

Devendra Kumar Sharma
Sheng-Lung Peng
Rohit Sharma
Dmitry A. Zaitsev *Editors*

Micro-Electronics and Telecommunication Engineering

Proceedings of 5th ICMETE 2021

Lecture Notes in Networks and Systems

Volume 373

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 worldwide 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>

Devendra Kumar Sharma · Sheng-Lung Peng ·
Rohit Sharma · Dmitry A. Zaitsev
Editors

Micro-Electronics and Telecommunication Engineering

Proceedings of 5th ICMETE 2021



Springer

Editors

Devendra Kumar Sharma
Department of Electronics
and Communication Engineering
SRM Institute of Science and Technology
Ghaziabad, India

Rohit Sharma
Department of Electronics
and Communication Engineering
SRM Institute of Science and Technology
Ghaziabad, India

Sheng-Lung Peng
National Taipei University of Business
Taipei, Taiwan

Dmitry A. Zaitsev
Odessa State Environmental University
Odessa, Ukraine

ISSN 2367-3370

Lecture Notes in Networks and Systems

ISBN 978-981-16-8720-4

<https://doi.org/10.1007/978-981-16-8721-1>

ISSN 2367-3389 (electronic)

ISBN 978-981-16-8721-1 (eBook)

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022, corrected publication 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

The book presents high-quality papers from the Fifth International Conference on Microelectronics and Telecommunication Engineering (ICMTE 2021). It discusses the latest technological trends and advances in major research areas such as microelectronics, wireless communications, optical communication, signal processing, image processing, big data, cloud computing, artificial intelligence, and sensor network applications. This book includes the contributions of national/international scientists, researchers, and engineers from both academia and the industry. The contents of this volume will be useful to researchers, professionals, and students alike.

Ghaziabad, India
Taipei, Taiwan
Ghaziabad, India
Odessa, Ukraine

Devendra Kumar Sharma
Sheng-Lung Peng
Rohit Sharma
Dmitry A. Zaitsev

Contents

Opportunistic Spectrum Distribution Protocol for Wireless Sensor Networks	1
Nellor Kapileswar, Polasi Phani Kumar, N. Uttam Reddy, B. Akhil Jaichandra Reddy, and B. Padmavathi	
Harmonic Reduction in a Three-Phase Voltage Source Inverter Using RLC Filter and FFT	17
Vivek Pratap Singh, M. A. Ansari, and Nivedita Singh	
Fault Detection and Classification Using Fuzzy Logic Controller and ANN	25
Rishabh Verma and M. A. Ansari	
Comparative Analysis of Phasor-Phasor and Detailed-Phasor Models of Regulating Transformer	37
Nitin Karnwal, Mukul Singh, Nidhi Singh, and M. A. Ansari	
Intelligent Condition Monitoring of Electrical Assets Using Infrared Thermography and Image Processing Techniques	47
Rishabh Anand and M. A. Ansari	
Elderly Care Using Generative Adversarial Networks (GANs) on Deep Video Analysis	61
S. Rajasekaran and G. Kousalya	
Performance Analysis of kNN-Based Image Demosaicing for Variable Window Sizes	71
Gurjot Kaur Walia and Jagroop Singh Sidhu	
A Novel Surface Crack Detection and Dimension Estimation Using Image Processing Technique	81
K. Shreyank, K. Yukta, N. Sowmya, M. Komal, V. S. Saroja, and S. Suhas	

FDTD Method-Based One-Dimensional Interaction of Electromagnetic Wave with Skin Tissue	91
Danvir Mandal and Indu Bala	
Performance Enhancement of UAV-Based Cognitive Radio Network	97
Indu Bala, Danvir Mandal, and Ankur Singhal	
Blockchain Technology with Supply Chain Management: Components, Opportunities and Possible Challenges	107
Ayasha Malik, Abhijit Kumar, Jaya Srivastava, and Bharat Bhushan	
Preventing and Detecting Intrusion of Cyberattacks in Smart Grid by Integrating Blockchain	119
Avinash Kumar, Bharat Bhushan, and Parma Nand	
Design of Low Switching Pattern Generator for BIST Architecture	131
Sachin Kumar Pandey and C. Paramasivam	
Connecting Blockchain with IoT— A Review	141
R. Anusha, Mohamed Yousuff, Bharat Bhushan, J. Deepa, J. Vijayashree, and J. Jayashree	
DAMONA: A Multi-robot System for Collection of Waste in Ocean and Sea	149
Riya Sil, Anwesha Das, Firdous Shamim, Aninda Chowdhury, Ritam Mukherjee, Sazid Ali, and Rohit Sharma	
Hardware Implementation and Comparison of OE Routing Algorithm with Extended XY Routing Algorithm for 2D Mesh on Network on Chip	159
Radha Velangi and S. S. Kerur	
Smart Medical Robotic Kit	173
Devesh Sonker, Akanksha, and Ranjeeta Yadav	
Design of a Secure Blockchain-Based Toll-Tax Collection System	183
Debashis Das, Sourav Banerjee, and Utpal Biswas	
An Overview of Security Services and Trust-Based Authentication Schemes in VANET	193
M. Gayathri and C. Gomathy	
Design of High-Speed Latched Comparator Used in Analog to Digital Converters by Using 180 nm Technology	207
Krishna Mohan Pandey, Yogendra Narayan Prajapati, and Naresh Kumar	
Healthcare Assistant—A Tool to Predict Disease Using Machine Learning	221
Sujata Joshi, Harish Kumar, Jagadish Babu, Akhil Raju, and Mahammad Nihaz	

