect. 4. Section 5 provides the analysis and discussions of the results. Finally, we conclude our work in Sect. 6.

2 Related Work

Reading comprehension, a challenging task for machines requires not only the natural language understanding but also the real-world knowledge of the context used in the paragraphs or sentences as stated by [1]. Moreover, in this direction, the researchers have been contributing in making many datasets available so that the training of this task to the machine could be easily done. Weston et al. [9] have created bAbI, a synthetic dataset which has features for twenty different tasks. Hermann et al. [10] have created a corpus for cloze questions by replacing the required entities with their placeholders in abstractive summaries generated from CNN/Daily Mail news articles. Chen et al. [11] have claimed that the former dataset is easier and the system performs very well using the CNN/Daily Mail dataset. Richardson et al. citeRich has created MCTest consisting of natural questions paired by four options. But this dataset is too small to be used for training answering models. Of recent advancements, Rajpurkar

[1] has relased SQuAD dataset which overcomes all the aforementioned problems in terms of size and quality.

Question generation is the most sought research of this generation and has attracted many of the researchers to actively promote experiments in this direction. Most works tackle the methods by imposing rules and patterns which transform the input into a representation which could be then transformed into an interrogative approach. The related works target at reviewing the research which have made use of models to train the machine and then automatically generate the questions. Xingwu Sun et al. [3] for question generation used a sequence-to-sequence approach, along with a pointer generator model. It combined a position aware and answer focused model. In position aware model, they directly incorporated relative word position embedding. This gave the model information about where the model is and what context words are close to it. In answer focused model, they added a restricted set of question word vocabularies in order to explicitly model the question words. The model produces a question word based on the answer embedding. The two models are combined using their context vectors, and an output is generated. Due to the encoding of answers, the model is better at generating a question word relevant to the answer. However, as the position of the answer is paid more attention to, the other words which are far from the answer in the context, but still relevant, are not taken into consideration all the time.

Xiyao Ma et al. [4] use a sequence-to-sequence model, with the inputs as the named entity recognition (NER) features, the parts of speech (POS) features and answer position features along with the context. The output of the encoder is passed through an answer aware gated function and then passed through the decoder. An attention mechanism along with a copy mechanism is implemented which performs sentence semantic matching using the last hidden state of the decoder output and the output of the answer gated function. The researcher then uses an attention flow network, in order to improve answer position inferring. The model improves upon previous work by improving the question words and copying only the parts of the context that are of consequence. Yanghoon Kim et al. [5] use a sequence-to-sequence model with answer separation, which instead of adding features to the input passage replaces the corresponding answer with a special token. It then uses another encoder to encode answer information. An answer separated decoder is used along with a keyword net, which extracts the keyword feature from the given passage. Finally, it uses a retrieval style word generator, which then replaces the output layer of the decoder.

A sequence-to-sequence approach is again made use by Vrindavan and Marilyn [6]. The inputs to the encoder are varied, wherein not only are the tokens passed to the encoder but also the features regarding the data are given along with details about the answer. The answer signals are fed by the use of a binary signal, along with the tokens. Case features (whether or not the token has a capital letter), NER features and coreferences feature are also added to the input of the encoder. After the encoder completes the calculations, the output is concatenated with the token level encoding, which allows the answer encoding to affect the final output more. For training initially, an encoder–decoder model is given the question as the input. After the training, the model learns to "copy" the input as the target. The encoder is then decoupled from the model. A new encoder is trained with the sentences, and the

previous encoder's embeddings are used as the ground truth, training the model to maximize the similarity between the embeddings of the new and the old encoders. The model generates factual questions only and also requires answers from the context. Therefore, future work on generating answer tokens from context is required.

One more research contribution in this direction was from Zhou et al. [7] who build upon the neural question generation (NQG) framework for their study, then went on making NQG++ which was similar to NQG system with a difference that it encorporated a feature rich encoder into the sequence-to-sequence architecture. It also used a maxout layer as stated by Goodfellow et a. [12], along with a copy mechanism which was implemented by Gulcehre et al. [13]. This research in terms of question generation improves BLEU scores over the previously used NQG framework. Likewise Xinya Du et al. [8] used a simple sequence-to-sequence mechanism by incorporating an attention mechanism to the model. This caused the model to pay more attention to the relevant words in the context given and hence improves performance. The researcher placed an attention layer between the encoder and decoder, to generate a context vector and used it for the final prediction. It, however, does not produce the best kind of question for all types of questions and thus can be improved more.

In the model developed by Wang et al. [14], linguistic features have been used, along with answers encoded with the context as the input to the network. It uses an attention mechanism and a copy mechanism in order to generate readable questions. The model generates human readable questions, which are better in quality than the work before it. However, it only generates factual questions. More work is required for more advanced questions. To the best of our knowledge our research given potential results and has outperformed some of the baseline models in this direction. We use SQuAD dataset in our work and similarly focus on the generation of natural questions for reading comprehension materials through automatic approach.

3 Task Definition

Given an input text, we have to generate a meaningful question which pertains to the input text and is coherent in its grammatical structure. The question length can be arbitrary, and the lengths chosen for the same are elaborated further in the implementation details section. In the next section, we will elaborate on the architecture used for generating questions.

3.1 Model

Our model which is partially inspired by the sequence-to-sequence model done by Sutskever [2] is just similar to a human who could solve the task of reading comprehension. In order to generate an interrogative sentence from the declarative, there is

Fig. 2 Proposed model

a need to pay attention to certain parts of the input sentence as well as to the context in which the sentence has been written. Hence, along with the attention mechanism with the encoder and decoder architecture, the model is able to focus on certain parts of the input sentence while generating the word during decoding.

Model as shown in Fig. 2 is inspired from the seq2seq model Sutskever et al. [2]. The model has two encoders instead of the usual one, where one encoder is used to encode the paragraph, while the other is used to encode answers. The decoder takes hidden state from the second encoder (paragraph encoder) and is trained using a shifted version of the sequence.

3.2 Sequence-to-Sequence Network

The RNN encoder–decoder model is used for generating task-specific sequential output given sequential input. The architecture is widely used for problems such as machine translation, text summarization and dialogue models. In this case, we are using an input of paragraphs and generating questions from the paragraph which are relevant to it. There are two parts to the model, namely the encoder and the decoder. The encoder, an RNN network, takes in a variable length input and gives an output which encodes the data in the input text into a single vector. The decoder uses this output as the input along with a shifted version of the target to generate the questions.

3.3 LSTM

LSTM networks are RNNs which can learn long-term dependencies. They have three gates, namely input, forget and output. Forget gates are used in order for the LSTM network to decide whether or not the information has to be passed on to the next cell, or not. Information from previous hidden state and the current input is passed through a sigmoid function, and depending on the value received, the forget gate is set to 0 or 1. Input gate is used to update the cell state. First the input and the hidden state are passed through a sigmoid. Both the two are also passed through a tanh activation function. Lastly, the tanh output and the sigmoid output are multiplied together to give the input gate output. Then, the cell state is updated. This is done by first multiplying the cell state with the forget vector. Then, the input gate output is added to the cell state, thus updating its values. Finally, the output gate is used to decide the hidden state to be passed along to the next LSTM cell. We pass the previous hidden state and the current input through a sigmoid function. The cell state is passed through a tanh. The output of the previous two calculations is then multiplied together to give us the hidden state. A bidirectional LSTM uses the output of two LSTMs to form a final vector, with both the LSTMs going in opposite directions of each other and then are combined (concatenated).

3.4 Embeddings

Word embeddings are used to model language data. Each word is represented by a list of numbers, which correspond to some information about the word. They are made such that words which are similar have similar "embeddings" that is the lists of numbers are closely related. The closer the embeddings are, the similar the words. A word embedding was used to represent each of the words in the vocabulary. The dimension used was 512, and it was trained from scratch on the data.

3.5 Answer Encoder

The answer encoder is a bidirectional LSTM with hidden size and two layers. At any timestep, the LSTM cell of the encoder takes in the input from the embedding, w_t, at that timestep, and the hidden states of the previous timestep. The final output of the encoder as given in Eq. 3 is the concatenated output of both the forward as shown in Eq. 1 and the backward LSTMs as given in Eq. 2, giving us an output with a hidden size of 2*hidden size of input. The output vector is denoted as h_t.

$$\overrightarrow{h_t} = \mathrm{LSTM}(w_t, \overrightarrow{h_(t-1)}) \tag{1}$$

$$\overleftarrow{h_t} = \text{LSTM}(w_t, \overleftarrow{h_(t+1)}) \tag{2}$$

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{3}$$

3.6 Paragraph Encoder

The context encoder is also a bidirectional LSTM with the same hidden size and two layers. The input to this encoder is the paragraph. The hidden state is initialized by the output hidden state of the previous encoder. The final output of this encoder is denoted by p_t. The output of this encoder is passed through the attention layer.

3.7 Attention

Attention is used to allow the model to learn the dependencies between the input texts used and the questions generated. It is usually used to give the decoder information about each of the encoder hidden states and learn which of those are important; however, we have used it to also find the area of importance between the answer and the input text. It learns the alignments, that is, the relation between the two vectors. This is especially useful for long input data. A scoring function is first used on each of the hidden vectors as shown in Eq. 4, which gives the importance of each of the vectors. This is then passed through a softmax layer, forming the alignment vectors given in Eq. 5. Then, alignment vector is multiplied with the hidden vector output of the encoder to form the context vector given in Eq. 6. The final context vector is then fed into the decoder, by various ways depending on architecture to architecture.

$$\text{score} = (h_t, h_s) = h_t^T \cdot h_s \tag{4}$$

$$\alpha_t = \text{Softmax}(\text{score}(h_t, h_s)) \tag{5}$$

$$c_t = h_t^T \cdot \alpha_t \tag{6}$$

3.8 Answer Attention

The hidden states from the answer encoder, ht, and the paragraph encoder, pt are passed through the attention layer, to give the context vector $c^1{}_t$. The context vector is concatenated with the hidden state output of the paragraph encoder. This is then passed through a dense layer, to align the dimensions with the decoder. This gives us the context vector $c^1{}_t$ as shown in Eq. 7.

Fig. 3 Vanilla encoder–decoder architecture without attention

$$c_t^1 = \text{Attention}(h_t, p_t) \tag{7}$$

3.9 Decoder

The decoder is an LSTM with hidden size twice the hidden size of the encoders (for correct dimensions). It is initialized with the hidden state of the paragraph encoder. The questions, shifted by one, are fed into the decoder as inputs, giving us the output vector d_t as shown in Eq. 8. The output of the first LSTM layer of the decoder and the context vector $c^1{}_t$ is passed through the attention layer, giving us context vector c_t as given in Eq. 9. Then, c_t and the output of the decoder d_t are concatenated together and passed through a second LSTM layer as given in Eq. 10. The output of which is passed through dense layer with a softmax activation shown in Eq. 11 to finally get the output in the shape of (batch_size, sequence_length, vocab_size). P_v is the probability distribution of the predicted word over the entire vocabulary, and W_v and b_v are parameters. The architecture of encoder and decoder has been adapted from the vanilla encoder–decoder architecture without attention shown in Fig. 3. As shown in Fig. 3, C and C' are answer context vector and paragraph context vector, respectively.

$$d_t = \text{LSTM}(q_t, d_(t-1) \tag{8}$$

$$c_t = \text{Attention}(c_t^1, d_t) \tag{9}$$

$$s_t = \text{LSTM}||([c_t; d_t], s_(t-1)) \tag{10}$$

$$P_v = \text{Softmax}(W_v \cdot s_t + b_v) \tag{11}$$

4 Implementation

The source vocabulary size is set to 50k, and the target vocabulary size is set to 35k. The rest are mapped to an OOV token. The model was implemented on Keras with tensorflow as backend. Dropout between layers was set to 0.4 implemented as in the work of Zaremba et al. [15]. Optimizer used was Nadam, with the learning rate at 0.001, with a batch size of 32. The model was trained for ten epochs. Sparse cross entropy was chosen as the loss function so as to save memory. During decoding, a beam search with a beam width of 5 is used. Decoding is stopped when an <eos> token is generated or the maximum permissible length of the question is reached.

4.1 Parameters

Number of paragraph tokens : 130
Number of question tokens : 13
Number of answer tokens : 5
Training triplets : 79488.

5 Evaluation

We have used automatic evaluation wherein the metrics used are BLEU Papineni et al.[16], METEOR Denkowski and Alon [17] and ROUGE-L Lin et al. [18]. BLEU 1–BLEU-4 gives a score to the hypothesis text, by counting the matching 4-gram in the hypothesis to the 4-gram in the reference text. METEOR—This metric uses synonym matching and stemming and is the harmonic mean between the 1-gram precision and recall, with more weightage being given to recall. It was used to find correlation at the sentence level. ROUGE-L measures the longest matching subsequence between the reference and the hypothesis text.

6 Results and Analysis

We have compared our model perfomance with two models of Quinyu et al. [7] and of Yanghoon et al. [5] using METEOR, BLEU-4 and ROUGE-L metrics. Table 1 shows the comparison of the models on the above metrics.

As shown in Table 1, the proposed model works better with sentence input; however, paragraph input also gives results which are better than the SOTA algorithms. Our proposed model outperforms all the previous NQG models in all metrics except ROUGE-L. There is a significant increase in the BLEU-4 and the METEOR metric. Figure 4 shows the examples of questions generated by our model. Here, source is the input text; reference text is the gold standard question, and the generated question is the question generated by our proposed model. The text in bold is the answer for the question. The model predicts questions which are grammatically coherent and fluent in the language, even if the gold standard is somewhat incorrect as shown in the first example of Fig. 4.

The questions generated by the model are coherent and take the answer into account. However, the model predicts a different question than the one in the test set, thus bringing down the BLEU and METEOR score even when the question in and of itself is acceptable. In some cases, however, the model recognizes a date or a quantity

Table 1 Comparison of the models on the METEOR, BLEU-4 and ROUGE-L metrics

Models	BLEU-4	METEOR	ROUGE-L
NQG++ [7]	13.29	17.59	40.75
ASs2s [5]	16.20	19.92	43.96
Our model	18.27	23.08	42.97
Our model with paragraph	17.26	20.90	42.22

Source	The Aare also contains the waters from the 4,274 m (14,022 ft) summit of **Finsteraarhorn**,the highest point of the Rhine basin.
Reference	Where is the highest point of the rhine basin?
Generated	What is the highest point of the rhine basin called?
Source	In 1960 **Maurice Hilleman** of Merck Sharp Dohme identified the sv40 virus, shown to cause tumors in many mammalian species .
Reference	Who identified the sv40 virus?
Generated	Who identified the sv40 virus?
Source	The word insect comes from the latin word **insectum** meaning with a notched or divided body or literally cut into from the neuter singular perfect passive participle of insectare (to cut into to cut up from in – into) and secare (to cut) ; because insects appear cut into three sections. .
Reference	What is the latin term for insect?
Generated	What is the latin term for insect?

Fig. 4 Examples of questions generated by the model

Source	Apollo and his sister **Artemis** can bring death with their arrows.
Reference	It was believed that this woman could bring death with her
Generated	Who is Apollo's sister?
Source	Indic scripts such as tamil and devanagari are each allocated only **128 code points** matching the ISCII standard .
Reference	What is the ISCII standard?
Generated	How many code points are Tamil and Devanagari allocated?

Fig. 5 Examples where the generated questions gets a lower score

and provides a question which has a separate date or quantity as an answer, but is not the same as the one provided. The model works well for sentence input; however, its efficacy reduces if the input length becomes larger. Fig. 5 shows examples where the model predicts fluent questions; however, it gets a very poor score as the generated question is not the one given as the gold standard. This indicates a need for a better way to score questions, as there exists no metric solely for scoring questions. The model also generates questions which are not grammatically correct, but the meaning does come through sometimes. However, it does not produce the best question all the time and fails to copy the correct words from context on several occasions.

7 Conclusion and Future Work

In this paper, we presented a novel method of generating questions from input text, which is one of the key indications of intelligence. We use a dual attention-based model to generate the questions. The model works with input sentences to automatically generate questions for reading comprehension. Using an attention-based neural network model for this task has proved to be useful to generate fluent and coherent questions which has been compared with automatic evaluation metrics. The model produces questions which are grammatically coherent and relevant to the text input. It outperforms state-of-the-art models and produces good questions even with paragraph input. As there is no effective metric that can classify questions, it is hard to train the model.

Attention implemented was global; however, as the paragraph size is large, the model would benefit from local attention. The model will also benefit with a copy network, as it tends to not get the correct word from the context. Giving the model which type of questions to predict would also allow it to generate specific questions and not just random ones the model tries to. Future work is to extend the model to predict specific questions rather than the random question which the model generates. We plan to extend our research by generating options in an automatic mode in future. Benefit with respect to local attention and copy network as the size of paragraph increases needs to be checked as a future extension of the research.

References

1. P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang, SQuAD: 100,000+ Questions for machine comprehension of text, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)Association for Computational Linguistics* (2016), pp. 2383–2392
2. I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks. Adv. Neural Inf. 3104–3112 (2014)
3. S. Xingwu, L. Jing, L. Yajuan, H. Wei, M. Yanjun, W. Shi, Answer-focused and position-aware neural question generation, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (2018), pp. 3930–3939
4. X. Ma, Q. Zhu, Y. Zhou, X. Li, Improving question generation with sentence-level semantic matching and answer position inferring, in *Proceedings of the AAAI Conference on Artificial Intelligence* (2020), 8464–8471
5. K. Yanghoon, L. Hwanhee, S. Joongbo, J. Kyomin, Improving Neural Question Generation using Answer Separation (2018). arXiv preprint arXiv:1809.02393
6. H. Vrindavan, W., Marilyn, Neural generation of diverse questions using answer focus, contextual and linguistic features, in *International Conference on Natural Language Generation (INLG)* (2018), pp. 296–306
7. Z. Qingyu, Y. Nan, W. Furu, T. Chuanqi, B. Hangbo, Z. Ming, Neural question generation from text: a preliminary study. CoRR, abs/1704.01792, 2017. http://arxiv.org/abs/1704.01792
8. D. Xinya, S., Junru, C. Claire, Learning to ask: neural question generation for reading comprehension, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017* (1: Long Papers), pp. 1342–1352
9. J. Weston, B. Antoine, C. Sumit, R. Alexander, M. Bart, J. Armand, M. Tomas, Towards AI-complete question answering: a set of prerequisite toy tasks (2016)
10. M. Hermann, K. Tomas, G. Edward, E. Lasse, K. Will, S. Mustafa, B. Phil, Teaching machines to read and comprehend, in *Advances in Neural Information Processing Systems (NIPS)* (2015), pp. 1693–1701
11. D. Chen, D., B. Jason, M. Christopher, A thorough examination of the cnn/daily mail reading comprehension task, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (vol 1: Long Papers). Association for Computational Linguistics (2016), pp. 2358–2367. http://www.aclweb.org/anthology/P16-1223
12. I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio, Maxout networks, in *Proceedings of The 30th International Conference on Machine Learning* (2013), pp. 1319–1327
13. Ç.S. Gülçehre, S. Ahn, R. Nallapati, B. Zhou, Y. Bengio, Pointing the unknown words. CoRR abs/1603.08148 (2016)
14. W. Zich, A. Lan, W. Nie, A. Waters, P. Grimaldi, R. Baraniuk, QG-Net: a data-driven question generation model for educational content, in*Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (2018), pp 1–10. https://doi.org/10.1145/3231644.3231654
15. W. Zaremba, I. Sutskever, O. Vinyals, Recurrent neural network regularization (September 2014). arXiv preprint arXiv:1409.2329
16. K. Papineni, R. Salim, W. Todd, W. Wei-Jing and Z.: Bleu: a method for automatic evaluation of machine translation, in *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics (2002), pp. 311–318. https://doi.org/10.3115/1073083.1073135
17. M. Denkowski, L. Alon, Meteor universal: language specific translation evaluation for any target language, in *Proceedings of the Ninth Workshop on Statistical Machine Translation*. Association for Computational Linguistics (2014), pp. 376–380. http://www.aclweb.org/anthology/W14-3348
18. C. Lin, Rouge: a package for automatic evaluation of summaries, in ed. by S. Szpakowicz, M.-F. Moens, *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*. Association

for Computational Linguistics (2004), pp. 74–81. http://aclweb.org/anthology/W/W04/W04-1013.pdf

A Comparative Study of Efficient Classification Models

Roopashri Shetty, M. Geetha, Dinesh U. Acharya, and G. Shyamala

Abstract The lung cancer is a principle cause of deaths in both women and men. Diagnosis and treatment depend on the cancer type, its stage, and the performance status of status. Depending on the cancer stage, treatments such as chemotherapy, radiotherapy or surgery will be decided. Patient's survival can be determined based on his overall health, stage and other factors. Only 14 out of 100 people survive around five years after the diagnosis. This paper gives a comparative study of various existing efficient classification models carried out for diagnosing lung cancer and K-Nearest Neighbor classifier outperforms other classifers.

Keywords Accuracy · Classification · Lung cancer · Precision · Prediction

1 Introduction

The classification is a main subfield of data mining [1] which builds a model for predicting the classes. The model is first trained, and it is tested for new data. Decision tree, probability based, k-nearest neighbor, logistic regression, etc., are some of the classification models. The main application of the classification algorithms is in the

R. Shetty (✉) · M. Geetha · D. U. Acharya
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
e-mail: roopashri.shetty@manipal.edu

M. Geetha
e-mail: geetha.maiya@manipal.edu

D. U. Acharya
e-mail: dinesh.acharya@manipal.edu

G. Shyamala
Department of Obstetrics and Gynaecology, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
e-mail: shyamala.g@manipal.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_35

area of medical science for predicting different diseases [2]. Lung cancer is one such disease which can be predicted using classification.

Some of the symptoms of lung cancer include [3]:

1. Bronchitis or pneumonia
2. Cachexia (fatigue, loss of appetite, weight loss)
3. Change in regular coughing pattern or Chronic coughing
4. Dysphasia(difficulty swallowing)
5. Shortness of breath with activity or Dyspnea
6. coughing up blood or Hemoptysis
7. Pain in shoulder ,chest , arm, abdomen
8. Wheezing.