A Comprehensive Survey on Topic Modeling in Text Summarization	231
G. Bharathi Mohan and R. Prasanna Kumar	
Design and Analysis of Microstrip High-Frequency Filter	241
V. Reji, R. Arthi, and C. T. Manimegalai	
Arduino UNO-Based Smart Hand Gloves for Physically Challenged People	251
B. V. Santhosh Krishna, L. Sowmya, Neetha Nataraj, Nivedita Salimath, and Shivani Yadav	
An Enhanced Haze Removal: Using DCP and Enriched-Invariant Features	257
Reenie Tanya, Faaiz Akhtar, and Mahipal	
Comparative Analysis on the Prediction of Road Accident Severity Using Machine Learning Algorithms	269
Manoj Kushwaha and M. S. Abirami	
Automatic Sickle Cell Anaemia Detection Using Image Processing Technique	281
V. Kiruthika, A. L. Vallikannu, and G. Vimalarani	
Recognition and Counting of an Object Using Yolo and CNN	289
R. Vallikannu, V. Kiruthika, and M. Kaviya	
A Novel Solution for Carrier Frequency Offset (CFO) Minimization for Efficient 5G Wireless Communication Using OFDM for Internet of Things (IoT)	297
G. Elavel Visuvanathan and T. Jaya	
Middleware and Security Requirements for Internet of Things	309
Bharat Bhushan	
Secure Multipath Key Establishment Solution in WSN	323
Charu Sharma and Rohit Vaid	
Controlling Node Failure Localization in Data Networks Using Probing Mechanisms	331
K. Saritha, Bingi Manorama Devi, Muralidhar Kurni, Debabrata Samanta, and Niju P. Joseph	
Rendering View of Kitchen Design Using Autodesk 3Ds Max	339
Ritwika Das Gupta, Debabrata Samanta, and Niju P. Joseph	
Preserving Security and Privacy in IoT Using Machine Learning and Trust Management	349
Avinash Kumar, Trisha Bhowmik, Rohit Sharma, and Abhishek Bhardwaj	

Design and Implementation Wireless Sensor Node with Security Algorithm Based on Microcontroller ESP8266	363
Zahraa A. Msekh and Alyaa A. Msekh	
Anomaly Detection of Ceramic Images Using Bag of Features	373
Zaid T. Omer and Amel H. Abbas	
An Analysis of the Effect of Wireless Network Channel on Radio Fingerprint Authentication	383
Mohammed Mahdi Salih Altufaili, Hussein Ali Mousa, and Yasir Abdulzahra Flaiyঃ Alaabedi	
Analysis of IoT Device Network Traffic: Thinking Toward Machine Learning	393
Vian Adnan Ferman and Mohammed Ali Tawfeeq	
A Survey of Scheduling Tasks in Big Data: Apache Spark	405
Balqees Talal Hasan and Dhuha Basheer Abdullah	
Unmanned Ground Vehicle Prototype Development for Search Evacuation and Defense Based on IoRT	415
Hiba A. Gumar, Baraa M. Albaker, and Mohammed N. Al-Turfi	
Data Mining Analysis Models Based on Prospective Detection of Infectious Disease	425
Ahmed J. Obaid	
Intelligent Parameter Tuning Using Deep Q-Network for RED Algorithm in Adaptive Queue Management Systems	439
Ayman Basheer, Hassan Jaleel Hassan, and Gaida Muttasher	
Secure Smart Contract Based on Elliptic Curve in Property Exchange Applications Using Blockchain	447
Noor Sabah	
Performance Investigation of Short Channel Impacts and Analog/RF Figure of Merits (FOMs) of SOI-FinFET	457
Nishant Srivastava and Prashant Mani	
IoT-Based Surveillance and Face Detection BOT	467
Veral Agarwal, Nipun Tyagi, and Rachit Patel	
Atmospheric Turbulence Effects on Bit Error Rate in Lognormal and Negative Exponential Channel in FSO Link	477
Priyanka Bhardwaj, Manidipa Roy, and Sanjay Kumar Singh	
Simulation and Design of Mach-Zehnder Interferometer	485
Priyanka Bhardwaj, Manidipa Roy, and Sanjay Kumar Singh	
IoT-Based Smart Home Security and Automation System	497
Partha Chakraborty and Sajeda Sultana	

A Fuzzy Model for Selection of Information and Location-Based Security Attributes in Cloud Environment	507
Deepika, Rajneesh Kumar, and Dalip	
Effect of esports Among Students in COVID Era	517
Ankit Bisht, Hitesh Kumar Sharma, and Tanupriya Choudhury	
Handwritten Digit Recognition with Neural Network	525
Satyanarayana Malla V, Hitesh Kumar Sharma, and Tanupriya Choudhury	
Metastability Mitigation and Error Masking of High-Speed Flip-Flop	533
Reshma Kumari, Sneha Pandey, Swarnima, and Surya Deo Choudhary	
Student Performance Prediction Using Technology of Machine Learning	541
Kaushal Kishor, Rahul Sharma, and Manish Chhabra	
Low-Power IoT Architecture, Challenges, and Future Aspects	553
Saurabh Sambhav and Shilpi Singh	
ANURL: An Innovative Management Scheme for Web Uniform Resource Locators	561
Ashish Karn, Suyash Thakur, and Pankaj Badoni	
A Novel Voice Recognition System with Artificial Intelligence	573
Sabhav Gupta, Adarsh Pandey, Shivam Naruka, and Keshav Gupta	
Investigation on Malware Detection Using Deep Learning Methods for Sustainable Development	581
M. Anusha and M. Karthika	
RSSI Strength Measurement in Wireless Sensor Network with and Without Obstacles	593
Santosh Anand, Roshan Muralidharan, K. Manoj, C. Shreyas, and Hruthik Kariappa	
Anatomy of Virtual Machine Placement Techniques in Cloud	609
Chayan Bhatt and Sunita Singhal	
Effectual Attendance Application for Remote Education During Era of COVID-19	627
Mohitsinh Parmar, Shailesh Khant, and Atul Patel	
Anomaly Detection in Thermal Images of Perishable Items Using Deep Learning	647
G. Ramyapriyanandhini, T. Bagyammal, Latha Parameswaran, and Karthikeyan Vaiapury	

Median Filtering Detection Using Ensemble Methods	661
Sajjad Ahmed and Saiful Islam	
Creation and Segmentation of Image Dataset of Mung Bean Plant Leaf	669
Akruti Naik, Hetal Thaker, and Nirav Desai	
Road Accident Analysis Using ML Classification Algorithms and Plotting Black Spot Areas on Map	685
Manu Tiwari, Piyush Nagar, Gautam Arya, and Surendra Singh Chauhan	
Detecting Hate Speech and Offensive Language Using Transformer Techniques	703
Mahin Bindra, Bhavya Sharma, and Nipun Bansal	
Enabling Multi-Factor Authentication and Verification in Searchable Encryption	717
Sai Deepika Panguluri, K. V. Lakshmy, and Chungath Srinivasan	
Music Genre Classification Using CNN and RNN-LSTM	729
Rohan Gupta, Shivam Ashish, Himanshu Shekhar, and MS. Deepica S. Dominic	
House Prices Using Machine Learning Algorithms	747
Vishal, Amit Singh, and Chirag Chaudhary	
HEVC Encoding and Decoding using Fast Algorithm for Intra Frame Partition-Prediction Using DTCWT	759
V. Madhurima and K. Padmapriya	
A Perspective Toward 6G Connecting Technology	775
Neha Katiyar, Jyoti Srivastava, and Kushall Pal Singh	
MRI Breast Tumor Extraction Using Possibilistic C Means and Classification Using Convolutional Neural Network	795
R. Sumathi and V. Vasudevan	
An EEG Atomized Artefact Removal Algorithm: A Review	805
Rudra Bhanu Satpathy and G. P. Ramesh	
A Journey of Artificial Intelligence and Its Evolution to Edge Intelligence	817
P. Britto Corthis and G. P. Ramesh	
Correction to: Micro-Electronics and Telecommunication Engineering	C1
Devendra Kumar Sharma, Sheng-Lung Peng, Rohit Sharma, and Dmitry A. Zaitsev	
Author Index	827

Editors and Contributors

About the Editors

Devendra Kumar Sharma received his B.E. degree in Electronics Engineering from Motilal Nehru National Institute of Technology, Allahabad in 1989, M.E. degree from Indian Institute of Technology Roorkee, Roorkee in 1992 and Ph.D. degree from National Institute of Technology, Kurukshetra, India in 2016. He served PSU in different positions for more than 8 years in QA and Testing/R&D departments. Dr. Sharma joined the department of Electronics and Communication Engineering, Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India in October 2000 as Senior Lecturer and worked there till April 2018 at different capacities of Assistant Professor, Associate Professor, Professor and Dean Academics. He is currently working as Professor and Dean of SRM Institute of Science and Technology, Delhi-NCR Campus, Ghaziabad, India. Dr. Sharma has authored many papers in several international journals and conferences of repute. His research interests include VLSI interconnects, Electronic Circuits, Digital Design, Testing and Signal Processing. He is a reviewer of many international journals belonging to various publication houses such as IEEE, Elsevier, Emerald, World Scientific, and Springer. Dr. Sharma has participated in many International and National conferences as Session Chair, and member of Steering, Advisory and Technical Program Committees. He has been the Editor for several books/conference proceedings and has been the Organizing Chair for several international conferences. He is life member of ISTE, Fellow of IETE and Senior Member of IEEE.