With the increase in the use of tobacco, mortality and morbidity rate also increase. Lung cancer usually develops in the epithelium of the bronchial tree, and it may affect any part of the respiratory system. Lung cancer often takes many years to develop and people between the ages of 55 and 65 are most likely affected.

This work undertakes the experimental study to investigate the influence of symptoms and associated risk factors in the lung cancer detection in the absence of experts so that the rate of mortality and prevalence could be reduced. A comparative study of various existing efficient classification models such as K-Nearest Neighbor (KNN), Logistic Regression (LR), Naive Bayes (NB) and Support Vector Machine (SVM) is carried out for the diagnosis of lung cancer and KNN classifier outperforms other classifers. Lung cancer dataset from UCI repository is taken as input data for all the classifiers. The data is preprocessed which drops the tuples with missing values and selects the attributes which contribute to the lung cancer. Then, different classification algorithms are used to classify the lung cancer dataset into 2 classes: Yes and No.

2 Literature Review

A model was proposed by Krishnaiah et al. [4] for correct and early diagnosis of lung cancer. Naive Creedal Classifier 2 (NCC2) and One Dependency Augmented Naïve Bayes classifier (ODANB) are used in data preprocessing and its classification. The data is preprocessed by replacing the missing values with the modes and mean. Using the symptoms such as age, pain in arm and chest, sex, shortness of breath, wheezing, etc., it can predict the probability of patients getting the disease.

Hsu et al. [5] compared different methods for multiple class support vector machines which are basically designed for classifying into two classes. This model extended to obtain multi class binary classifiers by combining several SVM models. However in large dataset, generating multiple classes using SVM is difficult to achieve hence is restricted to be used only to classify smaller data set.

Yadav, Tomar [6] developed a foggy k-means algorithm for accurate clustering of real-time datasets. This algorithm considers the lung cancer data set provided by

SGPGI, Lucknow which contains nine attributes like age, sex, BMI, family history, tumor size, tuberculosis, smoking, lymph node involvement, and radiation and forms two clusters, one cluster containing the records of the patients suffering from lung cancer and the other cluster containing the record of those who are not suffering from cancer. This method also handles outliers. Using lung cancer data set, foggy k-means algorithms selects some prominent attributes and depending on the value of these attributes number of clusters are decided. The attributes are plotted and clusters are obtained.

Bala and Dr. Kumar [7] explained a classification method ANN (Artificial Neural Networks) and its variants. ANN is a collection of input-output connections having an input layer, one or more number of hidden layers associating weights and an output layer. Weight of the connection could be iteratively modified to get the proper output from ANN. And it is found that ANN is not time-efficient for large datasets.

The performance of the KNN classifier [8] uses Euclidean distance as a standard distance metric to determine the k nearest neighbors of the query data points. To make the diagnosis based on historical data,large storage of information is used in this study. The probability of occurrence of a particular ailment is computed. The accuracy of such diagnosis is increased using KNN.

Christopher and Banu [9] carried out a study of classification algorithms such as the Bayesian network,48 and Naïve Bayes, algorithms for lung cancer prediction. The dataset used has 100 records with 25 attributes. This paper analyzes the performance of classifiers to provide an early warning to the users. The Naïve Bayes algorithm outperformed other classification algorithms.

It is understood from the literature review that

1. There is a lack of efficient diagnostic method for diagnosing lung cancer.
2. The prediction of the disease becomes complex as the symptoms and impact of risk factors associated with lung cancer are not implicit and vary.

2.1 Objectives

The objective of this work is to make a comparative study of various efficient classification models with respect to their classification accuracy for diagnosis of lung cancer.

3 Methodology

3.1 Database Description

Dataset used in this study is taken from UCCI repository. 568 records with 32 initial attributes are chosen in the dataset. The attribute such as, Age, Air Pollution, Bal-

anced Diet, Chest pain, Chronic Lung Disease, Coughing of blood, Dry Cough, Dust Allergy, Fatigue, Frequent Cold, Gender, Genetic Risk, Occupational Hazards, Obesity, Passive smoker, Shortness of breath, Smoking, Snoring, Swallowing difficulty, Use of alcohol, Weight loss, Wheezing, etc., are considered for predicting the lung cancer.

3.2 Preprocessing

The dataset is preprocessed using iPython Jupyter Notebook to drop the attributes that do not contribute to lung cancer and to drop the duplicate records and records with missing values. Finally 22 prominent attributes are considered for the prediction.

3.3 Experimental Setup

The preprocessed data is then given as input to different classifiers such as NB, KNN, SVM and LR. For the experimental analysis, and the classifiers are tested with 80:20, 50:50 and 70:30 data distributions for training and testing and performances of the algorithms are analyzed.

4 Analysis and Results

All the models have been built using Python. The performance of each model has been compared with respect to the accuracy. Accuracy is calculated as

$$\text{Accuracy} = \frac{(\text{True}_P + \text{True}_N)}{(\text{True}_P + \text{True}_N + \text{False}_P + \text{False}_N)} \tag{1}$$

where True_P = True Positive, True_N =True Negative, False_P=False Positive and False_N=False Negative

Accuracy of the models with different data distribution is shown in Table 1.

Table 1 Accuracy of the algorithms

Data distribution	SVM (%)	NB (%)	KNN (%)	LR (%)
80:20	84	82.5	91	87
50:50	78	80	87	85
70:30	81	86	89	85

Fig. 1 Accuracy of the models in 80:20 distribution

Fig. 2 Accuracy of the models in 50:50 distribution

From Table 1, it is clear that KNN algorithm outperformed when compared with other classifiers. Out of three different data distributions, 80:20 gave the better result.

The graphical representation of the accuracy values with different data distribution are shown in Figs. 1, 2 and 3.

Table 2 shows the Specificity (True_N Rate), sensitivity (True_P Rate), Precision, Recall and F - Measure of different classifiers with respect to 80:20 distribution.

$$\text{Sensitivity(Recall)} = \text{True}_P/(\text{True}_P + \text{False}_N) \tag{2}$$

$$\text{Specificity} = \text{True}_N/(\text{True}_N + \text{False}_P) \tag{3}$$

$$\text{Precision} = \text{True}_P/(\text{True}_P + \text{False}_P) \tag{4}$$

Fig. 3 Accuracy of the models in 70:30 distribution

Table 2 The performance evaluation of various classifiers

Classification algorithm	Specificity	Sensitivity/recall	Precision	F-measure
SVM	0.609	0.955	0.866	0.908
KNN	0.655	0.972	0.855	0.91
LR	0.564	0.938	0.877	0.907
NB	0.645	0.869	0.866	0.867

$$\text{F-Measure} = 2 * (\text{Recall} * \text{Precision})/(\text{Precision} + \text{Recall}) \quad (5)$$

From Table 2 it is clear that KNN classifier performed better than other three classifiers for the Lung cancer dataset considered.

5 Conclusion and Future Scope

A comparative study of various classification techniques for predicting the lung cancer is carried out. The classifiers take data from lung cancer disease database and data is preprocessed and classified using various classifiers. Finally, K-Nearest Neighbor classification model came out to be the effective model in predicting the lung cancer in patients.

This comparative study uses a lung cancer dataset with 568 records. This can be extended for large datasets. Other data mining techniques, like Association Rules, Clustering and Time Series and Ensembled classification algorithms can be used to improve the accuracy.

References

1. N. Rikhi, Data mining and knowledge discovery in database. Int. J. Eng. Trends Technol. (IJETT) **2**(23) (2015)
2. J. Han, J. Pei, M. Kamber, *Data mining: concepts and techniques* (Elsevier, Amsterdam, 2011)
3. N. Singh, S. K. S. Bhadauria, Early detection of cancer using data mining. Int. J. Educ. Manag. Eng. **1**(9), 47–52 (2016)
4. V. Krishnaiah, G .Narsimha, N. Subhash Chandra, Diagnosis of lung cancer prediction system using data mining classification techniques. Int. J. Comput. Sci. Inf. Technol. **1**(4), 39–45 (2013)
5. C.-W. Hsu, C.-J. Lin, Comparison of methods for multiclass support vector machines. IEEE Trans. Neural Netw. **2**(13)(2002)
6. A.K. Yadav, D. Tomar, Clustering of Lung cancer using Foggy K-means, in *International Conference on Recent Trends in Information Technology (ICRTIT)* (2013), pp. 13–18
7. R. Bala, D. Kumar, Classification using KNN. Int. J. Comput. Intell. **7**(13) (2017)
8. H.S. Khamis, K.W. Cheruiyot, S. Kimani, Application of k-nearest neighbour classification in medical data mining. Int. J. Inf. Commun. Technol. Res. **4**(4) (2014)
9. T. Christopher, J.J. Banu, Study of classification algorithm for lung cancer prediction. Int. J. Innov. Sci. Eng. Technol. (IIJISET) **2**(3) (2016)

Sentiment Classification on Twitter Media: A Novel Approach Using the Desired Information from User

B. Shravani, Chandana R. Yadav, S. Kavana, Dikshitha Rao, and Sharmila Shanthi Sequeira

Abstract All around the globe brimming with the IT companies, each company has different data that needs to be analyzed. It is impractical to build models for different data available. Hence, the work carried out serves for this need to build a model that works for different data efficiently. An application will be created, wherein a search box is included that takes an input (a word or any phrase) and gives the related tweets. This application solves the problem in two sections. Primarily, the training section is where we train the given dataset with Naive Bayes, KNN, and random forest algorithms, and in the second section, this trained model is used to test the dynamically extracted tweets from the Twitter media. The results are visualized with the help of a confusion matrix and the pie chart.

Keywords Naive Bayes · KNN · Random forest · Confusion matrix · Pie chart

1 Introduction

The Web 2.0 communities have given rise to plenty of publicly accessible Web sites that highlighted the user-generated content. User-generated content on the Web matters a lot to know how the information is being circulated in the real-time. These endless applications laid down a path for many studies and researches in the area of text mining, machine learning, and natural language processing. Sentiment analysis (SA) is one of the recent developments in the research area used in mining and natural language processing. Sentiment analysis targets identifying, extracting, and categorizing the sentiments from user-generated data [1]. In the task carried out, sentiment analysis is performed on Twitter data. The desirable advantage of Twitter is its accessibility that is the ease to use nature. This made Twitter a principle source for data extraction [2]. Also, the Twitter Web site restricts the user-tweets to 140 characters, thereby forcing the user to stay focused on the point and provides us with the qualitative insights about a topic.

B. Shravani (✉) · C. R. Yadav · S. Kavana · D. Rao · S. S. Sequeira
Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

P. Shetty D. and S. Shetty (eds.), *Recent Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1386,
https://doi.org/10.1007/978-981-16-3342-3_36

Twitter media is an open-source platform where one can find ample amount of data that are different from each other (movie review, brand endorsement, etc.). Every data communicated is dominant concerning the associated organization or company. But the regretful reality is that a single trained model is not sufficient to analyze the different types of data available. Is it possible to create separate models for all the data domains? It is highly impossible to design an enormous number of classifier models. It is impractical to achieve accurate models with high-performance rates. The fundamental issue here is the domain of the text. The work developed takes a progression to solve this issue. A simple, interactive sentiment-based approach that opens up the possibility to operate on different data domains is established. The problem discussed above can be solved by developing a simple interactive environment which collects the user responses and performs the training according to the specified algorithm on the selected data domain. In addition to this, many sentiment analyzers developed out in the market works with static data. In practical, one has to work with the real-time data. Hence, there is a need for a dynamic sentiment analyzer. The project developed also serves for this purpose, which works with the dynamic data by extracting the tweets from Twitter API. These dynamically extracted tweets are considered as the test data and are analyzed with the trained classifiers, and the results are visualized with pie chart.

2 Architecture

The block diagram of project work carried out is displayed in Fig. 1. The brief overview of work flow to solve the task is demonstrated sequentially.

3 Methods

- *Data Collection for analysis*
  The labeled datasets from different domains (airlines, movie,…) are collected, which are used to train the classifiers on different data domains. The US Airline Tweets [3] is one such labeled dataset which is collected to train the model for airlines domain. This dataset contains 14,640 rows and 15 columns.
  The columns which are not useful for analysis are removed and only the required columns are considered. The airline dataset which is reduced to two columns (airline sentiment and text) is displayed in Fig. 2.

- *Preprocessing Steps*

The purpose of the work incorporates raw textual data from Twitter sites. This is often considered as unstructured form of data. Thus, there is a need to convert the unstructured format of text into a structured fashion, which is formally known

Fig. 1 Representing the flow of classification process

Fig. 2 Dataset used in training phase

as cleaning the data [4]. To clean the data and make it more machine-readable, one should follow preprocessing steps. These steps help us with transformation of textual data to extract meaningful information from the textual tweets. The preprocessing steps adopted for cleaning the data are:

- Removing @ Symbols: The '@' symbols are one of the commonly used signs while tweeting online. However, these symbols will not give any useful information in analyzing the tweet. Thus, it is required to eliminate all these unwanted symbols. One can accomplish this task with the help of regular expressions.
- Negation Handling: Negation handling is the replacement of the contracted words with its full form. It is a common practice to use apostrophes for connecting words. For example, do not is written as don't using an apostrophe. It is suggested to split these negated words to maintain the uniform structure throughout the process.
- Removing Http Links: The raw form of textual data contains huge number of http links which does not add any weight in analysis part [5]. Hence, http links should be removed.

- Replacing Emojis: The electronic messages on the Web pages encourage the use of emoticons which are shortly called as emojis. These add meaning to the text [6], for example, smiling face, thumbs up, and red heart represents positive emotions; angry face and cross mark represents negative emotions. Notably, emojis can clearly describe what a person is trying to communicate. These emojis should be replaced with their respective English phrases.
- Removing URLs: The raw form of textual data contains a lot of URL mentions. This does not add any weight in analysis part [5]. Hence, the URLs should be removed.
- Removing Stop Words: The analysis of textual tweet is at the word level. Hence, there is a need to remove frequently occurring unwanted words which are generally known as stop words [4]. The, is, and, for, of, etc., are few stop words. One can remove stop words using the inbuilt libraries or by collecting the stop words list.
- Feature Extraction

Feature extraction is a dimensionality reduction technique that decreases the number of elements to be processed without eliminating the originality from it. In addition, to this, feature extraction also solves the problem of over-fitting, helps in increasing the speed of training model, and increases the accuracy rates. Bag-of-words (BOW) is one of the frequently used techniques for feature extraction and is easy to understand. In simple words, BOW is the storehouse of collection of words irrespective of their orders [2]. This method serves our purpose by extracting the words from different tweets. The number of occurrences which is mathematically called as frequency of each word is considered as the feature for training the classifier. In BOW technique, there are three tasks: 1. Tokenization, 2. Vocabulary of Words, and 3.Generating the Vectors. Tokenization aims at dividing the given text into pieces of words called as tokens. A vocabulary of words is formed by gathering all the unique words from the dataset and recording their respective occurrences. Eventually, the numerical vectors are generated which can serve as input for machine learning algorithms.

- *Training The Model*

Three different ML algorithms—Naive Bayes, K-nearest neighbors (KNN), and random forest are used to train the classifier.

Naive Bayes

Naive Bayes is the most frequently used ML algorithm. It is simple and easy to code. It involves two main mathematical concepts from probability [5]. They are conditional probability and Bayes theorem.

Conditional probability of an event, Y is the probability that an event will occur given the knowledge that an event X has already occurred. The conditional probability of y given X is easily computed by:

$$P(Y|X) = P(X \cap Y)/P(X) \tag{1}$$

The Bayes theorem is frequently used probability theorem. This calculates the unknown probability with the other know probabilities. This theorem is based on the knowledge of prior probabilities to calculate the posterior probabilities.

The prior probability is the estimation of the probability of event with current knowledge that is, before the new data is collected, while posterior probability is the reformed probability of an event considered after the occurrence of new data and is calculated by updating the prior probabilities using Bayes theorems [10]. The Bayes theorem is given by:

$$P(Y|X) = P(X \mid Y) * P(Y)/P(X) \tag{2}$$

Description of terms used:

$P(Y|X) \rightarrow$ The posterior probability of class Y given the evidence X is seen.

$P(X|\mathrm{Y}) \rightarrow$ The probability of likelihood for evidence X given class Y.

$P(Y) \rightarrow$ The prior probability of class Y (probability before the evidence X is seen).

$P(X) \rightarrow$ The probability of evidence X.

Naive Bayes is the probabilistic ML algorithm that can be used in prediction problems, spam filtering, and sentiment analysis. The major assumption adopted by the Naive Bayes is that features are independent that is one feature does not affect the other feature [5]. The Bayes rule is simplified with naive assumption that Xs (features) are independent to each other. The simplified calculation is shown below.

$$P(Y|X) = P(X1|Y) * \cdots * P(Xn|Y) * P(Y)/P(X) \tag{3}$$

As the probability of evidence remains same for all the target classes, this denominator can be ignored during the computation. Hence, the naive equation is given by:

$$P(Y|X) = \pi_i P(X_i|Y)\ P(Y) \tag{4}$$

Terms used in the algorithm are described below:

- $X = (X_1, X_2, \ldots, X_n)$ represents the set of features or attributes
- Y represents the target class (either positive or negative sentiment)
- P represents the probability function

Naive Bayes Algorithm

Step1: Compute prior probabilities

The prior probability of each target class (positive and negative) is calculated. It can be examined as the probabilistic measure of portion of each sentiment out of all the sentiments from the populated dataset. This can be calculated by:

Prior probability of class Y = total number of examples belonging to class Y/total number of examples in the set.

Step 2: Compute the probability of likelihood

The probability of likelihood is calculated for each target class. This can be mathematically performed by applying the dot product for the conditional probabilities of all features. This can be formulated as:

$$P(X|Y = k)\ P(X1|Y = k) * \cdots *\ P(Xn|Y = k) \tag{5}$$

Step 3: Compute the posterior probability

The posterior probability of each target class is calculated by using the naive Bayes formula. The task is accomplished by substituting the above two computations in naive equation. Mathematically, it is calculated as:

$$P(Y = k|X) = \pi_i P(X_i|Y)\ P(Y = k) \tag{6}$$

Step 4: Predict the sentiment

Once the posterior probability of each target class is computed, a max function is applied over these posterior probabilities to predict the sentiment. The one with maximum posterior probability is considered as the predicted class.

K-Nearest Neighbors (KNN)

KNN is the most simple and easy-to-use ML algorithms. It is used to solve the classification problems. KNN stands for K-nearest neighbors. The whole algorithm depends on the parameter 'K'. The target class is predicted based on the K-nearest neighbors of the test instance. It is termed as Lazy learner since it works on dataset at the time of classification. Before that it just stores the entire dataset [5]. The distances between train and test data are calculated. The k-most similar instances are considered to predict the sentiment of the class. The most frequent class label among the K instances is considered to be the predicted class. A step-by-step procedure involved in performing the KNN algorithm is discussed.

K-Nearest Neighbors Algorithm

Step 1: Handling the data

KNN is lazy learning algorithm, thus it starts the process from the scratch. Initially, the dataset is loaded, and the split operation is performed on the entire dataset which results in partitioning the complete dataset into train and test data.

Step 2: Initialize k

The value of K should be selected in such way that it improves the effectiveness of the classifier. The value of K represents the number of nearest neighbors which are used in voting process to predict the class label of new test instance.

Step 3: Compute the distance

For each example in the training data.

3.1 The distance between the test instance and each training instance is calculated. This calculation is performed with the help of Euclidean distance formula. The formula used to find the distance between two data points $a(x_1,y_1)$ and $b(x_2,y_2)$ is:

$$d(a,b) = \sqrt{\left\{(x_2 - x_1)^2 + (y_2 - y_1)^2\right\}} \tag{7}$$

3.2 The calculated distance is appended to the collection, which stores all the calculated Euclidean distances between the points.

Step 4 Sort the collection

The collection of calculated distances from all the training instances is ordered in an ascending fashion based on the values of computed distance.

Step 5: Selecting the points

The first K entries from the sorted collection are considered for the predicting class label of test instance.

Step 6: Predict the sentiment

The K class labels are pulled out from the K entries, and the test instance belongs to the most frequent class label.

Random Forest

The random forest algorithm is introduced to overcome the 'over-fitting' drawback of decision tree algorithm. The concept is based on the collection of multiple decision tree classifiers randomly, hence the name random forest. As the number of trees increases, the accuracy of the classifier also increases [5]. The prediction is done by considering the majority voting from all the decision tree classifiers. It is an ensemble learning method, which learns from multiple decision trees and thereby solves the problem of over-fitting by averaging the result. A step-by-step procedure involved in performing the random forest algorithm is discussed.

Random Forest Algorithm

Step 1: Select samples

In random forest algorithm, the very first step is to select random number of samples from the entire dataset.

Step 2: Construct decision tree

A decision tree is constructed for each randomly selected sample which results into multiple decision trees.

Step 3: Collect the predictions

A prediction result is collected from each decision tree, and a voting process is performed for each predicted result.

Step 4: Predict the sentiment.

A prediction result with more number of votes is considered as the final prediction.

- *Performance Tuning*

Hyperparameter Using GridSearchCV
The machine learning models are mathematical methods where one can learn about parameters by working with data. But there are some specified parameters which are tuned to obtain better predictions. Hyperparameter is one such parameter whose value is decided before the training process starts and is given to the model [6]. These parameters should be tuned to improve the performance of our model. This can be achieved by performing trails with different parameter values. The value that increases the performance of model and reduces the complexity of operation is selected as the hyperparameter value. GridSearch is one of the simplest methods used for tuning the hyperparameters. To simplify, the task a grid of hyperparameters is created, and this will allow us to try all the combinations of parameter values.

Cross Validation
Cross validation is the test performed on the model to check if the classifier predicts the sentiment correctly for unknown data (i.e., other than the training data). K-fold cross validation is widely used cross-validation techniques that works on splitting the overall data (train/test split) [7]. As the name indicates, this method contains a single parameter 'K' which is selected randomly or by previous experiences. Generally, the value of 'K' ranges from 5 to 10. Here, the data is shuffled in a random manner and is split into 'K' groups which are technically called as 'K' folds. Next, each unique group is taken as a test data (i.e., the holdout data) and rest (K-1) groups are considered as trained data. The process is repeated k times, and the evaluation score is noted for each group by fitting 'K' different models. These evaluation scores can describe the skill of the trained model.

- *Twitter Streaming API*

Twitter API (Application Programming Interface) serves for extracting the tweets dynamically from the Twitter sites. Tweepy is the Python tool used for extraction process [8]. This process generates the credential keys. By using these keys, one can search the required tweets using a hash-tag (Ex: #airlines). The output tweets of this extraction process will serve as the test data for our project. Once the dynamic data is available for testing, the data undergoes the same steps of preprocessing, feature extraction, and so on. One of the three trained classifiers is used to classify the sentiment. The results of test data are visualized with a pie chart, which is discussed in the results section.

Fig. 3 ROC curve of random forest classifier

- *Performance Evaluation*

In machine learning, use of an AUC (area under the curve) ROC (receiver operating characteristics) curve to visualize the performance of classification model. AUC-ROC curve aims at calculating the performance of classification model at various thresholds [9]. Area under the graph measures how better the model classifies the sentiments. The graph is plotted with false positive rate (FPR) on *X*-axis and true positive rate (TPR) on *Y*-axis. A model is said to have a good measure of separability, when AOC of plot nears to 1 and poor if it nears to 0. The ROC curve of random forest classifier is shown in Fig. 3.

4 Flowchart

The sequence of steps required to perform sentiment analysis is illustrated in this chapter. The control flow and inter connection between the different modules involved in entire process is visualized (Fig. 4).

5 Results

- *Confusion Matrix*

A confusion matrix which is also called as an error matrix represents the performance of classifiers on the test data and allows visualizing the confusion made by the classifier while predicting the sentiments (i.e., the miss-classification of positive to negative and negative to positive). In this matrix, the principle diagonal entries represent the true predictions, while the secondary diagonal entries represent the false predictions made by the model [10]. As shown in Fig. 5, error matrix is depicted as a table with four entries of actual and predicted sentiment combinations, where

Fig. 4 Flowchart of sentiment analysis

		PREDICTED VALUE	
		NEGATIVE 0	POSITIVE 1
ACTUAL VALUE	NEGATIVE 0	TN	FP
	POSITIVE 1	FN	TP

Fig. 5 Confusion matrix representation

the row elements belong to the actual class while the column elements belong to the predicted class.

Describing the terms from the matrix:

- TN: It stands for true negative, the case in which the sentiment predicted by the classifier is negative and the actual sentiment is also negative.
- FP: It stands for false positive, the case in which the sentiment predicted by the classifier is positive, but the actual sentiment is negative. The classifier made an error by miss-classifying the negative sentiment as positive.

- FN: It stands for false negative, the case in which the sentiment predicted by the classifier is negative, but the actual sentiment is positive. The classifier made an error by miss-classifying the positive sentiment as negative.
- TP: It stands for true negative, the case in which the sentiment predicted by the classifier is negative and the actual sentiment is also negative.

Confusion matrix allows us to perform calculations for finding different measuring rates that evaluates effectiveness of the model. In the work carried out, confusion matrix is used to calculate the accuracy of classifiers. It is the measure of how often classifier classifies the sentiment correct (i.e., the positive sentiment is predicted as positive and the negative sentiment is predicted as negative).

- *Accuracy of the Classifiers*

It is the measure of how often classifier classifies the sentiment correct (i.e., the positive sentiment is predicted as positive, and the negative sentiment is predicted as negative). Accuracy is calculated by performing the division operation on number of true predictions (TN & TP) and overall predictions (TN, FP, FN& TP). Accuracy is calculated as:

$$\text{Accuracy} = (\text{TN} + \text{TP}) - (\text{TN} + \text{FP} + \text{FN} + \text{TP}) \tag{8}$$

As soon as the classifiers are trained, there is a need to find accuracy of classifiers. Here, confusion matrix is used to compute the accuracies.

The confusion matrix of Naive Bayes classifier is shown in Fig. 6

Accuracy of Naive Bayes classifier is calculated as:

$$\text{Accuracy} = (1808 + 211) - (1808 + 40 + 250 + 211) = 87\%$$

$$\textbf{Accuracy of naive bayes classifier} = \textbf{87}\%$$

Fig. 6 Confusion matrix of Naive Bayes classifier

The confusion matrix of KNN classifier is shown in Fig. 7
Accuracy of KNN classifier is calculated as:

$$\text{Accuracy} = (1506 + 383) \div (1506 + 361 + 59 + 383) = 82\%$$

$$\textbf{Accuracy of KNN classifier} = \textbf{82}\%$$

The confusion matrix of random forest classifier is shown in Fig. 8
Accuracy of random forest classifier is calculated as:

Fig. 7 Confusion matrix of KNN classifier

Fig. 8 Confusion matrix of random forest classifier

Fig. 9 Extracted tweets from the Twitter

$$\text{Accuracy} = (1791 + 275) \div (1791 + 27 + 216 + 275) = 89\%$$

Accuracy of Random Forest classifier = 89%

- *Pie chart of test data*

As discussed earlier, dynamic data is used in testing phase. The tweets are extracted from Twitter based on certain hash-tags as per the requirements. The tweets gathered from extraction process are shown in Fig. 9.

The dynamically extracted tweets of united airlines have more of negative feedback because the flights have been canceled due to corona pandemic situation. These extracted tweets are analyzed with the three different classifiers developed in the training phase. The results are observed in the form of pie charts.

The prediction obtained from Naive Bayes classifier is visualized with a pie chart in Fig. 10

The prediction obtained from random forest classifier is visualized with a pie chart in Fig. 11

The prediction obtained from random forest classifier is visualized with a pie chart in Fig. 12.

- *Comparison of predicted results*

Fig. 10 Pie chart of Naive Bayes classifier

Fig. 11 Pie chart of KNN classifier

Fig. 12 Pie chart of random forest classifier

Table 1 Accuracy rates of classifiers

Proposed classifiers	Dataset	Accuracy (%)
Naive Bayes	US Airline Tweets	87
KNN	US Airline Tweets	82
Random forest	US Airline Tweets	89

The accuracy measurements of three different machine learning algorithms are calculated and listed in the Table 1.

6 Conclusion

In the work carried out, a classifier is developed that successfully classified the sentiments from different data domains. The model is trained with three different machine learning algorithms—Naive Bayes, KNN, and random forest. The results show that random forest gives better performance than the other two ML algorithms. The reason is KNN always depends on the parameter 'K', and it behaves abruptly with irrelevant attributes; and on the other hand, precision of Naive Bayes algorithm

is not makeable with smaller dataset, and the naive assumption makes it fall behind as the features may not be independent in real-time data. According to the calculations made, random forest is the best among three machine learning algorithms with 89% accuracy.

7 Challenges and Future Enhancements

The classifier was not able to classify the tweets which had both negative and positive sentiments. The processing speed becomes slower, due to the hidden noise in the text, with the increase in dimension of data. In recent times, sentiment analysis has gained significance, but it is restricted to English texts. The work can be extended to perform sentiment analysis on other languages. There is a huge requirement in improving the analysis on sarcastic sentences. There is an on-going research with deep learning as they produce good accurate models.

References

1. N. Jivane, Twitter Sentiment Analysis of Movie Reviews Using Machine Learning Techniques, 20 Nov 2018 [Online]. Available https://medium.com/datadriveninvestor/twitter-sentiment-analysis-of-movie-reviews-using-machine-learning-techniques-23d4724e7b05
2. R. Wahome, This Is How Twitter Sees The World: Sentiment Analysis Part One, 8 September 2018. [Online]. Available https://towardsdatascience.com/the-real-world-as-seen-on-twitter-sentiment-analysis-part-one-5ac2d06b63fb
3. Figure Eight, "Twitter US Airline Sentiment" [Online]. Available https://www.kaggle.com/crowdflower/twitter-airline-sentiment
4. S. Bhansal, Steps for effective text data cleaning (with case study using Python, November 16, 2014 [Online]. Available https://www.analyticsvidhya.com/blog/2014/11/text-data-cleaning-steps-python/
5. S. Ray, Commonly used Machine Learning Algorithms (with Python and R Codes), September 9, 2017 [Online]. Available https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
6. E. Lee, An Intro to Hyper-parameter Optimization using Grid Search and Random Search, June 6, 2019 [Online]. Available https://medium.com/@cjl2fv/an-intro-to-hyper-parameter-optimization-using-grid-search-and-random-search-d73b9834ca0a
7. Srinivasan, The Importance Of Cross Validation In Machine Learning November 23rd, 2018, November 23rd, 2018 [Online]. Available https://www.digitalvidya.com/blog/cross-validation-in-machine-learning/
8. A. Moujahid, An Introduction to Text Mining using Twitter Streaming API and Python, July 21, 2014 [Online]. Available http://adilmoujahid.com/posts/2014/07/twitter-analytics/
9. S. Narkhede, Understanding AUC-ROC Curve, June 26, 2018 [Online]. Available https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5
10. J. Brownlee, What is a Confusion Matrix in Machine Learning, November 18, 2016. [Online]. Available https://machinelearningmastery.com/confusion-matrix-machine-learning/

Object Detection System for Visually Impaired Persons Using Smartphone

D. Ravi Kumar, Hiren Kumar Thakkar, Suresh Merugu, Vinit Kumar Gunjan, and Suneet K. Gupta

Abstract Portable assistive technology systems are developed to enhance the capability of persons with disabilities. One of the most important senses for humans is vision. Vision is the most important sense which helps us in understanding the perception of the surrounding environment. Visually impaired persons face a lot of difficulty in understanding the perception around them, particularly in outdoor environment where objects are continuously changing and moving from one place to another. Object detection solutions would greatly assist visually impaired persons in avoiding from the barriers which they face in their daily routine life. The aim of the object detection system is to provide a simple, user-friendly, handy, economical and efficient solution for the visually impaired persons. Motive of this system is to develop a solution that detects the objects present using camera, as the input device in real time and communicate the same to the user using smartphone through headphones. The system would be using an audio device such as speakers or headphones in providing the information about objects to assist the visually impaired persons. The proposed system helps in identifying and avoiding the objects both in outdoor and indoor environments that affect day to day life activities and occupational performance of the visually impaired persons. The information about the objects in surrounding environment would be very much helpful to the visually impaired persons in their daily life.

Keywords Object detection system · Visually impaired · Blindness · Assistive system · Computer vision · Image recognition

D. R. Kumar (✉) · S. Merugu
Department of Research & Development, CMR College of Engineering & Technology, Hyderabad, India
e-mail: dravikumar@cmrcet.org

H. K. Thakkar · S. K. Gupta
Department of Computer Science Engineering, Bennett University, Greater Noida, India

V. K. Gunjan
Department of Research & Development, CMR Institute of Technology, Hyderabad, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

A. Kumar et al. (eds.), *ICDSMLA 2020*, Lecture Notes in Electrical Engineering 783,
https://doi.org/10.1007/978-981-16-3690-5_154

1 Introduction

There are several organizations that take cognizance of the welfare of the blind and visually impaired. [1] The World Health Organization has reported that there are 2.2 billion people who are either visually impaired or completely blind through a recent study conducted by them. [2] In India, the Ministry of Health and Family Welfare of the Government of India has conducted a survey called "The National Blindness & Visual Impairment Survey India 2015–2019". It was conducted for approximately 85,000 people. The results of the survey showed the estimation of the number of people with vision problems. The results are presented for the age group of 50 and above. The prevalence for blindness is 1.99% and the prevalence for visual impairment is 13.76%. The survey data supports that there is an astonishingly large number of people with vision related problems. The numbers of people that need support are huge and a simple, efficient system is needed in place for their assistance. A successful execution of a supporting system will have a large impact on the ways and livelihood of the people with visual impairment.

The challenges that the completely blind and visually impaired people face directly impact the quality of their lives and also how do they perform their daily activities. The traditional method is that of a supportive guide. A person is needed who can efficiently guide them in travelling from one place to another. The white cane is a helpful aid in navigating with obstacles in near proximity of the person. It can only help identify obstacles that are very close to the person. This brings up the issue of safety of the blind and visually impaired. Independent mobility also is matter of concern. An object detection system is the need of the hour to combat the safety issues and also enable them to have mobility that is not dependent on anyone. The object detection system has to be a simple and most importantly user friendly. It also should be budget friendly so that it can be afforded by them.

We are fortunate to have technologies that can make a huge difference in the life of the blind and visually impaired. Some advanced technologies like computer vision, object detection and object identification are available in today's world. The camera of a smart phone can be used as the vision apparatus which helps in the object detection system. It then detects the objects that are in front of the camera. Yolo V3 is an object detection model that can be used for detection of objects in a real time video. The detected object can be identified and then communicated as an audio message to the user. It can be done either on the speaker or using a head phones. This application is designed to be used on an Android based smartphone. Smart phones not only help in the regular telephonic conversations but also help the blind and visually impaired if the smart phones have a camera. The primary goal of the application is to enable the user identify the object in front of the camera. The computer vision techniques supported by the object detection systems can enhance the quality of lives of the blind and visually impaired by many fold. We are proposing a system that is simple and user-friendly. It is not only handy but also economical and efficient. This overcomes the issues of an expensive, additional sophisticated system. The use of smartphones is not just for the super advanced applications. A

common man is able to use a smartphone with a camera in a budget friendly manner. The proposed system will not need any additional hardware other than those are already available in a smartphone. Smartphone comes with an integrated camera. It can be used to capture the real time video. This in turn is used to detect the objects. The identified object is communicated to the user as an audio instruction. This paper highlights the computer vision techniques and presents the overview of the object detection system, which can enhance the quality of lives of the visually impaired persons.

2 Literature Survey

Various aids have been designed to develop a system that provide visual assistance and which help the persons facing vision problems in their daily routine life by communicating about the obstacles present in front of them. The following literature review presents the outcome and the limitations of some of the existing assistive solutions for blind and visually impaired.

In [3], authors proposed an android based system which captures the images through the smart phone camera and communicates the user about the object detected, the distance and direction of the object to the user via audio instructions using speakers or headphones. The proposed android based system provides the perception of the surrounding environment and enables them to navigate independently. The system has three modules: Object detection module, Distance calculation module and Direction. The object is classified and module isolate different objects present in the image is done in the object detection module. Object detection is done using TensorFlow's object detection API. Open source computer vision library (OpenCV) and Triangle similarity law are used to determine the distance of the object. Using 'x' coordinates of the bounding box of the object, the direction 'Left', 'Center' and 'Right' is determined and communicated the same to the visually impaired persons. The proposed system was able to detect with an accuracy of 87%. Some of the limitations identified are the distance calculation functionality is yet to be included in the system. The proposed system can be retrained to detect more objects.

The aim of the system [4] is to minimize the need of dedicated and wearable devices by providing an android smartphone based application system that assist visually impaired persons in recognizing objects and in moving around. The R-CNN algorithm is used for classification of real time objects. Image processing algorithms are used to process the video captured by the inbuilt camera of the smartphone. After objects are scanned and segmented, Fast R-CNN algorithm is used to identify multiple regions. Later, using Google API text to speech conversion is done to communicate the same to the user. The proposed system revealed that better accuracy is achieved in detecting with low latency by SSDMobileNetV2 model and performed 35% better in speed than MobileNetV1 when paired with TensorFlow Lite. The system has a limitation to focus on better performance in detecting multiple objects with better accuracy.

The goal of the proposed systems in [5–7] is to develop an Android application which provides necessary information about the objects and environment to the visually impaired persons. The proposed system minimizes the usage of the conventional dedicated devices and wearable devices. The paper also provides an overview of the object detection and significance of ORB over SIFTS and SURFS in different cases. The video is captured by the camera then the video is divided into frames. Based on the key matching in the database the image edge detection is processed then text to speech conversion is done using synthesizer which instructs the user.

The objective of the papers [8, 9] is to review the different navigation solutions available for visually impaired persons that work both in indoor and outdoor environment. A thorough review of the existing navigation solutions in both indoor and outdoor is made. In Assistive Travel Aids, the author classified available aids into three major categories: electronic orientation aids (EOAs), position locator devices (PLDs) and electronic travel aids (ETAs). Electronic orientation aids (EOAs)—used in finding the navigation path to the visually impaired persons. Position locator devices (PLDs)—both Global positioning system (GPS) and Geographic information system technologies are used to guide user navigation. Electronic travel aids (ETAs)—help visually impaired persons to avoid obstacles. Some of the limitations observed are in EOAs, it is difficult to add complex computing devices. Using PLDs, the device cannot detect objects and ETAs can be improved in the detection range of the obstacles. The study found lacking of some core features like portability, comfort ability, adaptation time and user learning which are very helpful and important for independent mobility in the existing navigation systems for visually impaired persons.

To develop an intelligent assistance system [10] composed of wearable smart glasses, mobile device app, an intelligent walking stick and an online information platform that assist in object detection. The components wearable smart glasses, mobile device app, an intelligent walking stick and an online information platform are composed in the system. Whenever the smart glasses detect objects it reminds user by smart walking stick. If the user falls down then the related information is stored online and shares the same information with their family members. Apart from object detection assistance, the major advantage in using this device is that whenever user falls down or any collision event happens then related information is sent to family members. A few limitations observed are features like recognizing front images like traffic signs and to develop intelligent walking guide are yet to be included.

The aim of the device in [11] is to provide the safe navigation assistance and perception of the environment in the real time. The obstacle avoidance system in the device learns from the RGBD, pilots choice-of-action input and semantic map, it ensures to provide safe and appropriate feedback about the surrounding environment to the visually impaired. A few limitations of the device are the sonar and bump sensors are yet to be incorporated to ensure the safety of the user in exception cases.

Matusiak et al. [12, 13] explained about the main features of the smartphone Android application software developed to assist visually impaired persons. The software has three modules, where the main module in the system recognizes and

matches the scanned objects with the objects stored in the system database. The other two modules are capable of identifying direction of the major colors regions in the captured images. The SIFT (Scale-Invariant Feature Transform), a computer vision algorithm is used in the proposed application to describe image features.

Devi et al. [14], Bhatnagar et al. [15] explains about determining obstacles and potholes and notify the user through audio output and to share active location of the user with the care taker. The obstacles detection part is carried out by ultrasonic sensors attached to the device interacts with the Microcontroller then the respective voice commands are given to the user. Potholes are identified by the infrared sensors and vibrates the aid to let the user know. GSM gathers the active location of the user [16, 17]. Whenever user press panic button then emergency assistance request is made. Apart from travelling assistance, emergency request is also included. The limitation observed in the system is since all components are placed on stick, weight may be a constraint and repeated hitting of stick on ground may cause malfunction of the device.

3 Proposed System

Visually impaired persons face many challenges in identifying objects and performing the day to day activities. Most of the time they are dependent on others in moving from one place to other. They face a lot of difficulty especially when moving in outdoor environment, where objects are continuously moving. This object detection system would assist the visually impaired persons by providing the perception of the surrounding environment or position of the objects. This device can assist visually impaired persons in avoiding the obstacles in both indoor and outdoor environment. The proposed system also helps the user in identifying the objects around them. It would minimize the visually impaired person's difficulties and help them lead a quality life.

The proposed system aims to be simple, user-friendly, handy, economical and efficient solution. Most of the existing assistive systems are highly sophisticated and expensive. Smartphones are widely used these days and the usage of smartphones by all persons has become very usual in the recent past. Therefore this system uses all the advanced built-in features of a smartphone. Smartphone's integrated camera is used to capture the real time video to detect the objects present and headphones or smartphone's speaker is used to communicate to the user through audio instructions.

The proposed application can easily be accessed by the visually impaired persons. The Yolo V3 (You only look once) algorithm is used to detect the real time objects captured from the continuous streaming by the smartphone camera. It is considered as one of the most powerful real time object detector algorithms. Unlike other algorithms like R-CNN and Faster R-CNN which examine several regions of the image to identify objects, Yolo passes image or video only one time through its network and uses a unique neural network using the characteristics of the complete image to predict multiple boxes containing an object. This is a significant feature in Yolo, which reduces the processing speed when compared with other algorithms. Processing speed plays a key role particularly while detecting objects in real time video stream. OpenCV provides a function that facilitates image pre-processing for deep-learning classification. Group of connected pixels in an image that share common properties (Blob ex: Grayscale value) of each frame captured is identified. Bounding boxes are created for each object identified and based on the Yolo V3 pre-trained weights, confidence score and coco dataset each object is processed and labelled. The coco dataset contains all the objects or class names on which the model is trained. Non-maximum suppression (NMS) uses function called "Intersection over Union (IoU)" which is used to determine the best boxes. In order to select the best box, NMS follows three steps. 1. It selects the box with highest score. 2. Computes the overlap with other boxes, removes the overlap which has more than a certain

threshold. 3. Iterates the process until there are no more boxes with a lower score than the currently selected box. The above NMS steps will remove all the boxes that have a large overlap with the selected boxes. Only the best boxes remain. After detecting the object by the proposed model, it produces the text format output. Then using text to speech converter, the label of the object detected is converted to speech and communicated the same to the user through headphones.

4 Methodology

The object detection, recognition and communicating speed of the object to the user should be very fast. Since the proposed system is intended for visually impaired persons, the response time of the system plays a key role. Delay in communicating about the obstacles to the user will not meet the purpose. To overcome processing speed and delay issues, Yolo V3 algorithm is used in the proposed system. Yolo V3 algorithm is faster than many other real time object detecting algorithms.

Methodology

1. Captures the objects in the surrounding environment in real time and this is done by using the camera of the smartphone. The video will get divided into frames
2. The RGB color image is converted into greyscale image
3. Group of connected pixels in an image that share common properties of each frame captured is identified (Blob)
4. Based on the Yolo V3 pre-trained weights, confidence score and coco dataset each object is processed and labeled (Yolo V3 algorithm)
5. Using text to speech converter, the label of the object detected is converted to speech

5 Results

The main motive of the proposed system is to assist visually impaired persons by providing the perception of the environment, which helps them in avoiding obstacles or barriers and in moving from one place to another. The goal to provide a simple, user-friendly and handy solution is achieved. The proposed system is capable of detecting the objects present in the surrounding environment with good speed and accuracy. It detects objects effectively in both outdoor and indoor environment. The system is able to successfully detect the multiple objects present in the surrounding environment and communicate the same to the user in audio through headphones or speaker.