Sheng-Lung Peng is a Professor of the Department of Creative Technologies and Product Design in National Taipei University of Business Taiwan, Honorary Professor in Beijing Information Science and Technology University, and Visiting Professor in Ningxia Institute of Science and Technology, China. He received a BS degree in Mathematics from National Tsing Hua University, and the MS and Ph.D. degrees in Computer Science from the National Chung Cheng University and National Tsing Hua University, Taiwan, respectively. He is an honorary Professor

of Beijing Information Science and Technology University, China, and a visiting Professor of the Ningxia Institute of Science and Technology, China. He is also an adjunct Professor at Mandsaur University, India. He serves as the secretary-general of the ACM-ICPC Contest Council for Taiwan and the regional director of the ICPC AsiaTaipei-Hsinchu site. He is a director of the Institute of Information and Computing Machinery, of Information Service Association of Chinese Colleges and Taiwan Association of Cloud Computing. He is also a supervisor of the Chinese Information Literacy Association, of Association of Algorithms and Computation Theory. Dr. Peng has edited several special issues at journals, such as *Soft Computing*, *Journal of Internet Technology*, *Journal of Real-Time Image Processing*, *International Journal of Knowledge and System Science*, *MDPI Algorithms*, and so on. He is also a reviewer for more than 10 journals such as *IEEE Access and Transactions on Emerging Topics in Computing*, *IEEE/ACM Transactions on Networking*, *Theoretical Computer Science*, *Journal of Computer and System Sciences*, *Journal of Combinatorial Optimization*, *Journal of Modelling in Management*, *Soft Computing*, *Information Processing Letters*, *Discrete Mathematics*, *Discrete Applied Mathematics*, *Discussiones Mathematicae Graph Theory*, and so on. His research interests are in designing and analyzing algorithms for Bioinformatics, Combinatorics, Data Mining, and Networks areas in which he has published over 100 research papers.

Rohit Sharma is currently an Assistant Professor in the Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Delhi NCR Campus Ghaziabad, India. He is an active member of ISTE, IEEE, ICS, IAENG, and IACSIT. He is an editorial board member and reviewer of more than 12 international journals and conferences, including the topmost journal *IEEE Access* and *IEEE Internet of Things Journal*. He serves as a Book Editor for 7 different titles to be published by CRC Press, Taylor & Francis Group, USA and Apple Academic Press, CRC Press, Taylor & Francis Group, USA, Springer, etc. He has received the Young Researcher Award in “2nd Global Outreach Research and Education Summit and Awards 2019” hosted by Global Outreach Research and Education Association (GOREA). He is serving as Guest Editor in SCI journal of Elsevier. He has actively been an organizing end of various reputed International conferences. He is serving as an Editor and Organizing Chair to 3rd Springer International Conference on Microelectronics and Telecommunication (2019), and have served as the Editor and Organizing Chair to 2nd IEEE International Conference on Microelectronics and Telecommunication (2018), Editor and Organizing Chair to IEEE International Conference on Microelectronics and Telecommunication (ICMETE-2016) held in India, Technical Committee member in “CSMA2017, Wuhan, Hubei, China”, “EEWC 2017, Tianjin, China” IWMSE2017 “Guangzhou, Guangdong, China”, “ICG2016, Guangzhou, Guangdong, China” “ICCEIS2016 Dalian Liaoning Province, China”.

Dmitry A. Zaitsev received the Engineering degree in Applied Mathematics from Donetsk Polytechnic Institute, Donetsk, Ukraine, in 1986, the Ph.D. degree in Automated Control from the Kiev Institute of Cybernetics, Kiev, Ukraine, in 1991, and the Dr. Sc. degree in Telecommunications from the Odessa National Academy of

Telecommunications, Odessa, Ukraine, in 2006. He is a Professor of Information Technology at Odessa State Environmental University, Ukraine since 2019. He developed the analysis of infinite Petri nets with regular structure, the decomposition of Petri nets in clans, generalized neighborhood for cellular automata, and the method of synthesis of fuzzy logic function given by tables. He developed Opera-Topaz software for manufacture operative planning and control; a new stack of networking protocols E6 and its implementation within Linux kernel; Petri net analysis software Deborah, Adriana, and ParAd; models of TCP, BGP, IOTP protocols, Ethernet, IP, MPLS, PBB, and Bluetooth networks. His current research interests include Petri net theory and its application in networking, computing and automated manufacture. Recently he started working in the area of exascale computing applying his theory of clans to speed-up solving sparse linear systems on parallel and distributed architectures. He was a co-director of joint projects with China and Austria. Recently he has been a visiting professor to Technical University of Dortmund, Germany on DAAD scholarship, to University of Tennessee Knoxville, USA on Fulbright scholarship and to Eindhoven University of Technology, Netherlands. He published a monograph, 3 book chapters and more than a hundred of papers including issues listed in JCR. He is a senior member of ACM and IEEE.