The proposed system is tested in detecting objects in indoor environment, outdoor environment and objects which are more than 10 m from the camera. The system is capable of detecting the objects in the surrounding environment and provide audio output to the user. The performance of the application in the above mentioned three categories is satisfactory. The results captured are presented below.

5.1 Indoor Environment

The following figure show the objects captured by the built in camera and the detected objects with the labels and the respective audio output of the proposed system.

5.2 Outdoor Environment

The following figure shows the objects captured by the built in camera and the detected objects with the labels and the respective audio output of the proposed system.

The proposed system successfully detected the objects which are far away in the outdoor environment with good accuracy. Following are the objects that are detected in the outdoor environment. The system was successful in detecting the objects which were more than ten meters away in the outdoor environment.

6 Future Scope

The performance of the system can be further improved by re-training on larger datasets. Providing information about direction of the object will help the user to

understand the perception better. The system can be improved in detection time and better accuracy.

7 Conclusion

In recent years, there were many assistive solutions developed for the visually impaired persons to provide assistance in detecting objects present in their surrounding environment and in moving from one place to another. But most of the existing solutions are expensive, highly sophisticated, difficult to handle, designed as a dedicated aid, require training etc. Our primary goal is to provide an assistive solution to the visually impaired persons which is simple, user-friendly, affordable and handy and assist the visually impaired persons in understanding the surrounding environment and help them in moving from one place to another place independently. The real time video stream is captured by the smartphone camera and then the label of the object detected using object detection algorithm is communicated to the user in speech through the speakers or headphones. The audio output communicated to the user would help them in performing their day to day activities and lead a quality life.

Acknowledgement & Funding Statement: Authors would like to thank the Deanship of Scientific Research at CMR College of Engineering & Technology, Hyderabad for supporting this work under TIDEDST Project No. SEED/TIDE/2018/55.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. World Health Organization, "World Report on Vision" https://www.who.int/publications-detail/world-report-on-vision
2. National Programme for Control of Blindness & Visual Impairment (NPCBVI), Government of India "The National Blindness & Visual Impairment Survey India 2015–2019". https://npcbvi.gov.in/writeReadData/mainlinkFile/File341.pdf
3. Badave A, Jagtap R, Kaovasia R et al (2020) Android based object detection system for visually impaired. In: International conference on industry 4.0 technology 2020, pp 34–38. https://doi.org/10.1109/I4Tech48345.2020.9102694
4. Anitha J, Subalaxmi A, Vijayalakshmi G (2019) Real time object detection for visually challenged persons. Int. J. Innov. Technol. Explor. Eng. 8:312–314
5. Jadhav PSP, Tomy S, Jayswal SS et al (2016) Object detection in android smartphone for visually impaired users. 5:332–334. https://doi.org/10.17148/IJARCCE.2016.51171
6. Khairnar DP, Karad RB, Kapse A et al (2020) PARTHA: A visually impaired assistance system. In: Proceedings of 2020 3rd international conference on communication system computing IT application CSCITA 2020, pp 32–37. https://doi.org/10.1109/CSCITA47329.2020.9137791

7. Sait U, Ravishankar V, Kumar T et al (2020) Design and development of an assistive device for the visually impaired. Procedia Comput. Sci. 167:2244–2252. https://doi.org/10.1016/j.procs.2020.03.277
8. Kuriakose B, Shrestha R, Sandnes FE (2020) Tools and technologies for blind and visually impaired navigation support: a review. IETE Tech Rev (Institution Electron Telecommun Eng India) 1–16. https://doi.org/10.1080/02564602.2020.1819893
9. Manjari K, Verma M, Singal G (2020) A survey on assistive technology for visually impaired. Internet of Things 11:100188. https://doi.org/10.1016/j.iot.2020.100188
10. Chen LB, Su JP, Chen MC, et al (2019) An Implementation of an intelligent assistance system for visually impaired/blind people. In: 2019 IEEE International conference on consumer electronics ICCE 2019, pp 4–5. https://doi.org/10.1109/ICCE.2019.8661943
11. Lin Y, Wang K, Yi W, Lian S (2019) Deep learning based wearable assistive system for visually impaired people. In: Proceedings of 2019 International conference on computer vision workshop ICCVW 2019, pp 2549–2557. https://doi.org/10.1109/ICCVW.2019.00312
12. Matusiak K, Skulimowski P, Strurnillo P (2013) Object recognition in a mobile phone application for visually impaired users. In: 2013 6th international conference on human system interaction (HSI 2013), vol 17, pp 479–484. https://doi.org/10.1109/HSI.2013.6577868
13. Matusiak K, Skulimowski P, Strurnillo P (2013) Object recognition in a mobile phone application for visually impaired users. In: 2013 6th International conference on human system interaction (HIS), pp 479–484. https://doi.org/10.1109/HSI.2013.6577868
14. Devi RA, Uthaman I, Shanthanam RR (2020) An IOT Security based electronic aid for visually impaired detection with navigation assistance system, p 6
15. Bhatnagar V, Chandra R, Jain V (2019) IoT Based alert system for visually impaired persons. Springer, Singapore
16. Anand S, Kumar A, Tripathi M, Gaur MS (2019) Human face detection enabled smart stick for visually impaired people. Springer, Singapore
17. Jain S, Sushanth D, Varsha V et al (2019) Design and implementation of the smart glove to aid the visually impaired. In: Proceedings of 2019 IEEE international conference on communication and signal process (ICCSP 2019), pp 662–666. https://doi.org/10.1109/ICCSP.2019.8698009

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 67-70
Published Online June 2021 in IJEAST (http://www.ijeast.com)

OBJECT DETECTION SYSTEM FOR THE BLIND WITH VOICEGUIDANCE

Miss Rajeshvaree Ravindra Karmarkar
Department Of Computer Science And Engineering Walchand College Of Engineering
Sangli, India

Prof.V.N. Honmane
Department Of Computer Science And Engineering Walchand College Of Engineering
Sangli, India

***Abstract*—As object recognition technology has developed recently, various technologies have been applied to autonomous vehicles, robots, and industrial facilities. However, the benefits of these technologies are not reaching the visually impaired, who need it the most. This paper proposed an object detection system for the blind using deep learning technologies. Furthermore, a voice guidance technique is used to inform sight impaired persons as to the location of objects. The object recognition deep learning model utilizes the You Only Look Once(YOLO) algorithm and a voice announcement is synthesized using text-to-speech (TTS) to make it easier for the blind to get information about objects. As a result, it implements an efficient object-detection system that helps the blind find objects in a specific space without help from others, and the system is analyzed through experiments to verify performance.**

***Index Terms*—object detection, TTS technology, YOLO algorithm, position estimation**

I. INTRODUCTION

With the recent rapid development of information technology (IT), a lot of research has been carried out to solve inconveniences in everyday life, and as a result, various con- veniences for people have been provided. Nevertheless, there are still many inconveniences for the visually impaired. The greatest inconveniences that a blind person feels in everyday life include finding information about objects and indoor mobility problems. Previous studies included object analysis using ultrasonic sensors. However, with these methods, it is difficult to know exactly where an object is located, especially in the presence of obstacles. In this paper, we analyze accurate object information and obtain a location using a deep learning object recognition technique. Object recognition algorithms are designed based on the You Only Look Once architecture, an object recognition deep learning model, to detect objects using a camera. In this dissertation, system analyzes accurate object information and obtain a location using a deep learning object recognition technique. In addition, voice guidance technologies are synthesized so the visually impaired can know the location of the objects with voice guidance. In the voice guidance technology, the technique of synthesizing the position of the object so it can be output, and synthesizing the name of the object, is done by using text-to-speech (TTS).

II. LITERATURE SURVEY

[1] This paper explained the current methods for detection models and also gave the standard datasets. This paper coveredddifferent detectors like one-stage, two-stage which helped in analyzing different object detection methods and also gatheredsome traditional as well as new applications. It also listed somebranches related to object detection. Also pointed out some development trends to better follow the set of art algorithmand further process

[2] This paper presented region based fully convolution network. R-FCNN is used for precise and efficient object detection. Thus this paper can easily adopt ResNets like fully convolution image classifier backbones for detecting the object. This paper provided a simple but efficient R-FCNN framework for object detection. This method achieves same accuracy as compared to faster R-FCNN. Thus this helped to adopt the state of art image classification backbones.

[3] This Challenge is considered as the guideline for object classification and object detection .This paper classified and detected more 100+object categories and 1M+ images. This paper provided the large scale data collection process. And also described the most efficient algorithm for this data and also discussed the success and failure of other algorithms.

[4] The result of this research showed that grids of

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 67-70
Published Online June 2021 in IJEAST (http://www.ijeast.com)

oriented gradient performs better than current feature set for human recognition.

III. METHODOLOGY

A. Preprocessing Image:

•Read image: -Very first step of image processing is readthe training as well as testing images.

•Converting color image to grayscale image: -color isn't required to recognize and detect an image -Color images are represented in three channels (i.e. Red, Blue and Green) perhapsgray image contains only one channel. - converting to gray scale image reduces time consumption for processing imageas well as extra space required to store in memory

•Standardize Images:-Images must be resized to basic heights and widths before they are fed to algorithm. -for that standardizing is done in case of preprocessing and scaling of image.

Fig. 1. System Design

B. Identifying objects in image

•YOLO(You Only Look Once): YOLO algorithm predicts from which class object is from and also helps to give the location of object Following components of bounding boxes are used to get the location of object center, width, height, value c defines which class object is from and In addition pc value is used to get the possibility of object present in the bounding box.

In this algorithm only region containing object is not taken into account but the whole image is divided into grids. This is an algorithm used for object recognition Other algorithms such as variants of cnn does not predict the location of object. But in case of YOLO it useful to predict the location using bounding boxes. Also YOLO can detect multiple objects in a single image unlike other algorithms which classifies objects into categories only. YOLO uses single neural network on the image and not uses multiple multiple neural network.

Fig. 2. YOLO Architecture

fig 2 represents the architecture of YOLO Algorithm.

There are 2 FC layers with 24 convolutional layers in detection n/w. That makes alternating covolutional layer. These conolutional layers decrease the feature space of previous layers.

In this algorithm, network divides the images into mxm grids which is also called as regions.

Further this algorithm decides the bounding boxes and also gives the class probability map which predicts the possibility of object present in certain grid.

These bounding boxes contains weight depending on how much object is present in the region.

YOLO estimates the type and location of objects using regression inference on the problem of area selection and classification.

YOLO processes 45 frames per second so it is incredibly fast and also efficient as it predicts more than one objects fromone image.

•Bounding Box Regression: Bounding Box is used to high- light the object

Here,

Pc:Probability of object present in region bx,by,bh,bw: Dimension of bounding box c1,c2,c3:Classes of predicted object

YOLO uses above dimensions to predict the label of object.

•Intersection Over Union(IOU): It determines how predictedbox overlaps on determined box. YOLO ensures that these twoboxes are overlapped equally.

•Combination of three techniques: Given image shows how three techniques are combined by YOLO.

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 67-70
Published Online June 2021 in IJEAST (http://www.ijeast.com)

Fig. 3. YOLO Algorithm

As Shown in fig 3, it will divide the input image in S xS grid, then on each grid edge detection filters (Canny Filter) will be applied. With Edges and their connectedness bounding boxes (The rectangular box in which object is enclosed) are generated. Features of each grid in particular bounding box is concatenated and object is predicted using trained model Dataset: In YOLO we will train our object detector model on COCO dataset.

Dataset: The COCO dataset consists of 80 labels and 330kimages. Dataset consists labels like:

-Animals
-Flowers
-People
-Traffic signals
-buildings
-devices like mobile phone, computer, laptop, etc.
-Food items
And many like this…

C. Estimating the position of image:

To estimate the position of image, we need to create the bounding box for every detected object. Then with the height width of particular bounding box, with respect to image frame.

5 values are used to estimate the position of object in bounding box.

The first 4 values bx, by, bw, bh shows the position of object.

The 5th value is BC: defines how much box contains an object

BC is calculated as multiplication of object exist in box andIOU (i.e Intersection Over Union)

As labelled data should be passed forward for training ’y’is considered as label and it can be defined as,

•Triangle Similarity [11] As object gets closer to the camera its width increases and angle at the camera. triangle similarity has formula as:

F=(PxD)/W

As continue to move the camera both closer and farther away from the object/marker, applied the triangle similarity to determine the distance of the object to the camera:

D’ = (W x F) / P

D. Generating audio signal for identified object:

Voice guidance is a function that conveniently provides information to specific users, such as people who are visually impaired, when the system finds the location of the object theywant to find

The identified class label of detected object and the estimated portion of object is concatenated as text and for voice feedback Google Text-to-Speech API can be used

Google Cloud TTS API:- It allows users to merge hundreds of natural voices

-It is also available in multiple languages.

-It is an easy to use api as developer can interact with user with multiple languages, applications and devices.

IV. Result and Discussion

In this study we have developed and android application forobject detection

Fig. 4. Object detection with Position

The system displays output as the name of the object and the probability of the object in percentage. So the system will only detect those objects having probability greater than giventhreshold probability.

As well as system uses google API to convert text into speech and gives the position of object in inches using device’sspeakers.

tiny YOLO is used in order to implement the system on android platform as a result it reduces the accuracy of object detection.

Fig. 5. Android Application screenshot

Fig represents the snapshot of android application. It displaysthe name of object and its probability.

whilst the application also tells the class label and how far the object from camera through device's speakers.

V. Significance

Current system provides help for blind people.

-This system can be embed in mobile phones or any similarhandy devices.

-This system will help visually impaired people to know theinformation (like name and location) of object.

VI. Conclusion and future work

Develop the object detection system with the help of deep learning algorithm such as YOLO and estimate the positionof object with the help of same.

-This system provides the voice guidance for visually impaired people. This system is solely designed to help the blind people effectively.

However the accuracy can be increased. Also the current sys- tem works on android Operating system which can be further developed in order to be compatible on every convenience devices.

VII. Acknowledgement

This paper is inspired from "Object Detection System for the Blind with Voice Command and Guidance" presented from Wonju National University / Gangneung, Gangwon 25457, Korea.

References

[1] Choi D., and Kim M. ' , "Trends on Object Detection Techniques Based on Deep Learning,." *Electronics and Telecommunications Trends*, Vol. 33, No. 4, pp. 23-32, Aug. 2018.

[2] J. Dai et al.' , "R-FCN: Object Detection via Region-based Fully Con- volutional Networks." *Conf. Neural Inform. Process. Syst., Barcelona, Spain,*Dec. 4-6, 2016, pp. 379-387

[3] Dalal N. and Triggs B., ' , "Histograms of Oriented Gradients for Human Detection,." *IEEE Comput. Soc. Conf. Comput. Vision Pattern Recogn., San Diego, CA, USA,*,June 20-25, 2015, pp. 886-893

[4] O. Russakovsky et al., ' , "ImageNet Large Scale Visual Recognition Challenge,." *Int.J. Comput. Vision,*Vision, vol. 115, no. 3, Dec. 2015, pp. 211-252.

[5] Ahonen T., and Pietika¨inen M., ' , "Face Description with Local Binary Patterns: Application to Face Recognition." *" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,,* , VOL. 28, NO.12, DECEMBER 2006

[6] Raj A.,Kannaujiya M.,Bharti A.,Prasad R.,Singh N.,Bhardwaj I., ' ," Model for Object Detection using Computer Vision and Machine Learning for Decision Making ,." *International Journal of Computer Applications (0975 – 8887),*Volume 181 – No. 43, March 2019

[7] W. Liu et al., ' , "SSD: Single Shot MultiBox Detector,." *European Conference on Computer Vision,,*Vol. 9905, pp. 21-37, Sept. 2016

[8] J. Redmon et al., ' , "You Only Look Once: Unified, Real-Time Object Detection." *in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR),*, pp. 779-788, Jun. 2016

[9] Moonsik K., ' , "Object Detection System for the Blind with Voice Command and Guidance." *IEIE Transactions on Smart Processing and Computing,* vol. 8, no. 5, October 2019

[10] YOLO: Real-Time Object Detection, https://pjreddie.com/darknet/yolo

[11] Triangle Similarity https://www.pyimagesearch.com/

International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 8 Issue 5, Sep-Oct 2024 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Object Detection and Recognition Using TensorFlow for Blind People

Darshan A. Mirapure[1], Dipam Pakhamode[2], Prof. Rina Shirpurkar[3]

[1,2]School of Science, G H Raisoni University, Amravati, Maharashtra, India
[3]Assistance Professor, G H Raisoni University, Amaravati, Maharashtra, India

ABSTRACT

Computer Vision impairment or blindness is one such top ten disabilities in humans, and unfortunately, India has the world's largest visually impaired population. For this we are creating a framework to guide the visually impaired on object detecting and recognition, so that they can navigate without others support, and be safe within their surroundings. In this system the captured image is taken and sent it as input using camera. SSD Architecture is used here for the detection of objects based on deep neural networks to make precise detection. This input will be given to the software and it will be processed under the COCO datasets which are predefined in the Tensor flow library used as training dataset for the system in general this data set consist of features for nighty percent of real world data objects and distance is calculated by depth estimation and also by using voice assistance packages the software will produce the output in the way of Audio. The System is implemented completely using Python Programming Language since python consist of many inbuilt packages and libraries which will make the complication of writing code more number of lines into simple any less number of lines.

KEYWORDS: *Object Detection, Tensor Flow, COCO Datasets*

How to cite this paper: Darshan A. Mirapure | Dipam Pakhamode | Prof. Rina Shirpurkar "Object Detection and Recognition Using TensorFlow for Blind People" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-8 | Issue-5, October 2024, pp.747-752, URL: www.ijtsrd.com/papers/ijtsrd69446.pdf

IJTSRD69446

Copyright © 2024 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0)

I. INTRODUCTION

The fast progress of data and organize technology has advanced from the Internet and computerization systems that were initially utilized for authoritative offices and mechanical and economical applications to the application of these innovations every one's life. Once you think of technology like augmented reality, one of the key components to consider is object acknowledged innovation, moreover known as object detection. This term specifies to a capacity to distinguish the frame and shape of diverse objects and their position in space caught by the camera. It's a known reality that the number of visually disabled individuals within the world is almost more than 280 million, roughly break-even with the 25% of the Indian population. They suffer-normal and difficult challenges in regular activities particularly when they are on their own. They are generally dependent on somebody for gettingto their day-to-day works.

So, it's a very challenging and the nonphysical arrangement for them is of most extreme significance and much required. One such solution from our sideis that we came up with an Machine Learning Framework permits the blind activities to distinguish and classify general Time Based day-to-day object and produce voice outputs and calculates distance using mathematical calculationswhich produces alerts whether user is exceptionally near or far absent from the source. The same framework can be used for Obstacle Detection Instrument. The primary reason for object detection is to find different things, which draw rectangular bounding boxes around them, and decide the course of each item found.

Applications of object discovery emerge in large no of diverse domains counting recognizing people on foot for self-driving vehicals, checking crops, and indeed real-time ball following the basket.

TensorFlow, an open-source machine learning framework developed by Google, provides robust tools and libraries specifically designed for building and deploying object detection models. With its TensorFlow Object Detection API, users can access a

variety of pre-trained models, facilitating rapid development and fine-tuning for specific tasks.

This framework simplifies the entire workflow, from data preprocessing and model training to and inference. TensorFlow's integration with Keras also allows for easy model customization and experimentation, making it an ideal choice for both beginners and experienced practitioners.

II. OBJECTIVE

The project goal is to incorporate an art of techniques for object detection to achieve high accuracy with realtime detecting performance. In this project, we use Python programming language with an TensorFlow-based solution for solving the problem of object detection in an end-to-end solving proposed system will be fast and effiecient. A TensorFlow based application approach for an mobile device, using its built-in hardware component camera is used for detecting objects, more specifically: The framework is prepared in such a way where an mobile application (assuming you're using it on an Android/ios device) will capture real-time frames and will send them to the backend of the application where all the predefind computations takes place.

Fig 1: Objects for object recognition which consist a dog and a duck on the beach

- Along with the object finding, we have used an alert framework where distance will get calculated. In case the Blind Person is especially close to the object or is far away at a safe put, it'll produce voice-based outputs yields alongside distance units. The backend of the application is where the video clip is sent and is taken as an input, which goes through
- Victim is especially close to the source or is distant far away at a more secure place, it'll generate voice notes alongside distance measure units.
- The COCO DATASETS object detection model one of the datasets predefined and which tests and detects with accurate metrics. After testing output of the application will sent to voice modules the course of the object will be changed Into default voice notes which can at that point be sent to the users for their needs.
- Along with the object finding, we have an alert voice framework where figure out distance. In case the Blind victim is especially close to the source or is distant far away at a more secure place, it'll generate voice notes alongside distance measure units

III. LITERATURE SURVEY

OBJECT DETECTION USING CONVOLUTIONAL NEURAL NETWORK

In 2019, "Object Detection using convolutional Neural Networks". As Vision systems are essential in building a mobile robot. That will complete a certain task like navigation, surveillance, and explosive ordnance disposal (EOD). Vision systems are essential in building a mobile robot. A project was proposed based on CNN, which is used to detect objects in the environment. Methodology used- Two state of art models are compared for object detection, Single Shot Multi-Box Detector (SSD) with MobileNetV1. Another methodology is A Faster Convolutional Neural Network (Faster-RCNN) with the help of InceptionV2.

IMAGE BASED REAL TIME OBJECT DETECTION ALONG WITH RECOGNITION IN IMAGE PROCESSING

In 2019, "Image Based Real Time Object Detection and Recognition, In Image Processing" Object detection and tracking mainly for human and vehicle is presently most active research topic. It is used in applications such as surveillance,image retrieval. A solution was proposed which has reviewed recent technologies for each phase of the object detection. The methodology used here is four different methods for object detection which is nothing but a computer technology related to computer vision with image processing that deals with detecting instances of semantic objects of a certain class in digital images and videos and, they are feature based detection, region based detection outline based detection illustrations and model based detection. applications in everyday life. It is used in applications such as surveillance, image retrieval.

A solution was proposed which introduced a new fast method for saliency object detecting within images. The main aim was detection of objects in complex images. The methodology used has four steps: regional feature extraction, segment clustering, saliency score computation and post-processing.

REAL-TIME OBJECT DETECTION USING DEEP LEARNING

In 2020, "Real-Time Object Detection Using Deep Learning". Object detecting, recognition in images and videos is one such major thing today. For this a solution was proposed using deep learning. The methodology used here includes feature extraction with the help of Darknet-53 along with feature map up sampling and concatenation. Model includes various changes in object detection techniques.

ASSISTIVE OBJECT FINDING SYSTEM FOR VISUALLY IMPAIRED PEOPLE

In 2020 ,"Assistive object Recognition/finding System for visually impaired" The issue of visual impair or blind people is faced worldwide, for this a solution was proposed where two cameras placed on blind person's glasses, GPS free service, and ultrasonic sensors are employed. To give information about the surroundings.

The methodology used here is system takes real-time images as input, thenimages are pre-processed based on the job, their background and foreground work are separated and then the DNN module with the help of pre-trained YOLO model is applied for resulting in featured extraction.

IV. PROBLEM FINDING

The Populated number of people visually impaired in the world is more than 290 million. In this 42% are blind and 58% have no vision. They are an important part of our society. It's very difficult for them to live the outside world independently. Today in the fast moving society, visually impaired people require supportive instruments in their day-to-day life. Our thought primarily centered on developingan assistive framework for impaired people to detect objects effectively which can be helpful to live.

PROBLEM DESCRIPTION

The system is designed in such a way where an mobile application will capture real-time objects and will send it to a laptop based networked server where all the important computations take place and utilizing a pretrained SSD detection model which is trained on COCO datasets the objects will detect and recognized by the system. After that distance will be calculated and the output for this will bein audio form where the system gives warnings with calculated distance.

V. EXISTING SYSTEM

Most of the computer vision systems exist now-a-days to help the people who arevisually impaired in their life. These include technological Augmented Reality approached wearable goggles, video calling applications for the visually impaired to ask for assistance, AI and GPS based navigation systems, etc. These systems aredeveloped to work in specific cases or conditions, and cannot be used broadly. There are cases wherein the people with visual impairment have to accept about their surroundings, which is not possible with the existing systems.

DISADVANTAGES OF EXISTING SYSTEM

- They are expensive. Most of the visually impaired people (assume single person) cannot afford such highly economical products.
- These systems may be complex in functionality, making it difficult to be used bythe blind people.
- Some systems are not real-time.
- Developing and training object detection models can be complex and resource-intensive, requiring substantial expertise in machine learning and computer vision.
- High-quality, annotated datasets are essential for training effective models. Collecting and labeling this data can be time-consuming and costly.
- Object detection systems can struggle with accuracy in challenging conditions, such as poor lighting, occlusions, or variations in object appearance.
- Achieving real-time detection can be computationally demanding, requiring powerful hardware, which may not always be feasible for all applications.
- Models can inherit biases from the training data, leading to inaccuracies in detecting certain objects or demographics, which raises ethical concerns.
- The initial investment in technology and ongoing maintenance can be high, especially for small businesses or startups.
- In applications like surveillance, object detection can raise privacy issues, as it often involves monitoring individuals without their consent.

VI. PROPOSED SYSTEM

In this proposed system, we are using Python with an Tensor Flow-based approach to find the solution for the problem of object detection in an end-to- end fashion. We used SSD Detection Model for the detecting of items based on deep neural networks to make effective detection and OpenCV library for real time picture capturing. Among ImageNet, Google Open, COCO datasets we are using COCO since it will provided class of classified feature for more than 90% of the real world objects. The image is sent as an input to the model and meanwhile distance is calculated using depth estimation with the help of

voice modules predefined by python the output of the object name will be converted into default voice notes which are sent to the blind people for their help with calculation.

User starts the System, after that the system will activate the camera and capture instant real time images which will be considered as input. After capturing imagesit will store and send to the dataset where using SSD Architecture the internal computations will takes place mean while after the computations the model will detect the object and recognition will be done.

Fig 2 System Design

After detection of object next it will be displayed on the monitor where frames are captured to the detected object along will the labels. Next distance will be calculated using depth estimation by finding mid ranges to the frames. Now using speakers which are based on voice module packages the detected object images will be read as text and it will be the output.

➢ ADVANTAGES

- Easy to use
- Provides real-time results and this result is in the voice format with distance.
- Depending on the video quality, difference between various objects like chairand table etc can be easily differentiated .
- Due to usage of COCO datasets it will provide the 90% of results efficiently.
- It enables automated systems to identify and analyze objects in real time, streamlining processes in industries like manufacturing, security, and logistics.
- It allows for efficient analysis of large datasets, extracting valuable insights from images or videos for applications in surveillance, retail, and agriculture.
- By automating tasks like inventory management or quality control, object detection can reduce labor costs and minimize errors.
- It can assist visually impaired individuals by identifying objects in their environment through audio feedback.
- In marketing, object detection can be used to tailor advertisements and content based on user behavior and interactions with specific products.
- In applications like augmented reality and mobile apps, object detection enhances user interactions by recognizing and responding to real-world objects.
- If a model is too closely trained on specific data, it may not generalize well to new, unseen environments or variations of the objects.
- The initial investment in technology and ongoing maintenance can be high, especially for small businesses or startups.
- In applications like surveillance, object detection can raise privacy issues, as it often involves monitoring individuals without their consent.

VII. MODULES

Video Capturing Module:

When the system is turned on the system capture images using camera. We have to connect this as input to the COCO dataset and classification of pixels and features takes place. The captured frames can be seen in the monitor with drawn boundaries and label. The

method videocapture( ) is used to start the camera and capture the video.

Image Processing Module:
OpenCV (Open Source Computer Vision) is a library in python which functions mainly aimed at real-time computer vision. It is mainly used to do all the computational operation related to images. cv2 is used to perform image processing and make use of methods which are used to detect and capture the frames and specifies names. This module is processed after the input is taken from the camera.

Object Detecting Module:
The algorithm will take the image as input and all the computations will take place like divding the image into neurons nothing but pixels and classification of features which will be done on Neural Network. Image will be read as string for the next computation and it will be compared under trained dataset. This can be achieved here by using category index where 90 objects are trained separately. Here we used SSD Architecture which comes under Tensor Flow API .

Distance calculation Module:
To find the distance of the object numpy is used, which is pip package used for mathematical calculation. Finding distance can be approach by using depth estimation, using detected objects visible on the monitor frames the depth estimation will take place by finding mid ranges and rounding the estimation scaleto 0-10.

TensorFlow Object Detection API:
This is a powerful framework specifically designed for object detection tasks. It provides pre-trained models, utilities for training custom models, and tools for evaluation and visualization.

TensorFlow Core:
Basic TensorFlow libraries and functions are used for building, training, and evaluating neural networks. This includes layers, optimizers, and metrics.

TFRecord:
A data format optimized for TensorFlow that is commonly used for storing training data, including images and their corresponding annotations.

Audio Output Module:
Next after detecting the object and calculating the distance our aim is to give the output in the audio using voice notes. In the output we are going to specify the distance along with units and the warning messages to alert the user. For audio output the pyttxs3 pip package which is predefinded python built in module used for converting text to speech.

Fig 3: Object Detection

VIII. CONCLUSION

Previous studies have proposed a number of methods to detect object. After doing literature survey, different techniques has been found for detecting and Recognition of Object and they use different types of data as input for their methodology. After the survey of different types of methods, it is found that using SSD Architecture model which was trained under COCO datasets is the easy method which can be easily applied and appropriate in all conditions. We decide to explore this method of computer vision and proposed a noble method to detecting and recognize the objet based on Tensor flow and finding distance, sending output through voice assistance like speaker, by this blind person can live without depending on others for their day to day life on detection and recognizing the object and will alerted because of voice outputs. As per future work we are willing to make an application software for the IOS devices.

IX. REFERENCES

[1] Aditya Raj, "Model for Object Detection using Computer Vision and Machine Learning for Decision Making," International Journal of Computer Applications, 2019.

[2] Bhumika Gupta, "Study on Object Detection using Open CV Python," International Journal of Computer Applications Foundation of Computer Science, vol. 162, 2017.

[3] Abdul Muhswin M, "Online Blind Assistive System using Object Recognition," International Research Journal of Innovations in Engineering and Technology, vol. 4, pp. 49-51, 2018.

[4] "OpenCV," [Online]. Available on: www.opencv.org.

[5] "Python language," [Online]. Available on: www.python.org.

[6] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), “An Analytical Perspective on Various Deep Learning Techniques for Deepfake

Detection", *1st International Conference on Artificial Intelligence and Big Data Analytics (ICAIBDA)*, 10th & 11th June 2022, 2456-3463, Volume 7, PP. 25-30

[7] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), "Revealing and Classification of Deepfakes Videos Images using a Customize Convolution Neural Network Model", *International Conference on Machine Learning and Data Engineering (ICMLDE)*, 7th & 8th September 2022, 2636-2652, Volume 218, PP. 2636-2652, https://doi.org/10.1016/j.procs.2023.01.237

[8] Usha Kosarkar, Gopal Sakarkar (2023), "Unmasking Deep Fakes: Advancements, Challenges, and Ethical Considerations", *4th International Conference on Electrical and Electronics Engineering (ICEEE)*,19th & 20th August 2023, 978-981-99-8661-3, Volume 1115, PP. 249-262, https://doi.org/10.1007/978-981-99-8661-3_19

[9] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2021), "Deepfakes, a threat to society", *International Journal of Scientific Research in Science and Technology (IJSRST)*, 13th October 2021, 2395-602X, Volume 9, Issue 6, PP. 1132-1140, https://ijsrst.com/IJSRST219682

[10] Usha Kosarkar, Prachi Sasankar(2021), " A study for Face Recognition using techniques PCA and KNN", Journal of Computer Engineering (IOSR-JCE), 2278-0661,PP 2-5,

[11] Usha Kosarkar, Gopal Sakarkar (2024), "Design an efficient VARMA LSTM GRU model for identification of deep-fake images via dynamic window-based spatio-temporal analysis", Journal of Multimedia Tools and Applications, 1380-7501, https://doi.org/10.1007/s11042-024-19220-w

[12] Usha Kosarkar, Dipali Bhende, " Employing Artificial Intelligence Techniques in Mental Health Diagnostic Expert System", International Journal of Computer Engineering (IOSR-JCE),2278-0661, PP-40-45, https://www.iosrjournals.org/iosr-jce/papers/conf.15013/Volume%202/9.%2040-45.pdf?id=7557

IJTSRD International Journal of Trend in Scientific Research and Development ISSN: 2456-6470

iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: II Month of publication: February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49107

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Object Detection System with Voice Alert for Blind

Dr. M Y Babu[1], Akash Jatavath[2], G Yashwanth Kumar Reddy[3], Pittala Arun Kumar[4]
[1]*Associate Professor, Dept of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad*
[2, 3, 4]*UG Student, Dept of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad*

Abstract: As we can see, there are numerous blind persons nearby who encounter various challenges, such as difficulty in crossing roads and identifying objects in their environment. With the advancement in technology in several fields, human life is evolving to better standards. Unfortunately, those who are blind are unable to fully enjoy this kind of lifestyle. So, this project is one strategy for introducing blind individuals to a new way of living that makes them independent on others. The major goal of this project is to create a deep-learning algorithm that can be used to analyse the environment for people who are blind by using the rapidly evolving technology. We'll accomplish this using object detection and transform the data into speech alerts and warnings. Real-time object detection is one of the more challenging tasks since it requires continuous processing and takes a long time. The convolution neural network is the main backbone for any type of object detection (CNN). We can create algorithms based on photos and videos by employing a convolution neural network. We utilise the YOLO technique for object detection because it is simple and quick to process. In addition, for the voice warnings, we employed Text to Speech (TTS). The dataset used in this technique is the COCO dataset, which contains the names of things and objects in our daily lives. These algorithms have been thoroughly trained by the over 90 outdoor objects that we view every day in our daily lives.
Keywords: Deep Learning; YOLO; TTS; CNN

I. INTRODUCTION

One of the most important organs in human body are eyes. We enjoy the beauty of nature, various types of books, and many other aspects of our lives. We can go anywhere independently and have fun with friends and family . What if we are blind? Forgot about enjoying, what if we don't even do our own work independently? What if we must depend on some others for regular daily works? It is difficult to think and imagine these kinds of situations. However, some of us in society are visually impaired. They must depend on others for their regular work. The ability to visualize the surroundings is a gift.

The visually impaired people face many difficulties in their day-to-day life. They face many difficulties in object detection and analysing of their surroundings. While Walking on the streets, they face many difficulties in identification and recognizing the objects .These cause them many injuries and accidents etc . So, to put an end to these difficulties we came with an idea of recognizing the objects around them by using object detection and converting them in to voice messages to identify and understand the situations around the person. The use of technology in finding and recognizing objects is huge because of these rapid increase in the technology like AI, ML and Deep Learning etc facilitate many tools and libraries for the development new ideas which are useful in contemporary society like smart sticks, navigators etc.

Huge number of research and developments are going in the domain of machine learning and object detection .Large number of new kind of tools are also came in to the existence .Few of those developments are similar to our idea .But all those projects implementation has distinctions and differences in object detection like using of different algorithms and different libraries for the processing .Our dataset contain nearly 90 object names which are useful and observed by a common man in our day to day life , Which is enough for the real time object detection. we use YOLO algorithm for object detection and Text to Speech conversion technique for voice alerts.

II. RELATED WORK

A. Literature Survey

1) Prof. Seema Udgirkar, Shivaji Sarokar, Sujit Gore, Dinesh Kakuste, Suraj Chaskar, "Object Detection System for Blind People"

The author of this work attempts to demonstrate how their smart vision, the goal of which is to move about the environment using a user-friendly interface system, was proposed. These writers created a system that can detect obstacles that are close to his head, particularly while entering through a door.

Simply put, it is designed to protect his head from harm. This item is made to help blind people traverse any area. It uses the user's buzzer and vibration as two output modes to direct the user toward an object and provide information about an impediment. There are two different operating modes: buzzer mode and vibration mode. These outputs are made available for blind people.

2) Jigar Parmar, Vishal Pawar, Babul Rai, Prof. Siddhesh Khanvilkar," Voice Enable Blind Assistance System -Real time Object Detection".

In this study, authors tried to identify a presented object in front of a webcam. They trained and tested the TensorFlow Object Detection API frameworks using made a model. To alleviate Input / Output concerns, a good frames per second solution is needed because reading a frame from a web camera causes a lot of issues. As a result, they concentrated on threading technology, which dramatically reduces processing time for each item while improving frames per second. The item detected box takes about 3-5 seconds to move over the next object in the video, even though the application correctly identifies everything in front of the webcam.

III. METHODOLOGY

A. Object Recognition

Though similar approaches to object identification, object discovery and object recognition operate differently. Even though both are widely used for images and video. Object detection is considered a subset of object recognition in processing. Object detection and recognition are often employed in a wide range of sectors, from personal security to workplace productivity. It is used for autonomous driving systems, machine inspection, surveillance, security, and image retrieval, among other computer vision applications. In general, the text to speech conversion capability cannot be recommended for non-operating system devices. Thus, Android- or iPhone-based smartphones are the most popular choice among smart phone users who are blind or visually impaired.

Object discovery is the phenomenon of finding instances of items in both still images and films that contain the objects in question. Bounding boxes and information about the identified items' locations inside the frame are highlighted. Technology related to both image processing and computer-aided vision is object detection. It classifies and identifies a wide range of things from movies and digital photos, including people, animals, and cars. Multiple things in a video or digital picture can be swiftly classified using object detection. Although object detection has been around for a while, it is currently more prevalent than ever across a variety of sectors. The object detecting system has been put into action using a variety of techniques.

1) YOLO Algorithm

The YOLO algorithm was initially proposed by Joseph Redmon and his colleagues. In 2015, he released a paper on YOLO under the heading "You Only Look Once" Real-Time recognition, and it became immensely successful right away. CNN is followed by YOLO. When making predictions, the algorithm only "looks once" at the image since there is only one propagation that occurs throughout the neural network. Compared to other methods of object identification, the YOLO model is the fastest and most effective. The main benefit of YOLO is its quickness. There are 45 frames per second in this. The model is constructed in a concise manner to acquaint its network with an abstract description of things. The primary goal of object detection is to identify one or more specific things in audio visual or digital pictures. Contrarily, object class recognition classifies items into a certain category or class. Every thing has unique qualities that make it easier to distinguish it from other objects in movies or pictures. Additionally, it sets them apart from other classes. Object detection is the process of finding and defining objects, such as people, animals, objects, cars, and so on. Combining the You Only Look Once (YOLO) architectural algorithm with the COCO dataset results in a quick and effective deep learning technique for object recognition.

Figure 1: Yolo architecture

YOLO is made for comprehensive image processing and steadily raises the effectiveness of object detection. Frame identification is seen as a regression issue in this situation. In order to quietly store specialised information about groups and their looks, YOLO employs the entire background throughout training and testing periods while the networked is concentrated on recent images.
To continuously predict all bouncing boxes throughout all groups for a picture, it uses features from the full image. The method divides the informational image into a SxS example. The matrix cell can identify the point and choose the certainty scores for those containers when the focal point of an object falls within a network cell.

B. Estimation of Image Position

We must generate a bounding box for each identified object in order to approximate the position of the image. Using the specific bounding box's height and width in relation to the image frame. To estimate the position of an object within a bounding box, 5 values are used. The position of the item is shown by the first 4 values: bx, by, bw, bh, BC the fifth value, specifies how much of a box an object occupies.

$$BC=pr*IOU$$

Where IOU = Intersection over Union

Pr = Object existing in box

This estimates the predictability and likelihood of the box containing an object of any class. If there is no object in the box, BC=0, otherwise BC=1.

C. Voice Generation

When the system locates the desired object, voice guiding is a feature that offers information to specific users, such as those who are blind, in a convenient manner. It is crucial to alert the blind person heading in the way of the presence of an object when it has been discovered.PYTTSX3 is an essential part of voice generation module. Text to speech conversion can be done using the Python library Pyttsx3. Python version 2 and 3 are both compatible with this package. A straightforward tool for text to speech conversion is Pyttsx3. We also used Google Text to Speech (GTTS) for voice alerts. Google Text to speech contain many inbuilt English accents for the users who are from different parts of the world. It is very easy to use, it converts the text into audio which can be saved as a mp3 file It also supports many regional languages which is also useful for those who do not able to understand English.

1) *Dataset:* Here we used COCO dataset. COCO stands for common object in context. The COCO dataset contains challenging, high-quality visual datasets for computer vision. The images in the dataset were gathered from commonplace locations that provided specific information. In real-world situations, multiple items or things may be contained within the same frame, and each one needs to be distinguished as a distinct object and properly segmented. The identification and segmentation of the objects visible in the photographs are contained in the COCO dataset. We used this information to develop our item recognition and detection technology for persons with disabilities. This dataset contains approximately 90 objects or items.
2) *Process Model:* The working of this system is represented in the below process model. Input is taken from the user's camera to capture the images. The system checks if any objects are detected in the image. If an object is detected, the system identifies the object. Then the system calculates the distance of the object from a person. Based on the calculated distance, the system generates an audio output.

IV. RESULTS

Figure 3: Detection of a person with 86% Accuracy

Figure 4: Detection of a person with 84% Accuracy

Figure 5: Detection of a Bottle with 63% Accuracy

Figure 6: Detection of a Bottle with 56% Accuracy

V. CONCLUSION AND FUTURE SCOPE

In this project we used image recognition, voice generation modules for the development of the project. As of now accuracy is good but in case if we want to increase the accuracy we have to train the model with more object/images in the dataset. This project is a small experiment which is useful for blind persons, can be able to find the objects which are surrounded by them, and they are in a position of taking care of themselves when they are outside.

The ability of the blind person to stand alone and carry out tasks independently makes this blind assistance device useful for object detection by voice warnings. The device's camera serves as the blind person's virtual eye, capturing every detail of their environment. The voice alerts keep the person informed about the surroundings so that accidents are decreased. Reduced rely on other parties.

There are so many people present in the world who are visually impaired and illiterate from different parts of the world .Some of them do not understand other languages except their local language in their local accent . so, one of the future scope for this project is to develop the idea in which voice alerts in such a way that they can use their own local language .