Contributors

Amel H. Abbas Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq

Dhuha Basheer Abdullah Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

M. S. Abirami Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, India

Veral Agarwal Nagarro Software Pvt Ltd, Gurgaon, Haryana, India

Sajjad Ahmed Baba Ghulam Shah Badshah University, Rajouri, India

Akanksha ABES Engineering College, Ghaziabad, Uttar Pradesh, India

B. Akhil Jaichandra Reddy Department of Electronic and Communication Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India

Faaiz Akhtar SRM Institute of Science and Technology, Ramapuram, Chennai, India

Mohammed N. Al-Turfi College of Engineering, Al-Iraqia University, Baghdad, Iraq

Yasir Abdulzahra Flaiyh Alaabedi Department of Medical Laboratory Techniques, College of Medical and Health Techniques, University of Alkafeel, Al Najaf, Iraq

Baraa M. Albaker College of Engineering, Al-Iraqia University, Baghdad, Iraq

Sazid Ali Department of Computer Science & Engineering, Adamas University, Kolkata, India

Mohammed Mahdi Salih Altufaili Department of Computer Engineering Techniques, College of Engineering Techniques, University of Alkafeel, Al Najaf, Iraq

Rishabh Anand Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

Santosh Anand Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India

M. A. Ansari Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

M. Anusha PG & Research Department of Computer Science, National College (Autonomous), Affiliated To Bharathidasan University, Trichirappali, Tamilnadu, India

R. Anusha Vellore Institute of Technology, Vellore, Tamilnadu, India

R. Arthi Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India

Gautam Arya School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

Shivam Ashish School of Computing Science and Engineering, Galgotias University, Greater Noida, India

Jagadish Babu Nitte Meenakshi Institute of Technology, Bangalore, India

Pankaj Badoni School of Computer Science, University of Petroleum and Energy Studies, Bidholi, Dehradun, Uttarakhand, India

T. Bagyammal Computer Science Department, Amrita Vishwa Vidyapeetham, Coimbatore, India

Indu Bala Lovely Professional University, Phagwara, Punjab, India

Sourav Banerjee Kalyani Government Engineering College, Kalyani, India

Nipun Bansal Department of Computer Science, Delhi Technological University, New Delhi, India

Ayman Basheer Department of Computer Engineering, University of Technology, Baghdad, Iraq

Abhishek Bhardwaj School of Engineering and Technology (SET), Sharda University, Noida, India

Priyanka Bhardwaj ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Chayan Bhatt Department of Computer Science and Engineering, School of Computing & Information Technology, Manipal University Jaipur, Jaipur, Rajasthan, India

Trisha Bhowmik School of Engineering and Technology (SET), Sharda University, Noida, India

Bharat Bhushan School of Engineering and Technology (SET), Sharda University, Greater Noida, India

Mahin Bindra Department of Computer Science, Delhi Technological University, New Delhi, India

Ankit Bisht School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

Utpal Biswas University of Kalyani, Kalyani, India

P. Britto Corthis Research Scholar, Department of ECE, St. Peter's Institute of Higher Education and Research, Chennai, India

Partha Chakraborty Department of Computer Science and Engineering, Comilla University, Cumilla, Bangladesh

Chirag Chaudhary School of Computer Science and Engineering, Galgotias University, Greater Noida, India

Surendra Singh Chauhan School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

Manish Chhabra Department of CSE, Koneru Lakshmaiah Education Foundation, Hyderabad, India

Surya Deo Choudhary Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida, India

Tanupriya Choudhury School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

Aninda Chowdhury Department of Computer Science & Engineering, Adamas University, Kolkata, India

Dalip MMCTBM, M.M. (Deemed to be University), Mullana, Ambala, India

Ritwika Das Gupta Department of Computer Science, CHRIST Deemed to be University, Bengaluru, India

Anwesha Das Department of Computer Science & Engineering, Adamas University, Kolkata, India

Debashis Das University of Kalyani, Kalyani, India

J. Deepa Veltech Rangarajan & Dr Sakunthala R&D Institute of Science & Technology, Chennai, India

Deepika M.M. Engineering College, M.M. (Deemed to be University), Mullana, Ambala, India;
UIE-CSE, Chandigarh University, Mohali, India

Nirav Desai Department of Computer Applications, Atmiya University, Rajkot, Gujarat, India

Bingi Manorama Devi Department of CSE, K.S. R. M College of Engineering, Kadapa, Andhra Pradesh, India