REFERENCES

[1] Miss Rajeshvaree Ravindra Karmarkar, Prof.V.N. Honmane "Object Detection System For The Blind With Voiceguidance"- Published Online June 2021 in Ijeast
[2] Jigar Parmar, Vishal Pawar, Babul Rai, Prof. Siddhesh Khanvilkar "Voice Enable Blind Assistance System -Real time Object Detection"- IRJET, Apr 2022
[3] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam " Real-Time Object Detection with Yolo"- IJEAT,Feb 2019
[4] Mayuresh Banne, Rahul Vhatkar, Ruchita Tatkare "Object Detection and Translation for Blind People Using Deep Learning"- IRJET ,Mar 2020
[5] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh "Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People"- IJEER,June 2022
[6] Priya Kumari, Sonali Mitra, Suparna Biswas , Sunipa Roy, Sayan Roy Chaudhuri , Antara Ghosal , Palasri Dhar, Anurima Majumder "YOLO Algorithm Based Real-Time Object Detection"- IJIRT, June 2021
[7] N. V. N. Vaishnavi , Tummala Navya , Velagapudi Srilekha , Vinnakota Karthik , D. Leela Dharani "Blind Assistance In Object Detection And Generating Voice Alerts"- DRSR,Feb 2021
[8] Tanvir Ahmad , Yinglong Ma , Muhammad Yahya,Belal Ahmad,Shah Nazir , and Amin ul Haq "Object Detection through Modified YOLO Neural Network"- Published 6 June 2020 in Hindawi Scientific Programming
[9] Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time Object Detection",The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
[10] Dumitru Erhan, Christian Szegedy, Alexander Toshev, "Scalable Object Detection using Deep Neural Networks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147-2154
[11] N. V. N. VAISHNAVI , TUMMALA NAVYA , VELAGAPUDI SRILEKHA , VINNAKOTA KARTHIK , D. LEELA DHARANI ," BLIND ASSISTANCE IN OBJECT DETECTION AND GENERATING VOICE ALERTS" ,in 2021

10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

International Conference on Communication and Signal Processing, July 28 - 30, 2020, India

Reader and Object Detector for Blind

M. Murali, Shreya Sharma and Neel Nagansure

***Abstract*—This work aims to assist the visually impaired people for reading a text material and detect objects in their surroundings. The input is taken in the form of an image captured from the web camera. This image is then processed either for the purpose of text reading or for object detection based on user choice. The Raspberry Pi acts as the microcontroller for processing of the entire process. The text reading is supported by software named OCR. The read text is changed into an audio output using the TTS Synthesis. Other dependencies required for the process include Tesseract Library. The Object Detection is another aspect of the project which is implemented using a TensorFlow Object Detection API. It is able to detect various objects in its surroundings and provide an audio feedback about the same. The dataset can be trained on various different situations depending on the user needs, thus making it scalable**

***Index Terms*—Raspberry pi, OCR, tesseract, tensorflow**

I. Introduction

IN today's world, technology is growing at an alarming rate. It has found its way in every field of our life. But this technology is of no use if it couldn't provide itself to the aid of the disabled people. The aspect considered most important in human life is education. It is the education that you receive today shapes your tomorrow. But do blind people get the same level of education that we sighted people are privileged for? The answer is NO. Books read by blind people are scripted in Braille, but the cost of a simple book on counting shapes in Braille costs around rupees 1300 online. So imagine how much the academic textbooks would cost over the entire educational lifetime. Not everyone can afford this. Plus it takes much longer to print a Braille textbook than a regular textbook. The time and the money needed definitely fall into the expensive segment of the graph. Not everyone can afford this much time and money. Apart from time and money, the learning curve for Braille is also steep. Fig. 1 is showing the braille alphabets, it takes more training to learn the language of the blind than it takes to learn regular alphabets. This steep learning curve is another reason for Braille not being as efficient. Braille also fails to incorporate many new technological advances and innovations. It hasn't evolved much like a language. No new technology has incorporated Braille in its use case action.

M. Murali is with SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India (e-mail: muralim@srmist.edu.in).

Shreya Sharma is with SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India (e-mail: ss9090@srmist.edu.in).

Neel Nagansure is with SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India (e-mail: nn7295@srmist.edu.in).

Fig. 1. Braille alphabets

Thus there are several reasons why Braille couldn't be the most helpful language for the blind. With the aim to provide assistance to disabled people we decided to innovate something for the visually disabled or blind people of our society. The READER would be like a live tutor. It would recite the learner the entire page that he/she wishes to read. We would use computer vision coupled with the IoT technology to make what we call this, a reader for the blind. Now blind people can read whichever book they want to without spending tons of money and time to get it printed in Braille. Imagine how simple and efficient it would make their lives. The Object DETECTOR, on the other hand, will act like a pair of virtual eyes. They would detect the objects present in the field of vision using a camera. The live feed will be analyzed by Tensor Flow and trained using various datasets to maximize the detection rate. The Object Detector will pronounce the objects present in the vicinity through a speaker. This will enable the user to know exactly what is present around them.

The rest of the paper is organized as follows. The literature survey and the existing work problem are explained in Section II and Section III respectively. The proposed work and its implementation are described in Section IV and Section V respectively. The result analysis is discussed in Section VI. At last, the paper is concluded with the conclusion of the paper in Section VII. Finally future work is discussed in section VIII.

II. Literature Survey

In [1] author focused on creating a photo-to-speech application for the blind. The project is called Camera Reading for Blind People, and it aims in development of a mobile application that allows blind user to read text. To achieve this, Optical Character Recognition (OCR) and Text to Speech Synthesis (TTS) are integrated, which enables the user to take a picture using a camera and hear the text that exists in the picture.

In [2] this work focuses on the complete integration of Text to Speech system designed for blind people. The hardware

978-1-7281-4988-2/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: SRM University. Downloaded on September 10,2020 at 04:37:12 UTC from IEEE Xplore. Restrictions apply.

consists of a Webcam and a Raspberry Pi which accepts an image of the page to be readout.

The Software consists of OCR package that is responsible for the Text to Speech conversion. An audio amplifier then is fed with the output. MATLAB is used by the authors for the simulation of the proposed system. They used Libraries, Auditoriums, and Offices to read instructions and notices and assistance in filing application forms.

In [3] the hardware includes an Ultrasonic Sensor for environment scanning. A GPS service is also employed. Two Cameras are placed on the glasses that will generate a disparity map of the scene and a GPS is used to group objects based on the location. The Ultrasonic sensor detects any obstacle at a medium to long-range. The system is optimized to work effectively in a real-time system.

In [4] the system proposed in the above paper reads the text present on labels, products and printed notes. It involves the process of extraction of text from an image and its conversion to speech. The Raspberry Pi used is provided with a battery backup thus giving a portability aspect to the product.

In [5] various methods such as 'YOLO', 'SSD' and 'CNN' are used to provide the users with the best experience of object detection. However, accuracy could have been improved.

In [6] the proposed system aims to assist blind people in the detection of brightness and major colors in real-time by using RGBby the means of an external camera. It also helps the user in the identification of fundamental objects. The ability of the system to have a facial detection system also increases the scope of application. Hardware System includes the usage of a Raspberry Pi and a Pi camera for detecting the facial edges and objects. The software includes the YOLO Algorithm and MTCNN Networking.

III. Problems In The Existing Model

The existing model is capable of either detecting the objects around or reading the text from an image. No work has been done to inculcate both technologies into one single model. We realized that these two technologies can be beneficial to the blind people and its utilization to the maximum should be done by the integration of the OCR model and Object Detection model. We have also worked to improve the frame rate of object detection by limiting the training dataset to our everyday objects hence focusing more on a smooth and faster experience.

IV. Proposed Work

This project will present a document reader and object detector for blind people that is developed using Raspberry Pi. It uses the 'Optical Character Recognition' (OCR) technology to read the printed characters captured using USB camera or pi camera. OCR AND TTS (Text-To-Speech) are used to convert images of printed text into an intermediate form that is then changed to audio output. TensorFlow software library is used for Object Detetction. It can successfully detect majority of objects in our surroundings including animals, vehicles, humans etc.

Total visually impaired population – 285 million

Fig. 2. Distribution of Visually Impaired People across the world.

In the Fig. 2, the spread of visually impaired people in different areas of the world graph is shown. While the majority of it lies in South-Asian continent, lesser numbers are witnessed in Middle East & North Africa. Sub-Saharan Africa and Europe have 47.26% and 44.75% respectively.

Fig. 3. Procedure of the Proposed System

The procedure of the proposed system is shown in Fig. 3. Choose which module needs to be performed. If the text needs to be read, webcam clicks a photo of the printed text and fed into raspberry pi which converts it into audio output using libraries such as OCR and TTS. If objects are to be detected, webcam displays the name of the object around the rectangle that surrounds the object and that text is converted into audio output. This is achieved using Tensor Flow.

Authorized licensed use limited to: SRM University. Downloaded on September 10,2020 at 04:37:12 UTC from IEEE Xplore. Restrictions apply.

Abbreviations and Acronyms

OCR - Optical Character Recognition

gTTS – Google Text To Speech

Open CV- Open Computer Vision

TF- Tensor Flow

The hardware includes Raspberry Pi (INR 2300), USB Web Camera (INR 1500 approx.) and a speaker (INR 800 and above). Hence, the average cost adds up to roughly INR 5000. The Software used is open source thus no additional cost for the software is inculcated into the model.

V. Implementation

A. *Reading Text:*

A.1 A text document that needs to be read is placed in the camera view. This document should be printed. It should be written in the English language. The font size should not be too small to detect.

A.2 Use a USB web camera to click a picture of the written text document. This camera acts as an input device to our project model. It is connected to the Raspberry Pi via a USB connection. The picture clicked is converted into a .jpg format and sent for further processing.

A.3 The Raspberry Pi is the brain of the entire system. It is a miniaturized version of a computer. All the processing is done inside the Pi. The pi should be installed with all the required applications, libraries, dependencies, etc.

A.3.1 Update the Pi: The updating is shown in Fig. 4

Fig. 4. Updating Raspberry Pi

A.3.2 Installing the Raspbian OS :

Step 1: Download the Raspbian from the official website. The downloading details are shown in Fig. 5

Fig. 5. Downloading Raspbian OS from Official Website

Step 2: Unzip the downloaded file.

Step 3: Write the disc image to your microSD card

Step 4: Put the SD Card back into the Pi and boot up. The downloaded card flash is shown in Fig. 6

Fig. 6. Flashing the downloaded OS onto the SD card

A.3.3 Install Python: To Install Python on your Raspberry Pi you will have to build it yourself. Use the command

Sudo apt-get install python3

A.3.4 Install OpenCV: Use the original documentation present on the Official OpenCV website to install and download the packages.

https://www.learnopencv.com/install-opencv-4-on-raspberry-pi/

A.3.5 Install Tesseract OCR:

Step 1: Installation using the command

sudo apt-get install tesseract-ocr

Step 2: Check the version installed

tesseract –v

A.3.6 Install the gTTS :

Step 1: Install the gTTS on your Raspberry Pi using the command:

pip3 install gTTs

Now your Raspberry Pi is ready to perform the Text Reading. All the necessary components have been installed successfully.

A.4 Connect an amplifier to the Raspberry Pi using Bluetooth or Analog connection. The amplifier could be a speaker or any other device.

A.5 The clicked picture is then processed by the OCR and gTTS to read out the document.

B. *Object Detection*

B.1 Use a webcam, connected to the raspberry pi, to click a picture or get a live feed of the surroundings.

B.2 The Web Camera's input is given to the raspberry pi for further processing.

B.2.1 Install TensorFlow: Install the TensorFlow on your Raspberry Pi using the command

pip3 install TensorFlow

Authorized licensed use limited to: SRM University. Downloaded on September 10,2020 at 04:37:12 UTC from IEEE Xplore. Restrictions apply.

B.2.2 The TensorFlow Object Detection uses Google's API of Protobuf. Install it using the command

sudo apt-get install protobuf-compiler

B.2.3 set up the Directory and Python path. Download the repository from GitHub using

gitclone–recurse-submodules https://github.com/tensorflow/models.git

B.2.4 Run the script using the command:

python3 Object detection picamera.py –usbcam

B.2.5 The gTTS is already installed. Provide the output of the detected objects to the input of gTTS to provide an output in the audio format.

B.3 Connect an amplifier to the Raspberry Pi using Bluetooth or Analog connection. The amplifier could be a speaker or any other devices.

VI. RESULT ANALYSIS

Successful results were obtained while executing the project. Text was read out aloud in the first case whereas in the second objects were detected successfully.

Fig. 7. Reader Part

Fig. 8. Object detection

During the analysis of the Reader part it was noticed that font below the size of 16 was difficult to detect shown in Fig. 7. Also, a plain background behind the text produced better and more accurate results than an abstract background. The hundred percentage accuracy cannot be expected while performing any of the two modules. During the Object Detection the rectangular box that appears around the detected object adjusts itself according to the size of the image shown in Fig. 8. This is done automatically. The model can detect almost everyday objects in the surrounding like the Television set, Bed, Chair, Fan, Person, Tree, Bird, Animals etc. It is possible to train the model with other new objects for a wider spectrum of detection. The frame rate achieved is 1 and above, which is very fast compared to the processing capabilities of the Raspberry Pi. The input that is being fed to the program can sometimes produce inaccurate results due to bad lighting conditions, however to overcome this an accuracy meter has been provided which determines how close the detected object is to the actual object.

VII. CONCLUSION

This paper shows the implementation of the project 'Reader and Object Detector for Blind'. It was developed to aid blind in everyday life and help them to be independent. The project aimed to cover a broader aspect of life and hence we incorporated both the parts into one. This project aims to assist the blind people in reading the printed text on pamphlets, books, magazines and other printed material. One can be assisted in reading their everyday newspaper with the help of this device. The feature of Object Detection can be used to help the blind people know more about their surroundings without having to move around the place. The project faces certain limitations as well. The project can only identify words of English language. It can read words that have font size greater than or equal to 14. Only certain objects can be identified. The frame rate for object detection is slow due to the small computation power of Raspberry Pi.

VIII. FUTURE WORK

The frame rate for the object detection is very low. With better computational power we will be able to increase the accuracy and speed of the object detection. After training the model with other languages we would be able to read several other languages and reach out to more people with a language diversity.

REFERENCES

[1] Roberto Neto, Nuno Fonseca Camera Reading For Blind People. Procedia Technology 16 (2014) 1200 – 1209

[2] D.Velmurugan , M.S.Sonam , S.Umamaheswari , S.Parthasarathy , K.R.Arun A Smart Reader for Visually Impaired People Using Raspberry PI DOI 10.4010/2016.699 ISSN 2321 3361 © 2016 IJESC.

[3] Jamal S. Zraqou, Wissam M. Alkhadour and Mohammad Z. Siam Real-Time Objects Recognition Approach for Assisting Blind People. International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 – 5161.

[4] Amal Jojie , Ashbin George , Dhanya Dhanalal Nayana J. Book Reader for Blind. ISSN (e): 2250-3021, ISSN (p): 2278-8719 PP 33-38.

[5] P. Rajeshwari, P. Abhishek, P. Srikanth, T. Vinod. International Journal of Trend in Scientific Research and Development (IJTSRD) Volume: 3 | Issue: 3 | Mar-Apr 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

[6] Ferdousi Rahman, Israt Jahan Ritun, NafisaFarhin, JiaUddin .https://dl.acm.org/doi/proceedings/10.1145/3309074

Authorized licensed use limited to: SRM University. Downloaded on September 10,2020 at 04:37:12 UTC from IEEE Xplore. Restrictions apply.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Object Detection with Voice Feedback

Rajat Lilhare[1], Jitendra Meena[2], Nikhil More[3], Shubhangi Joshi[4]

[1,2,3]Student Final Year ECE, MIT School Of Engineering, Pune, Maharashtra, India
[4]Assistant Professor, Dept. Of Electronics & Communication Engineering MIT SOE, Pune, Maharashtra, India

---

Abstract - *Object Detection is a field of Computer Vision that detects instances of semantic objects in images or videos (by creating bounding boxes around them in our case). In this project, we will convert image to text and then text to speech for the visually impaired person who deserve to live independently by using You Only Look Once V3 (YOLO v3) algorithm that runs through a variation of an extremely complex Convolutional Neural Network architecture called the Darknet with OpenCV and Google Text to Speech, We can then convert the annotated text into audio responses and give the location of the objects in the camera's view. The system will continuously capture multiple frames using a camera on raspberry pi and the frames then converted to audio segment, the obtained results manage to achieve the success of the proposed prototype in giving visually impaired users the capability to understand unfamiliar surroundings, through a user friendly device with this profound object identification model.*

Key Words: *Object detection, YOLO, Deep neural network, Tensorflow, OpenCV, Python, Raspberry Pi3b+, Google Text To Speech.*

1.INTRODUCTION

With the recent rapid development of information technology (IT), a lot of research has been carried out to solve inconveniences ineveryday life, and as a result, various conveniences for people have been provided. Nevertheless, there are still many inconveniences for the visually impaired. The greatest inconveniences that a blind person feels in everyday life include finding information about objects and indoor mobility problems. They have difficulty recognizing simple objects, and it is not easy to distinguish objects thathave similar forms. Previous studies included object analysis using ultrasonic sensors. However, with these methods, it is difficult to know exactly where an object is located, especially in the presence of obstacles. In this paper, we analyze accurate object information and obtain a location using a deep learning object recognition technique. In addition, voice recognition and voice guidance technologies are synthesized so the visually impaired can know the location of the objects they want to find by speaking to the system. Object recognition algorithms are designed based on theSingle Shot MultiBox Detector (SSD) structure, an object recognition deep learning model, to detect objects using a camera. In addition, voice recognition technology designed to use speech-to-text (STT) technology converts a user's vocal commands into text, from which only specific words are extracted and retrieved by the system. In the voice guidance technology, the technique of synthesizing the position of the article so it can be output, and synthesizing the name of the article, is done by using text-to-speech (TTS). In this paper, we propose an efficient object detection system to help find objects in a certain space without help from others, with special consideration for the blind.

1.1 Object Detection

Before Object recognition has developed rapidly, starting with the deep learning–based convolutional neural network (CNN) technique [5] that drew attention at the ImageNet 2012 competition. The CNN, however, was accurate with object classification, but it was difficult to determine where inside the image the object was located. Subsequently, the model for solving this problem was the region-based consolidated neural network (R-CNN), which uses a linear regression method. However, due to the slow speed of theR-CNN, Fast R-CNN was developed. It utilizes a deep learning technique to not only classify the object but also to find the area the object is located in. Nonetheless, there was a limit in that the above model's object recognition processing speed was insufficientfor real-time object recognition. Since then, You Only Look Once (YOLO), which comprises all the processes of object recognitionas a deep learning network, has emerged, and technologies with fast detection speeds, such as Single Shot MultiBox Detector (SSD),have been developed. YOLO estimates the type and location of objects using regression inference on the problem of area selection and classification. On the other hand, SSD does not create candidate areas separately, but recognizes objects using a feature map of various sizes. Since it does not generate candidate areas, it is faster to train than the Faster R-CNN and is more accurate than YOLO because it uses different sizes of feature map.

1.2 Image Processing

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Image processing basically includes the following three steps:

- Importing the image to the system.

- Analyzing and manipulating the image.
- Output in which result can be altered image or report that is based on image analysis.

1.3 Yolo v3

In this project we have used YOLO v3 which is faster than the prior version. It works three times faster, at 320 ×320 YOLOv3 runs 22ms at 28.2 map. It has a similar performance but 3.8× faster. The most notable characteristic of v3 is that it makes 3 distinct scales of detections. YOLO v3 is a fully convolutional neural network and it generates its resultant output by applying a 1 x 1 kernel to a feature map.In YOLO v3, the recognition is obtained by implementing 1 x1 detection kernels to three-size feature maps at three different regions in the network.

Within each boundary the network predicts 4 coordinates tx,ty,tw,th. Whereas if a cell is offset in the upper left corner of the image by (cx,cy) and prior bounding boxes has pw, ph width and height respectively then the prediction is done .

1.4 OpenCV

Techniques for Object Recognition in Images and Multi-Object Detection and segmentation is the most significant and testing central undertaking of Computer vision. It is a basic part in numerous applications, for example, image search, scene understanding, and so far. However it is as yet an open issue because of the assortment and multifaceted nature of item classes and foundations.

The most effortless approach to identify and fragment an item from a picture is the shading based techniques. The tem and the foundation ought to have a critical shading distinction so as to effectively portion objects utilizing shading based strategies.

OpenCV usually captures images and videos in 8-bit, unsigned integer, BGR format. In other words, captured images can be considered as 3 matrices; BLUE, GREEN RED (hence the name BGR) with integer values ranging from 0 to 255.In genuine pictures, these pixels are little to such an extent that the natural eye can't separate.

1.5 Hardware

We are using Raspberry Pi 3 Model B+ which has Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz and 1GB LPDDR2 SDRAM with class 10 micro SD card where the OS and project is stored, the reason behind we are using the class 10 memory card is that it helps to retrieve data at higher speed so that the project can take lesser time to execute The camera used in our system is Raspberry Pi Camera Module v2 which has Sony IMX219 8-megapixel sensor to feed images at 30 frame per second to this trained model,Ultrasonic Distance Sensor HC-SR04.

2. Working

We are using Python3 for this project, the camera is initialized by using OpenCV library and the camera starts capturing frames with the rate of 30 frames per second to the algorithm. Then the system uses YOLO v3 which is trained on the COCO dataset and Dark Neural Network (DNN) to identify the object kept before the user.

The object identified is later converted to an audio segment using gTTs which is a python library. The audio segment is the output of our system that gives the spatial location and name of the object to the person. Now by using this information the person can have a visualization of the objects around him. The proposed system will even protectthe person from colliding to the objects around will secure him from injuries & Ultrasonic sensor will detect object & give the distance between came.

2.1 Block Diagram & Proposed Platform

Figure 1: Block Diagram

Figure 2: Proposed Platform

3. Result and Experiment

The proposed system will be able to identify the object in front of the camera and will later on convert it into mp3 using gTTS. The proposed system is very low cost, FIG 3 shows the whole system which is a Raspberry pi 3b+, Bluetooth headphones and a power bank in order to provide power to the raspberry pi. The hardware implementation is shown in the given figure.

Figure 3.1: Interfacing of Hardware.

Figure 3.2: Object detection with Python.

Figure 3.3: Multiple Object detected.

Figure 3.4: Working of the system which is detecting 2 person & 1 cellphone.

Figure 3.5: Working of the system which is detecting person & cellphone.

FUTURE SCOPE

This project is for the blind people who are incapable to see this colorful and beautiful world, our initiative will support them to have a better life. By this project one will be able to understand what object is present in front of him and by continuous research and development our team will be able improve this product by feeding more data to the Deep Learning algorithm by which the accuracy of the model will increase as well as the power of the algorithm to recognize more objects will increase. The object recognition system can be applied in the area of surveillance system, face recognition, fault detection, character recognition etc. The objective of this thesis is to develop an object recognition system to recognize the 2D and 3D objects in the image. The performance of the object recognition system depends on the features used and the classifier employed for recognition. This research work attempts to propose a novel feature extraction method for extracting global features and obtaining local features from the region of interest.

CONCLUSIONS

This study can be used widely to provide the blind with privacy and convenience in everyday life. Also, it is expected to be applied to industrial areas where diminished visibility occurs, such as coal mines and sea beds, to greatly help production and industrial development in extreme environments. The study aims to enable people with visual impairment to live more independently. Peoplewith visual impairment will be able to overcome some threats that they may come across in their day to day life that may be either while reading a book or traveling through the city by making efficient use of the application and its associative voice feedback. Therefore, it will help to prevent possible accidents. The mobile devices can be carried easily and the camera of the device can be used to detect object from the surroundings and give output in audio format. Thus, helping visually impaired people to 'See throughthe Ears'.

REFERENCES

[1] Kedar Potdar, Chinmay Pai and Sukrut Akolkar, "A Convolutional Neural Network based Live Object Recognition System as Blind Aid", arXiv:1811.10399v1 [cs.CV] 26 Nov 2018 https://arxiv.org/pdf/1811.10399.pdf

[2] Liam Betsworth, Nitendra Rajput, Saurabh Srivastava,and Matt Jones. Audvert: Using spatial audio to gain a sense of place. InHuman-Computer Interaction–INTERACT 2013, pages 455–462. Springer, 2013

[3] Evanitsky, Eugene. "Portable blind aid device." U.S. Patent No. 8,606,316, 10 Dec. 2013.

[4] A.Culjak, D.Abram, T. Pribanic, H. Dzapo and M. Cifrek, A brief introduction to OpenCV,"2012 Proceedings of the 35th International Convention MIPRO, Opatija, 2012, pp. 1725-1730.

[5] Rahul Kumar and Sukadev Meher, "Assistive System for Visually Impaired using Object Recognition, M.Sc. Thesis at Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, Rourkela, Odisha-769 008, India, May 2015.

[6] Khushboo Khurana, Reetu Awasthi, 2013, "Techniques for Object Recognition in Images and Multi-Object Detection," International Journal of Advanced Research in Computer Engineering & Technology.

[7] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", University of Washington, Allen Institute for AI, Facebook AI Research, 2016.

[8] J. Redmon and A. Farha, Yolov3: An incremental improvement. arXiv, 2018.

[9] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pages 6517–6525. IEEE, 2017.

BIOGRAPHIES

Rajat Lilhare
B.Tech Student at MIT School of Enigeering, Pune, Maharashtra, India

Jitendra Meena
B.Tech Student at MIT School of Enigeering, Pune, Maharashtra, India

Nikhil More
B.Tech Student at MIT School of Enigeering, Pune, Maharashtra, India

Shubhangi Joshi
Assistant Professor, Dept.of Electronics & Communication Engineering at MIT School of Enigeering, Pune, Maharashtra, India

3D CMOS sensor based acoustic object detection and navigation system for blind people

Larisa Dunai, Beatriz Defez Garcia, Ismael Lengua, Guillermo Peris-Fajarnés
Universitàt Politècnica de Valencia, Camino de Vera s/n, 46022, Valencia, Spain
ladu@upv.es

***Abstract- T*he paper presents a new wearable Cognitive Aid System for Blind People (CASBliP). The prototype device was developed as an obstacle detector, orientation and navigation Electronic Travel Aid (ETA) for blind people. The device provides a binaural acoustic image representation. The environmental information acquisition system is based on an array of 1×64 CMOS Time-of-Flight sensors. Through stereoscopic (binaural) acoustic sounds the device relays the surrounding near and far environment. Experimental results demonstrate that blind users are able to detect obstacles and navigate through unknown and known environments safety and confidently. CASBliP works accurately in range of 15m in distance and 64º in azimuth, providing significant advantages in comparison with currently existing ETA systems.**

I. INTRODUCTION

There are over 314 million of blind and partially impaired people in the world. Over 45 million of them are totally blind [1]. Besides independence and communication limitation, the main problem for blind people is their mobility. Mobility has been defined by Foulke [2] as "the ability to travel safety, comfortably, gracefully and independently through the environment".

Since early decades, together with the sensors development, many efforts have been made to develop new and sophisticated Electronic Travel Aids (ETAs) able to perceive and represent the surrounding environment. The idea underlying the development of such devices was to overcome human sense limitations, such as blindness. These devices would help blind people to perceive their surrounding environment through tactile, vibrotactile, vocal or acoustic senses.

Traditionally, three classes of ETAs are known: *obstacle detectors*, *environmental sensors* and *orientation and navigation aids*. The first two classes make use of ultrasonic, laser and artificial vision to measure the distance to the nearest objects or receptors. The third class uses global positioning systems (GPS) for route prediction and orientation task.

The first remarkable *obstacle detector* ETA is the Lindsay Russell Pathsounder [3], [4]. Also, the Mowat Sonar Sensor is one of the simplest and most popular [5]. Both devices use ultrasonic transducers which measure the distance to the nearest object, providing tactile or vibrations feedback [6], [7]. This subgroup of ultrasonic based obstacle detector ETAs also includes other devices such as Nothingham Obstacle Detector, Sonic Torch, Sonic Pathfinder, Polaron™ Sensory 6, WalkMate [8], [6], [9], [10]. The existing obstacle detectors ETA devices are able to detect the obstacles in a range covering up to 15 feet and 15º in azimuth.

On the other hand, the most popular laser obstacle detector ETA is the Laser Cane [8], [11]. Laser Cane is based on laser technology which is implemented into the cane, making use of the triangulation principle. Its photo detectors receive the light beams which, after collision with the obstacle surface, are reflected. This enables the device to calculate the distance up to the obstacle. The detection range is up to 4m. Through acoustic signals, the blind user is informed about detected object.

The second class, *environmental sensors*, includes devices that try to go further than obstacle detectors devices [6], [11], providing additional features. Bat K Sonar Cane [12], Ultracane, KASPA, Trisensor, Miniguide, Sonic Torch, Sonicguide and Navbelt [13] belong to this class. The most influential ETA from environmental sensors is the Binaural Sonar Electronic Travel Aid also known as Sonicguide [14]. Sonicguide is the improved Sonic Torch device, developed by Kay in 1959 [8], [15]. The device incorporates two acoustic channels which help blind users not only to detect the objects but also to recognize and count them in range up to 4m in distance and 55º in azimuth [16], [14]. On the other hand, vOICe device is one of the latest ETA systems using stereovision technology for environmental image acquisition. The device, via acoustic signals and a tactile interface, informs the users about their surrounding environment [17], [18].

The third class of ETAs corresponds to the *orientation and navigation aids*; the main objective of these devices is the familiarization with the environments as well as the navigation [19], [20], [21]. The pioneer orientation device is Talking Signs [22]. The device, by means of an infrared beam pattern, controls the range and coverage of each sign in a range up to 30m in outdoor environments. SONA – Sonic Orientation Navigation Aid is based on a similar idea as Talking Signs [23]. The system was developed at Georgia Institute of Technology in 1979. Other devices such as Verbal Landmark [24], Easy Walker [11], Pilot Light, etc…belong to this class too. Also, the Global Positioning Systems (GPS) [10] are included in this group, [25]; among them, one of the best known is the GPS ETA, developed at the University of

Fig.1. Measurement principle of the Time of Flight (TOF) method

California [26], and known as Personal Guidance System [27]. This device is based on radio signals from satellites, which provide real information about every point of the Earth surface [28], [29]. Through synthetic speech [30] or tactile [31] the user gets informed about his location. The system range is 20m. Within this group we can also find the Mobility for Blind and Elderly People Interacting with Computers (MoBIC) device [32], Makino [27], Electronic Guide [27], GPS Braille Note [33], Geotact [34] Easy Walker [35], Loadstone, Trekker [36] or Tormes [37], Tanea prototype [38], [39], System for Wearable Audio Navigation (SWAN) [40], [41], Tyflos [42], [43], [44], [45], [46]. All these devices provide artificial speech, acoustic or tactile feedback to the user.

The aim of this paper is to propose a new system able not only to enable mere object detection, but also to provide orientation and navigation aid. This system is specially suited for blind and partially blind people. It provides significant advantages against other conventional ETA systems, such as wider detection range, possibility of use both in indoor and outdoor environments as well as a simple and easy utilization. The paper is structured as follows: Section II provides a general description of the developed device, listing its main advantages and drawbacks. Section III details the technical implementation of the sensor module whereas Section IV describes the acoustic module. In Section V, the experimental results are presented and analyzed. Finally, Section VI shows the conclusions of the work.

II. DESCRIPTION OF THE ACOUSTIC SYSTEM

Advances in sensor technology and electronics allow providing environmental information to visually impaired people. In the present section, the developed object detector and navigation Electronic Travel Aid is described.

In general terms, the CASBliP system detects the objects in the frontal view of the user and converts the information regarding object location into binaural virtual sounds which are delivered to the user via stereophonic headphones. The object detection is achieved by emitting laser light beams which are reflected after colliding with the surrounding objects (see Fig.1). These light beams are received by the sensor system of the device and converted into sounds by means of its internal software.

As it will be later detailed, the object detection system of the device is based on an improved 1×64 3D CMOS image sensor, originally developed for pedestrian recognition by cars [47], [48]. This sensor is implemented into a pair of glasses which are also equipped with the laser electronics for light beam emission. Besides, the Field Programmable Gate Array (FPGA), in charge of information processing and conversion into sounds, is implemented in an additional module of the device. This FPGA contains functions regarding sensor control, data acquisition, distance calculation, application software, sounds files, sound reproducing system, power supply and connectors between the parts of the device. As a conclusion then, the device is based on two main parts; *sensor module*, which includes the glasses and part of FPGA (for laser beam emission and reception) and the *acoustic module* based on the part of the FPGA- in charge of information processing and conversion into sounds- and on the stereo headphones (to deliver the sounds to the user).

The aim of the Acoustic System is not only to inform the user about the object presence in the range of the device, but also to inform him about object direction and distance. The device operates in an on-line mode. This means that it

delivers environmental information in real-time. Moreover, the device works properly both in known and unknown environments, and both in simple and in complex situations. Acoustic sounds perceived via headphones by the human brain create a map of the detected object, leaving the cognition task to the user. By hearing virtual sounds, the user is in charge to decide whether to avoid the obstacles in his direction or not. In this context, the functions of the Acoustic System are the detection and information tasks, whereas the navigation and route planning tasks are carried out by the user.

Although there are many Electronic Travel Aids that implement acoustic sounds, such some of those commented in Section I, there are significant differences between the developed acoustic system and other acoustic ETAs.

- The acoustic system enables the user to decide route and path planning. The device detects all objects appearing in the area of view. At the same time it informs the user about their presence and the user is free to decide to decide his way.
- The device provides a full environmental image due to the 1×64 pixels 3D CMOS image sensor which covers 60º in azimuth and a distance range between 0,5m and 5m. The device simultaneously provides 64 sounds covering the whole detected environment.
- The device perceives the object depth, since the different distances are represented by specific sounds.
- Virtual acoustic sounds are binaural sounds; the human brain easily perceives them as sounds provided by the environmental objects.
- The device provides environmental information in real-time. Sensor and Acoustic modules operate so fast that the device does not require supplementary time to interpret the environmental information.
- No false noises; when there is no object in the area of view of the system, the device is silent, and so no sounds are reproduced by the system.
- The system works on darkness and sunny days, cold and hot weather, because the 3D CMOS sensor works with infrared laser impulses.
- Acoustic sounds are sounds which do not represent any impact material, also do not disturb the user when the device is used during a long time.
- The environmental acquisition module of the device is placed at the human eyes level, which allows the user to perceive the surrounding from the head level. Since it is a head mounted device, the user is able to move the head and scan the environment at the direction of view of the user.
- Because of the binaural cues of the acoustic sounds, the sounds do not overlap when there is more than one object in the area of view.
- With the device the user is able to perceive the object shape, i.e., if it has a plane, convex or concave surface.
- The used sound does not interfere with external noises such as traffic lights, human noises, etc.

Fig.2. Multiple Double Short-Time Integrated measurement principle

III. IMPLEMENTATION OF THE SENSOR MODULE

The environmental image acquisition device or *sensor module* consists of a 4×64 3D CMOS image sensor based on 1×64 photodetector pulses.

The design of the sensor module is based on three parts: interface, sensor electronics and laser electronics. Fig 2 shows the sensor system design, where the laser pulse emitters are placed at the left and right side of the 3D CMOS sensor close together on a pair of glasses. The processing unit of the sensor module is implemented into FPGA, which is in charge of sensor control, data acquisition, and distance calculation and application software.

Sensor module transmits and receives identical 1×64 pixels laser pulses with refreshing rate of 25fps in a horizontal line at the user eye level, where the distance range is strictly dependent of the sensor module. Sensor module range for object detection is between 0,5m and 5m in distance and up to 60º in azimuth. 3D CMOS sensor generates pulses in the range of 20 to 200ns, the rise and fall timing range is 10ns, maximum shutter speed is 30ns, pulse repetition frequency is up to 20 kHz and the laser pulse peak power is up to 2kW. The wavelength is 905 nm in the near-infrared laser pulse (NIR) range; therefore the laser light becomes invisible to the human eye. Maximum frame rate of the image sensor is 19500 fps for a synchronic clock of 66 MHz, each pixel clock is of 5MHz, and output swing is 1,9V. The generated pulses are strongly rectangular pulses. The pixel pitch is 130µm in horizontal direction resulting in a sensitive area of 64×130µm =8,32mm^2. The chip is completely controlled by the FPGA.

Laser pulses generated by the NIR laser diodes and defocused by diffractive optics are emitted to the environment illuminating entire field of view [49]. When an object appears in the laser range, the pulses are reflected to the camera lenses of the 3D CMOS image sensor calculating the distance between the object surface and the sensor system for every 1×64 pixels simultaneously (see Fig.1).

The time delay τ_{TOF} of the laser light between the sensor and the object for a distance *d* can be calculated as.

$$\tau_{TOF} = \frac{2*d}{c} \qquad (1)$$

The Time-of-Flight (TOF) measuring method is based on the fact that the emitted light pulse is not measured directly, but through accumulation of charges on a sensor chip. This procedure is equivalent to the Continuous wave distance measuring method when using a short single pulse [47]: Camera lenses transmit a part of the reflected laser pulse to the 3D CMOS sensor chip surface, which operates through its light sensitive pixels as if they were electronic short-time windowed integrators. Then the distance between object surface and 3D CMOS sensor is calculated as:

$$d_x = c\frac{\tau_{TOF}}{2} = \frac{v_c}{2}(T_1 - T_w\frac{U_1}{U_2}) \qquad (2)$$

where T_1 represents the short integration time, T_w is the pulse width of the linear laser, U_1 is the measured sensor signal at integration time T_1 and U_2 is the sensor signal measured for integration timeT_2, d is measured distance, v_c is light propagation speed. The propagation time can be calculated by computing the quotient of the two integrated shutter intensities U_1/U_2.

Fig.1 shows the measurement principle of indirect TOF method [47] used in the development of the Acoustic System.

The travel time between the emission and reception of the laser pulse at pixel x depends on the travel distance, and can be calculated as:

$$T_{Travel,x} = \frac{2v_c}{d_x} \qquad (3)$$

Due to the near-infrared laser pulses, which have the ability to illuminate the entire scene simultaneously avoiding any sequential scanning, Time-of-Flight system process images in real time. In order to reduce the required laser power and to increase the measurement accuracy, multiple double short-time integration algorithms (MDSI) were used.

IV. IMPLEMENTATION OF THE ACOUSTIC MODULE

Whereas the sensor module provides to the acoustic module the linear image of the surrounding environment, the *acoustic module*, is in charge of transmitting this information to the blind user, by using virtual acoustic sounds. The principle of the acoustic module is to assign an acoustic sound to each one of the 64 pixels, for different distances. The acoustic sounds will be reproduced through the headphones from the position of the detected object, always that the sensor will send to the acoustic module distance values. The sound module contains a bank of acoustic sounds previously generated for a spatial area between 0,5m and 5m, for 64 image pixels, where one pixel has the coordinate of 0,94º in azimuth occupying the 30º at the left ear and 30º at the right ear of the human head. A delta sound of 2040 samples at a frequency of 44,1kHz was used to generate acoustic information of the environment. In order to define the distances, 16 planes in distance were generated from 0,5m increasing exponentially up to 5m. Refreshing rate of the sounds is 2fps. 16 Mb memories are needed to process acoustic module. Distance displacement is strongly dependent on the sound intensity and on the pitch. At nearer distances, the sound is stronger than at farther distances. The more the distance increases, the more the sound intensity decreases. Virtual sounds were obtained by convolving an acoustic sound with non-individual Head-Related Transfer Functions (HRTF) previously measured using KEMAR manikin described below. Fig. 3 shows measurement model of Head-Related Transfer Function.

The working principle of the acoustic module is similar to 'read and play'. This is, the acoustic module reads the output data, which represent coordinates in distances and in azimuth, from the sensor module and plays the sound with the same coordinates. The time interval between sounds is 8ms when there are sounds playing. When there are no sounds, after 5ms the sound module recalls the sensor module.

To measure the HRTFs a Maximum Length Binary Sequence (MLBS) was used to generate the sound source. HRTF measurement is based on the calculation of the impulse response for both ears by using the filter in frequency domain.

For a sound signal $x_1(n)$ reproduced by the speakerphone, the registered response can be calculated as:

$$Y_1 = X_1 LFM \qquad (4)$$

where X_1 is the representation of the sound $x_1(n)$ in the transformed frequency domain, L is the transfer function of the speakerphone and all reproduction equipment, F is the transfer function of the space between the speakerphone and human ear and M is the transfer function of the microphone from the human ear and all recording equipment.

In order to calculate the filter, it was necessary to generate a $x_2(n)$ - *sound* reproduced by headphones- so that the registered response Y_2 would be equal to Y_1. Y_2 can be calculated as:

$$Y_2 = X_2 HM \qquad (5)$$

where H is transfer function of the headphone and all reproduction equipment.

Then, the filter T is obtained from (4) and (5).

$$T = LF / H \qquad (6)$$

In this way, the filter T is measured for the impulse responses for each ear.

The impulse response is obtained by circular cross-correlation between the MLBS from the system input and the response at the output.

$$h(n) = \Omega_{sy}(n) = s(n)\Phi y(n) = \frac{1}{L+1}\sum_{k=0}^{L-1} s(k)\cdot y(n+k) \qquad (7)$$

where Φ represents the circular cross correlation.

Fig.3. Head-Related Transfer Function measurement model

Direct implementation of (7), for large sequences becomes impossible due to its computational time. Then, the convolution method was used:

$$a(n)\Phi b(n) = \frac{1}{L+1} a(-n) * b(n) \tag{8}$$

And even then the computational time is great due to that the Fast Fourier Transform used is of 2^k-1 length. In that case, the Fast Hadamard Transform (FHT) was used to reduce twice the computational time for HRTF generation:

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2 \{ H_{L+1} [S_1 (P_1 y(n))] \} \rangle \tag{9}$$

where P are the permutation matrices, S are the redimension matrices and H_{L+1} are the Hadamard matrices of L+1 degree.

In conclusion, binaural cues had an influence on the generation of the virtual sounds. The azimuth displacement of the virtual sound source was obtained by the interaural phase shift of the left and right signals in ter

ms of time difference between the left and right ear. Each virtual sound corresponds to one pixel from the sensor module and covers an area of 60º.

V. EXPERIMENTAL RESULTS

In this section, the tests with the CASBliP System are described. Twenty blind users took part in the experiments. These experiments were developed both in laboratory and in outdoor environments not known 'a priori', under the supervision of blind people trainers and engineers. In order to avoid unnecessary displacements, two groups, group A (users

Fig.4. Laboratory experimental scenario

from Germany) and group B (users from Italy) simultaneously tested the device at their respective countries. For the experimental task, two identical devices were used, one for each group. Before proceeding to the experimental task of object detection and navigation, subjects were trained during a month. The aim of this training was to learn to externalize the sounds listened through the headphones as well as to comprehend how to relate one spatial coordinate from the near environment to one specific virtual acoustic sound listened through the headphones, noticing the difference between two different located sounds. Seven exercises with different levels of complexity were developed. These exercises included from single object detection to the detection of several objects, doors and walls. During the tests, users were placed at distance of 2m from the object and they were asked how they perceived the sounds. When the users were prepared, they were allowed to move towards the object and back, going around the object. The idea was to learn how to localize spatial sounds. It is important to remark that, for blind people, each spatial sound represents an obstacle which must be avoided. Training period depended of the user ability to localize the sounds. After the training period, blind users proceeded to the laboratory and outdoor experimental trials.

The experiments aimed to test the Acoustic System as obstacle detector and navigation mobility aid for blind people. The tests consisted of two tasks: navigation through a

path with 4 pairs of soft obstacles in a controlled environment and navigation in an open uncontrolled environment. In the first experiment, subjects walked through a 14m artificial laboratory path limited by soft carton boxes (see Fig.4). All obstacles were identical in weight and height. After the device had been connected and prepared, the users were faced to the scenario. The columns were placed in an asymmetric order, separated horizontally by 2m and at a distance of 2,5m between the last pair of columns and the wall. In the laboratory experiment all obstacles remained static. The objective of the experiment was to navigate through the path avoiding all the obstacles. Two routes were prepared for the experiment. The second route represents the opposite direction to the first, i.e., the way back direction of the first route.

Fig.5. Average Absolute Walking Time (AWT) for laboratory experiment

In the experiment, the absolute walking time (AWT), number of hits (NUH) and number of corrections (NOC) were measured. Blind subjects were able to perceive object thickness and texture. Also the device can reliably detect objects with small diameters due to its 64 pixels which scan only 60º in azimuth.

The data are presented for a clear inspection in Fig. 5. The runs 1 and 2 are the runs for the group A and the runs 3 and 4 are the runs for the group B. Note that after the training exercise, the participants performed the subsequent experiments without any white cane, trainer instructions or support. The value of time in minutes in which the subject reported the end of the trajectory is plotted in the y-axis. Fig.5 shows that the group A required a longer time to complete the experiment. This is due to the fact that some subjects which had difficulties on sound externalization and orientation took part in the experiment. The difficulties were due to the time of blindness; for example, in the group A, one of the subjects is blind from born and he never used a navigation device for blind people. This user find difficulties in learning that the sounds were representing something that he had never seen, or that the sounds listened via headphones were not in his head, but coming from the environment.

The goal of the second experiment, named Mobility Test I, was to test perceptual authenticity of the navigation, by evaluating if blind users were able to detect the objects and to navigate through a 29m scenario in an outdoor environment.

In the evaluation of object detection and spatial sound localization in an open environment, there is always the problem that the sounds are not be perceived exactly as in the laboratory since, in the laboratory, the user is protected from the external noises, such as traffic noises, human, animal and bird's speech, etc. Also, blind users require longer times to take decisions about the perceived objects and the environment.