MS. Deepica S. Dominic School of Computing Science and Engineering, Galgotias University, Greater Noida, India

G. Elavel Visuvanathan Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India

Vian Adnan Ferman Computer Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

M. Gayathri Electronics and Communication Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani Campus, Vadapalani, Chennai, Tamil Nadu, India

C. Gomathy Electronics and Communication Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani Campus, Vadapalani, Chennai, Tamil Nadu, India

Hiba A. Gumar College of Engineering, Al-Iraqia University, Baghdad, Iraq

Keshav Gupta Galgotias University, Greater Noida, India

Rohan Gupta School of Computing Science and Engineering, Galgotias University, Greater Noida, India

Sabhav Gupta Galgotias University, Greater Noida, India

Balqees Talal Hasan Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

Hassan Jaleel Hassan Department of Computer Engineering, University of Technology, Baghdad, Iraq

Saiful Islam ZHCET, Aligarh Muslim University, Aligarh, India

T. Jaya Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India

J. Jayashree Vellore Institute of Technology, Vellore, Tamilnadu, India

Niju P. Joseph Department of Computer Science, CHRIST Deemed to be University, Bengaluru, India

Sujata Joshi Nitte Meenakshi Institute of Technology, Bangalore, India

Nellor Kapileswar Department of Electronic and Communication Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India

Hruthik Kariappa Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India

Ashish Karn MultiPhase Flows Laboratory, Department of Mechanical Engineering, School of Engineering, University of Petroleum and Energy Studies, Bidholi, Dehradun, Uttarakhand, India

Nitin Karnwal Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

M. Karthika PG & Research Department of Computer Science, National College (Autonomous), Affiliated To Bharathidasan University, Trichirappalli, Tamilnadu, India

Neha Katiyar Indian Institute of Management, Rohtak, India

M. Kaviya Department of ECE, Hindustan Institute of Technology and Science, Chennai, India

S. S. Kerur SDMCET, Dharwad, India

Shailesh Khant Smt. ChandabenMohanbhai Patel Institute of Computer Applications, Charusat University, Changa, India

V. Kiruthika Department of ECE, Hindustan Institute of Technology and Science, Chennai, India

Kaushal Kishor Department of IT, ABES Institute of Technology, Ghaziabad, U.P, India

M. Komal KLE Technological University, Vidyanagar, Hubbali, India

G. Kousalya Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore, Tamilnadu, India

B. V. Santhosh Krishna Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, India

Abhijit Kumar Noida Institute of Engineering Technology (NIET), Greater Noida, India

Avinash Kumar School of Engineering and Technology (SET), Sharda University, Greater Noida, India

Harish Kumar Nitte Meenakshi Institute of Technology, Bangalore, India

Naresh Kumar Department of Computer Science and Engineering, Quantum University, Roorkee, U.K., India

R. Prasanna Kumar Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Chennai, India

Rajneesh Kumar M.M. Engineering College, M.M. (Deemed to be University), Mullana, Ambala, India

Reshma Kumari Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida, India

Muralidhar Kurni Department of Computer Science, SoS, GITAM (Deemed to be University), Hyderabad, Telangana, India

Manoj Kushwaha Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, India

K. V. Lakshmy TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

V. Madhurima Department of ECE, UCE, JNTUK, Kakinada, Andhra Pradesh, India

Mahipal SRM Institute of Science and Technology, Ramapuram, Chennai, India

Ayasha Malik Noida Institute of Engineering Technology (NIET), Greater Noida, India

Satyanarayana Malla V School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

Danvir Mandal Lovely Professional University, Phagwara, Punjab, India

Prashant Mani SRM Institute of Science and Technology, Modinagar, Ghaziabad, India

C. T. Manimegalai Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India

K. Manoj Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India

G. Bharathi Mohan Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Chennai, India

Hussein Ali Mousa Department of Computer Engineering Techniques, College of Engineering Techniques, University of Alkafel, Al Najaf, Iraq

Alyaa A. Msekh Computer Department, College of Science for Woman, Babylon University, Hillah, Iraq

Zahraa A. Msekh Energy Department, College of Engineering Al Musaib, Babylon University, Hillah, Iraq;

Engineering Technologies for Medical Devices Department, Hilla University College, Hillah, Iraq

Ritam Mukherjee Department of Computer Science & Engineering, Adamas University, Kolkata, India

Roshan Muralidharan Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India

Gaida Muttasher Department of Computer Engineering, University of Technology, Baghdad, Iraq

Piyush Nagar School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

Akruti Naik Department of Computer Applications, Atmiya University, Rajkot, Gujarat, India

Parma Nand School of Engineering and Technology (SET), Sharda University, Greater Noida, India