TABLE I
MOBILITY TEST RESULTS FOR TWENTY SUBJECTS REPRESENTING THE ABSOLUTE WALKING TIME (AWT), THE NUMBER OF HITS (NH) AND THE NUMBER OF CORRECTIONS (NOC)

	Mean time, min	
	Group A	**Group B**
AWT-0	1,00	-
AWT-1	**9,02**	**8,27**
AWT-11	**7,31**	**6,46**
AWT-2		**6,49**
AWT-21		**5,55**
NUH-1	2,80	0,70
NOC-1	1,40	0,20
NUH-11	2,00	0,60
NOC-11	1,20	1,00

Blind user tasks were to localize the objects and to judge the differences between the perceived objects, to analyze the environment, to select the free path and walk through the scenario. External noises had a slight influence in the object detection performance.

Four different scenarios were used for this test. The group A, from Germany, used two external scenarios. Both scenarios were located in the patio of the School for Blind People of Berlin (ABSV). The blind user should navigate through 29m long way where a variety of obstacles were placed, from the school entrance up to the school door. The test started with the blind user looking to a wide space where no object intersected the user direction of view. Because no objects appeared in the system direction of view, no signals were sent by the system to the user. Thus, the user knew that in an area of 5 meters, there were no obstacles.

Also, the user knew that when he was listening a sound, then there were objects are in his area of view. Despite the total silence of the system the user knew that this silence did not mean that the system was broken, so he must not worry. The group B used also two different scenarios. The tests were carried out in the IFC Institute courtyard in Italy.

Group A carried out two sessions of this test (in two different days) whereas the group B carried out four sessions. The location of the soft obstacles was modified except on the real obstacles such as building wall, columns and bench.

This time, the walking time, number of hits and number of corrections were compared between the two groups (Fig. 6). In average, the group A spent 90,19 minutes for the first run

and 75,09 minutes for the second run. The group B registered better results with respect to the total and mean time. For the first run the group B spent 84,32minutes, for the run 2 – 67,37minutes, for the 3rd run 68,09minutes and for the 4th 59,07minutes. Table I confirms that a slight improvement was detected for the repeated run. Also, we can observe that all subjects showed a very similar pattern in the first run, except on four subjects, which had great difficulties in orientation. Three of them, apart from requiring more time to navigate through the trajectory, were less accurate in the navigation. These subjects had difficulties in object detection making more hits and confusions and needing more corrections. According to the results from Fig.5 and Fig.6, navigation time is affected mainly by the lack of training.

Fig.6 Mean absolute walking times from Mobility Test I and mean number of hits. The ten clusters of bars show mean results for the ten subjects from the group A as well as the mean results from the four runs of the Group B. The cluster of bars from the second graph shows the average Number of Hits for both groups for the previous carried out runs

VI. Conclusions

In the paper, a new wearable cognitive real-time Acoustical System device is described, which was developed as obstacle detector, orientation and navigation Electronic Travel Aid for blind people. An array of 1×64 CMOS Time-of-Flight sensors scan the environment and through binaural acoustic sounds the device relays the surrounding near and far environment. The prototype device aims to be a portable device and constructed by commercially available components.

Experimental results in laboratory conditions and real uncontrolled outdoor environments demonstrate that blind users are able to detect obstacles and navigate through unknown and known environments safety and confidently. Acoustic System works accurately in range between 0 and 5 m in distance and 60º in azimuth. The resolution of the device is sufficiently high both in object detection and acoustic representation of the environment.

Finally, it must be also said that future improvements are possible in order to obtain more complete environment information:

- Improvement of the image acquisition techniques by adding more arrays of 1×64 CMOS image sensor, or using stereovision instead CMOS image sensor.
- Improvement of the acoustic module adding sounds in elevation.

Acknowledgment

The authors wish to make the following acknowledgments to the participants of CASBliP project and to all blind people for their active participation on the system development, tests and evaluation. The project was developed within the 6th European Frame Programme, Cognitive Systems and Robotics

References

[1] World Blind Union, “White cane safety day,” World Blind Union Press release, October, Canada, 2009.
[2] E. Foulke, “The perceptual basis for mobility,” AFB Res. Bull, vol. 23, pp. 1, 1971
[3] L. Russell, “Travel Path Sounder Proceedings,” Rotterdam Mobility Research Conference. N.Y: American Foundation for the Blind, 1965
[4] R.W. Mann, “Mobility aids for the blind—An argument for a computer-based, man-device-environment, interactive, simulation system,” Pro Conf on Evaluation of Mobility Aids for the Blind, Washington, DC: Com. on Interplay of Eng. with Biology and Medicine, National Academy of Engineering, pp. 101–16, 1970
[5] D. L. Morrissette, G.L. Goddrich, J.J. Henesey, “A follow-up-study of the Mowat sensors applications, frequency of use and maintenance reliability,” J. Vis. Impairment Blindness, Vol. 75, 6, pp. 244-247, 1981
[6] H. A. M. Freiberger, “Mobility Aids for the Blind,” Bulletin of Prosthetic Research, pp. 73-78, 1974
[7] Book, “Electronic travel aids: new directions for research,” National Research Council, Chap. 6, pp. 67-90, 1986
[8] L. W. Farmer, “Mobility devices,” Bulletin of Prosthetic Research, pp-47-118, 1978
[9] N. Debnath, J. B. Tangiah, S. Pararasaingam, A. S. A. Kader, “A mobility aid for the blind with discrete distance indicator and hanging object detection,” IEEE TENCON Vol.D, pp. 664-667, 2004
[10] L. Farmer, D. Smith, “Adaptive technology,” Book chapter, Foundations on Orientation and Mobility, Eds. 2, pp.231-260, 1997
[11] B. Ando, “Electronic Sensory systems for the Visually Impaired,” IEEE Instrumentation & Management Magazine, pp. 62-67, 2003
[12] B. S. Hoyle, “The Bathcane – mobility aid for the vision impaired and the blind,” IEE Symposium on Assistive Technology, pp. 18-22, 2003
[13] S. Shoval, J. Borenstein, Y. Koren, “The Navbelt-A computerized Travel Aid for the blind based on mobile robotics technology,“ IEEE Transactions on Biomedical Engineering, vol. 45, no. 11, November 1998, pp.1376-1386
[14] R. Kuc, “Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion and surface texture classification,” IEEE

Transactions on Biomedical Engineering, Vol. 49, No 10, pp. 1173-1180, 2002

[15] L. Kay, "An ultrasonic sensing probe as an aid to the blind," Ultrasonics, Vol. 2, No. 2, pp. 53-59, 1964

[16] L. Farmer, D. Smith, "Adaptive technology," Book chapter, Foundations on Orientation and Mobility, Eds. 2, pp.231-260, 1997

[17] P. B. L. Meijer, "An experimental system for auditory image representations," IEEE Transaction Biomedical Engineering, BME-39, (2), pp. 112-121, 1992

[18] P. B. L. Meijer, "A modular sysnthetic vision and navigation system for the totally blind," World Congress Proposals, 2005

[19] J. Brabyn, "New Developments in Mobility and Orientation Aids for the Blind," IEEE Transactions on Medical Engineering, Vol. BME-29, NO 4, pp. 285-289, 1982

[20] J. Brabyn, "A review of mobility aids and means of assessment," In D. H. Warren & E. R. Strelow (Eds) Electronic Spatial Sensing for the Blind, Martinus Nijhoff, Boston, pp.13-27, 1985

[21] J. Brabyn, "Technology as a support system for orientation and mobility," American Rehabilitation, 1997

[22] J. Brabyn, "New Developments in Mobility and Orientation Aids for the Blind," IEEE Transactions on Medical Engineering, Vol. BME-29, NO 4, pp. 285-289, 1982

[23] R. Kuc, "Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion and surface texture classification," IEEE Transactions on Biomedical Engineering, Vol. 49, No 10, pp. 1173-1180, 2002

[24] B. Bentzen, P. Mitchell, "Audible signage as a wayfinding aid: comparison of verbal landmark and talking signs. Draft Report to the American Council of the Blind, Sept. 1993

[25] C. C. Collins, "On mobility aids for the blind," *in Electronic spatial sensing for the blind*, D. H. Warren & E. R. Strelow, Eds., Dordrecht, Martinus Nijjoff, pp. 35-64, 1985

[26] J. Loomis, R. Golledge, "Personal guidance system using GPS, GIS, and VR technologies," Proceedings, CSUN Conference on Virtual Reality and Person with Disabilities, San Francisco, June17-18, 2003

[27] J. M Loomis, R. G. Golledge, & R. L. Klatzky, "GPS-based navigation systems for the visually impaired," In. W. Barfield & T. Caudell, Eds, *Fundamentals of wearable computers and augmented reality*, pp. 429-446, Mahwah, NJ: Lawrence Erlbaum Associates, 2001

[28] J. M. Loomis, R. G. Golledge, R. L. Klatzky, & J. R. Marston Assisting wayfinding in visually impaired travelers. In G. Allen (Ed.), Applied spatial cognition: From research to cognitive technology. Mahwah, N. J.: Lawrence Erlbaum Associates, 2006

[29] J. M. Loomis, R. G. Golledge, & R. L. Klatzky, "Navigation system for the blind: Auditory display modes and guidance," *Presence: Teleoperators and Virtual Environments*, 7, pp.193-203, 1998

[30] J. M. Loomis, J. R. Marston, R. G. Golledge, & R. L. Klatzky, "Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance," Journal of Visual Impairment & Blindness, 99, pp. 219-232, 2005

[31] R. L. Klatzky, J. R. Marston, N. A. Guidice, R. G. Golledge, J. M. Loomis, "Cognitive load of navigating without vision when guided by virtual sound versus spatial language," Journal of Experimental Psychology: Applied, Vol. 12, No. 4, pp. 223-232, 2006

[32] H. Petrie, V. Johnson, T. Strothotte, A. Raab, R. Michel, L. Reichert, A. Schalt, „MoBIC: an aid to increase the independent mobility of blind travellers," British Journal of Visual Impairment, Vol. 15, No. 2, pp. 63-66, 1997

[33] HumanWare "Braile Note GPS", www.humanware.com

[34] R. Fancy, R. Leroux, A. Jucha, "Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view," CVHI Conference, 2006

[35] Action for Blind People. Easy Walk – a new system for people with sight loss. http://www.actionforblindpeople.org.uk

[36] Y. Lagace, "Trekker GPS system creates handheld personal guide with HP ¡PAQ Pocket PC," HumanWare, HP notice, 2005 April

[37] T. M. Morales, A. M. Berrocal, "El GPS como sistema de ayuda a la orientacion de personas ciegas," IIICV-INTEREVISUAL Autonomia Personal, 2005, Octomber

[38] A. Hub, S. Kombrink, K. Bosse and T. Ertl, "TANIA – a tactile-acoustical navigation and information assistant," CSUN Conference. *Conference Proceedings of the California State University, Northridge Center on Disabilities' 22nd Annual International Technology and Persons with Disabilities Conference*, Los Angeles, CA, USA, 2007, March 19-24

[39] A. Hub, D. and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," Proceedings of the ACM SIGACCESS conference on Computers and accessibility, Designing for accessibility, Atlanta, GA, USA, pp.147-152, 2004

[40] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, F. Dellaert, "SWAN: system for wearable audio navigation,"11th IEEE International Symposium on Wearable Computers (ISWC 07), pp. 91-98, 2007, 11-13 Oct.

[41] B. N. Walker, J Lindsay, "Using virtual environments to prototype auditory navigation display," Asst. Technology, 17, pp. 72-81, 2005

[42] N. Bourbakis, D. Kavraki, "Intelligent assistants for handicapped people's independence: case study," IEEE International Joint Symposia on Intelligence and Systems, Rockville, MD, p. 337-344, 1996, Nov.4-5

[43] D. Dakopoulos, N. Bourbakis, "A 2D vibration array as an assistive device for visually impaired," IEEE Int. Conf. on BIBE07, Boston, MA, vol. II, pp. 930-937, 2007, Oct. 15-17

[44] N. Bourbakis (2008). Sensing surrounding 3-D space for navigation of the blind - A prototype system featuring vibration arrays and data fusion provides a near real-time feedback. IEEE Engineering in Medicine and Biology Magazine, v. 27, Issue 1, pp. 49-55.

[45] N. Bourbakis, P. Kakumanu, "Skin-based face detection- extraction and recognition of facial expressions," Applied Pattern Recognition, book chapter, Vol.91, 2008

[46] D. Dakopoulos, N. Bourbakis, "Preserving visual information in low resolution images during navigation of blind," 1st International Conference on Pervasive Technologies Related to Assistive Environments, Athens, Greece, 2008, July 15-19

[47] P. Mengel, G. Doemens, L. Listl, "Fast range imaging by CMOS sensor array through multiple double short time integration (MDSI)," Proc. IEEE International Conference in Image Processing (ICIP 2001), Thessaloniki (Greece), Oct.7-10, 2001, pp.169-172

[48] P. Mengel, L. Listl, B. Konig, C. Toepfer, M. Pellikofer, W. Brockherde, B. Hosticka, O. Elkhalili, O. Schrey, W. Ulfig, "Three-Dimensional CMOS image sensor for pedestrian protection and collision mitigation," J. Valldorf and W. Gessner, Eds., Advanced microsystems for automotive applications 2006. Berlin: Springer, 2006. (VDI-Buch), pp. 23-39

[49] O. Elkhalili, O.M. Scherey, P. Mengel, M. Petermann, W. Brockherde, „A 4×64 pixel CMOS image sensor for 3-D measurement applications," IEEE Journal of Solid Circuits, vol. 39, 7, pp. 1208-1212, 2004

Object detection and conversion of text to speech for visually impaired

Ankur Jyoti Sarmah[1], Kabindra Bhagawati[2], Kaustov Duwarah[3], Swetashree Dey Purkayastha[4], Antarjeeta Boro[5], Divika Muchahary[6]

[1,2]Assistant Professor, Department of Electronics and Telecommunication Engineering,
Assam Engineering College,
Jalukbari, Guwahati
ankur.et@aec.ac.in

[3,4,5,6]B.Tech 8th Semester, Department of Electronics and Telecommunication Engineering,
Assam Engineering College,
Jalukbari, Guwahati
work.kaustov@gmail.com

Abstract*: *Assistive technologies are being developed for visually impaired people in order to live confidently[1]. In this project work, we aim to develop a system which would help blind persons get information about objects present in their surroundings in their daily lives. The project work is framed into two stages. First, image is captured using a portable camera module, if the object is identified as cell phone, person, book or as such, then the detected object is matched with a predefined dataset. A predefined dataset is loaded in order to match the detected object with the captured image. Secondly, once it is matched the recognized text is synthesized for producing speech output. Text to speech conversion successfully converts the detected object into an audio signal using the gTTs Module with the help of the iPython audio library. The objective of the TTS is the conversion of text into a natural language. It is not only applicable for the visually impaired but also to any normal human beings who are willing to read the text as a speech as quickly as possible. The entire system is de-signed using the YOLOv4 model trained on the MS COCO dataset through a laptop. The entire system is processed by using Python as the programming language.

***Keywords*:Visually Challenged Person, Object detection, Optical Character Recognition (OCR), text detection, Text-to- Speech (TTS), Python, dataset.**

(Article history: Received: 8th March 2023 and accepted 11th August 2023)

I. INTRODUCTION

In our planet of 8 billion humans, at least 2.2 billion people have a near or distance vision impairment. In at least 1 billion or almost half of these cases, vision impairment could have been prevented or has yet to be addressed. This 1 billion people include those with moderate or severe distance vision impairment or blindness due to unaddressed refractive error (88.4 million), cataract (94 million), age-related muscular degeneration (8 million), glaucoma (7.7 million), diabetic retinopathy (3.9 million), as well as near vision impairment caused by un-addressed presbyopia (826 million). It is a fact that all over the world that the visually impaired (partially or completely blind) people face a lot of difficulties in identifying products and avoiding obstacles. The greatest inconveniences that a blind person feels in everyday life include finding information about objects and indoor mobility problems. They have difficulty recognizing simple objects, and it is not easy to distinguish objects that have similar forms[2]. Object Detection is a field of Computer Vision that detects instances of semantic objects in images or videos (by creating bounding boxes around them in our case. In this paper, we analyse object detection using a deep learning object recognition technique. Additionally, a voice assistance technology is introduced so the visually impaired people can know about the object or obstacle around them..

II. OBJECTIVE

The main objectives of the project are as follows:

1. To help specially abled people who aren't able to see through their eyes to read and feel the surroundings around them via audio signals.

2. To devise a system that converts the text to audio signals that can be heard via a speaker.

III. METHODOLOGY

The functional block diagram of the proposed system describes object/text to speech conversion which is shown below.

Figure 1Functional block diagram of Object/Text to Speech conversion

There are six major functional blocks in the block diagram which has been explained in detailed below:

A. *Object & Text Capture using Camera:*

The objects and texts to be detected are being captured using camera and converted to an OpenCV image.

B. *Adaptive thresholding:*

Adaptive thresholding typically takes a grayscale or color image as input and, in the simplest implementation, out-puts a binary image representing the segmentation. For each pixel in the image, a threshold has to be calculated. If the pixel value is below the threshold it is set to the background value, otherwise it assumes the foreground value.

There are two main approaches to finding the threshold: (i) the Chow and Kaneko approach and (ii) local thresholding. The assumption behind both methods is that smaller image regions are more likely to have approximately uniform illumination, thus being more suitable for thresholding. Chow and Kaneko divide an image into an array of overlapping sub images and then find the optimum threshold for each sub image by investigating its histogram. The threshold for each single pixel is found by interpolating the results of the sub images. The drawback of this method is that it is computational expensive and, therefore, is not appropriate for real-time applications.

An alternative approach to finding the local threshold is to statistically examine the intensity values of the local neighbourhood of each pixel. The statistic which is most appropriate depends largely on the input image. Simple and fast functions include the mean of the local intensity distribution,

$$T = mean$$

The median value

$$T = median$$

Or the mean of the minimum and maximum values

$$T = \frac{\max + min}{2}$$

The size of the neighbourhood has to be large enough to cover sufficient foreground and background pixels, otherwise a poor threshold is chosen. On the other hand, choosing regions which are too large can violate the assumption of approximately uniform illumination. This method is less computationally intensive than the Chow and Kaneko approach and produces good results for some applications [3].

C. *Text/ Object classifier*

Text classification is a common NLP task that assigns a label or class to text. We use various Machine Learning algorithms to classify the object/ text for the ease of processing the data and training the data to be able to extract information and convert the detected data into audio signals

D. *Output through Speaker SPI*

A serial peripheral interface (SPI) is an interface that enables the serial (one bit at a time) exchange of data between two devices, one called a master and the other called a slave. An SPI operates in full duplex mode. This means that data can be transferred in both directions at the same time. We use the SPI to output the audio signals processed by the iPython library.

IV. System Implementation

A. *Google Colaboratory*

Colaboratory, or "Colab" for short, is a product from Google Research. With colab we have executed the python code through the browser. It is especially well suited to machine learning, data analysis and education that allows us to combine Python code and rich text along with charts, images, HTML, LaTeX and more into a single document stored in Google Drive. It connects to powerful Google Cloud Platform runtimes and enables you to easily share your work and collaborate with others. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

B. *YOLO*

YOLO is real-time object detection. It applies one neural network to the complete image dividing the image into regions and predicts bounding boxes and possibilities for every region.

Predicted probabilities are the basis on which these bounding boxes are weighted. A single neural network predicts bounding boxes and class possibilities directly from full pictures in one evaluation. Since the full detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.

C. *Google Text to Speech (gTTS)*

TTS (Text-to-speech) is the process of converting words into an audio form. The tool takes an input text from the user, and performs logical inference on the text. This processed text is passed into the next block where digital signal processing is performed on the processed text. Using many algorithms and transformations this processed text is finally converted into a speech format. This entire process involves the synthesizing of speech. There are several APIs available to convert text to speech in Python. One of such APIs is the Google Text to Speech API commonly known as the gTTS API. gTTS is a very easy to use tool which converts the text entered, into audio which can be saved as a mp3 file. The gTTS API supports several languages including English, Hindi, Tamil, French, German and many more.

RESULT ANALYSIS

In this project, we have till now attained the required text to speech conversion to successfully convert the detected object into an audio signal using the gTTs Module with the help of the iPython audio library. The final detected output can be seen in the following figure where, using the YOLOv_4 model trained on the MS COCO dataset, we have detected the objects through the laptop camera and conversion of the same to speech is attained by using the gTTS module with the help of iPython audio library.

CONCLUSION

By implementing this system, visually impaired can easily listen whatever they want to listen. And with the help of the translation tools, he can convert the text to the desired language and then again by using the Google speech recognition tool he can convert that changed text into voice. By that they can be independent. And it is less cost compared to other implementations. Text-to-Speech device can change the text image input into sound with a performance that is high enough and a readability tolerance of less than 2%, with the average time processing less than three minutes for A4 paper size. This portable device, doesnot require internet connection, and can be used independently by people. Through this method, we can make editing process of books or web pages easier.

FUTURE SCOPE OF THIS WORK

The proposed model of this project is already in a feasible state of use but there are some future works for this model.

1) Recognizing process of large texts is sometimes slow, for that reason introducing of some distributed technique is reserved for future work.

2) Now this model can read only English language, but in future it should allow Assamese language as well.

3) If any image or diagram is present in between text, then recognition of that image or diagram is an important task, which is also reserved for future work.

By extending these future works, many numbers of people will get more benefits from this reader.

REFERENCES

[1] Abhishek, R.; Kumar, K. N.; Karthik, R.; Puskhal, R.; Kumar, S. A. Smart Gadget Product Label Reading Using OCR Algorithm & TTS Engine. International Journal of New Technology and Research 4 (4), 70–72.

[2] S, B.; D, L. Image to Audio Conversion Using Portable Camera. Journal of Electrical & Electronic Systems 2018, 07 (03). https://doi.org/10.4172/2332-0796.1000268.

[3] Fisher, R.; Perkins, S.; Walker, A.; Wolfart, E. Adaptive Thresholding https://homepages.inf.ed.ac.uk/rbf/HIPR2/adpthrsh.htm (accessed 2023 -02 -06).

.