Shivam Naruka Galgotias University, Greater Noida, India

Neetha Nataraj Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, India

Mahammad Nihaz Nitte Meenakshi Institute of Technology, Bangalore, India

Ahmed J. Obaid Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Kufa, Iraq

Zaid T. Omer Mustansiriyah University, Baghdad, Iraq

K. Padmapriya Department of ECE, UCE, JNTUK, Kakinada, Andhra Pradesh, India

B. Padmavathi Department of Electronic and Communication Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India

Adarsh Pandey Galgotias University, Greater Noida, India

Krishna Mohan Pandey Department of Computer Science and Engineering, DVSIEET, Meerut, U.P., India

Sachin Kumar Pandey Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

Sneha Pandey Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida, India

Sai Deepika Panguluri TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

C. Paramasivam Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

Latha Parameswaran Computer Science Department, Amrita Vishwa Vidyapeetham, Coimbatore, India

Mohitsinh Parmar Smt. ChandabenMohanbhai Patel Institute of Computer Applications, Charusat University, Changa, India

Atul Patel Smt. ChandabenMohanbhai Patel Institute of Computer Applications, Charusat University, Changa, India

Rachit Patel Department of Electronics & Communication Engineering, ABES Institute of Technology, Ghaziabad, Uttar Pradesh, India

Polasi Phani Kumar Department of Electronic and Communication Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India

Yogendra Narayan Prajapati Department of Computer Science and Engineering, DVSIEET, Meerut, U.P., India

S. Rajasekaran Department of Information Technology, KGSL Institute of Technology, Coimbatore, Tamilnadu, India

Akhil Raju Nitte Meenakshi Institute of Technology, Bangalore, India

G. P. Ramesh Department of ECE, St. Peters Institute of Higher Education and Research, Chennai, India

G. Ramyapriyanandhini Computer Science Department, Amrita Vishwa Vidyapeetham, Coimbatore, India

V. Reji Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India

Manidipa Roy ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Noor Sabah University of Anbar, Anbar, Iraq

Nivedita Salimath Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, India

Debabrata Samanta Department of Computer Science, CHRIST Deemed to be University, Bengaluru, India

Saurabh Sambhav Amity School of Engineering and Technology, Amity University Campus, Patna, India

K. Saritha S. V. Degree & P. G. College, Anantapur, Andhra Pradesh, India

V. S. Saroja KLE Technological University, Vidyanagar, Hubballi, India

Rudra Bhanu Satpathy Department of ECE, St. Peters Institute of Higher Education and Research, Chennai, India

Firdous Shamim Department of Computer Science & Engineering, Adamas University, Kolkata, India

Bhavya Sharma Department of Computer Science, Delhi Technological University, New Delhi, India

Charu Sharma CSE Department, M. M. Engineering College, M.M (Deemed to be University), Mullana, Ambala, Haryana, India

Hitesh Kumar Sharma School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

Rahul Sharma ABES Institute of Technology, Ghaziabad, U.P, India

Rohit Sharma Department of Electronics & Communication Engineering, SRM Institute of Science & Technology, Ghaziabad, India;

Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India

Himanshu Shekhar School of Computing Science and Engineering, Galgotias University, Greater Noida, India

ShivaniYadav Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, India

K. Shreyank KLE Technological University, Vidyanagar, Hubbali, India

C. Shreyas Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India

Jagroop Singh Sidhu Department of ECE, DAVIET, Jalandhar, Punjab, India

Riya Sil Department of Computer Science & Engineering, Adamas University, Kolkata, India

Ankur Singhal Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

Sunita Singhal Department of Computer Science and Engineering, School of Computing & Information Technology, Manipal University Jaipur, Jaipur, Rajasthan, India

Amit Singh School of Computer Science and Engineering, Galgotias University, Uttar Pradesh, Greater Noida, India

Kushall Pal Singh Malaviya National Institute of Technology, Jaipur, India

Mukul Singh Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

Nidhi Singh Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

Nivedita Singh Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

Sanjay Kumar Singh ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Shilpi Singh Amity School of Engineering and Technology, Amity University Campus, Patna, India

Vivek Pratap Singh Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

Devesh Sonker ABES Engineering College, Ghaziabad, Uttar Pradesh, India

L. Sowmya Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, India

N. Sowmya KLE Technological University, Vidyanagar, Hubballi, India

Chungath Srinivasan TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Jaya Srivastava Noida Institute of Engineering Technology (NIET), Greater Noida, India

Jyoti Srivastava Madan Mohan Malviya University of Technology, Gorakhpur, India

Nishant Srivastava SRM Institute of Science and Technology, Modinagar, Ghaziabad, India

S. Suhas KLE Technological University, Vidyanagar, Hubballi, India

Sajeda Sultana Department of Computer Science and Engineering, Comilla University, Cumilla, Bangladesh

R. Sumathi Department of Computer Science and Engineering, School of Computing, Kalasalingam Academy of Research and Education, Krishnankovil, Tamil Nadu, India

Swarnima Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida, India

Reenie Tanya SRM Institute of Science and Technology, Ramapuram, Chennai, India

Mohammed Ali Tawfeeq Computer Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

Hetal Thaker Department of Computer Applications, Atmiya University, Rajkot, Gujarat, India

Suyash Thakur School of Computer Science, University of Petroleum and Energy Studies, Bidholi, Dehradun, Uttarakhand, India

Manu Tiwari School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

Nipun Tyagi Cisco Systems India Pvt Ltd, Bangalore, Karnataka, India

N. Uttam Reddy Department of Electronic and Communication Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India

Karthikeyan Vaiapury TCS Research and Innovation, IITM Research Park, Chennai, India

Rohit Vaid CSE Department, M. M. Engineering College, M.M (Deemed to be University), Mullana, Ambala, Haryana, India

A. L. Vallikannu Hindustan Institute of Technology and Science, Chennai, India

R. Vallikannu Department of ECE, Hindustan Institute of Technology and Science, Chennai, India

V. Vasudevan School of Computing, Kalasalingam Academy of Research and Education, Krishnankovil, Tamil Nadu, India

Radha Velangi Govt Polytechnic, Hubballi, India

Rishabh Verma Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

J. Vijayashree Vellore Institute of Technology, Vellore, Tamilnadu, India

G. Vimalarani Hindustan Institute of Technology and Science, Chennai, India

Vishal School of Computer Science and Engineering, Galgotias University, Greater Noida, India

Gurjot Kaur Walia Electronics and Communication Engineering, I. K. Gujral Punjab Technical University, Kapurthala, Punjab, India

Ranjeeta Yadav ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Mohamed Yousuff Vellore Institute of Technology, Vellore, Tamilnadu, India

K. Yukta KLE Technological University, Vidyanagar, Hubbali, India

Opportunistic Spectrum Distribution Protocol for Wireless Sensor Networks



Nellor Kapileswar, Polasi Phani Kumar, N. Uttam Reddy,
B. Akhil Jaichandra Reddy, and B. Padmavathi

Abstract At present, the dynamic spectrum access (DSA) is the core technique that has been widely researched. As it allows the secondary users to access the white space, it increases the utilization of spectrum. This reduces the spectrum scarcity and detrimental interference induced by network glitch. It reworks on the occupied frequency bands and does not interfere the primary users. The physical layer defines the relationship between the device and transmission media in responsible of broadcasting the information. The medium access control (MAC) shares the media basis for many broadcast networks using DSA and static channelization. This MAC accords the data to multiple users by employing the spectrum sensing and spectrum accessing in the channel. MAC protocol broadcasts the channel by managing the traffic, operates the activities between the media, and avoids the collision in WSNs. Along with, the fuzzy logic develops the solution for the inaccuracy and insufficient network. In this paper, a novel protocol named opportunistic spectrum distribution protocol (OSDP) is proposed to provide a parallel communication process. The proposed protocol performs spectrum distribution on considering the transmission history, fuzziness present in the channel due to different paths, path changes, and priority of data transmission. It gives the dynamic response in transmitting the data with better efficiency and minimum error rate. The simulation results describe the benefits of the proposed OSDP over IEEE 802.15.4 and SMAC in parameters of end-to-end delay, throughput, and energy consumption.

Keywords Medium access control protocol · Opportunistic spectrum distribution protocol · End-to-end delay · Throughput · Energy consumption

N. Kapileswar (✉) · P. Phani Kumar · N. Uttam Reddy · B. Akhil Jaichandra Reddy ·
B. Padmavathi

Department of Electronic and Communication Engineering, SRM Institute of Science and
Technology, Ramapuram Campus, Chennai, Tamil Nadu, India
e-mail: kapilesn@srmist.edu.in

1 Introduction

Opportunistic spectrum access (OSA) has developed as a hopeful spectrum sharing model for the effective utilization of the spectrum [1–3]. It is an emerging technique with a huge demand as an increase in the numerous wireless protocols depends on and utilizing the internet. The secondary users in the OSA access the radio frequency spectrum and detect the unused broadcasted spectrum bands, namely spectrum holes or white space. The unused bands in the licensed user channels are accessed by the secondary users and opportunistically employed in transmission. This availability of spectrum in licensed user clears the problem of spectrum scarcity. Although the secondary user is unlicensed, interference in the spectrum holes disrupts the network. It cannot be controlled and more complicated to restrict and affects licensed users' communication by exploiting its spectrum band. To minimize the errors, the secondary users need high-frequency bands with frequency hopping. They should perform spectrum sensing for licensed users periodically. The report of the data undergoes to the control head, which decides the spectrum to be continued or switched to further channel. This phenomenon of observation and adjusting in a collaborative manner is called dynamic spectrum access. Thus OSA model fabricates the use of DSA opportunistically in various fields without additional spectrum frequency range.

The secondary user plays a pivotal role in spectrum sensing and spectrum accessing. The licensed user and secondary user would not permit simultaneous transmissions. The licensed user vacates when a secondary user exists and vice-versa. The wide transmission medium of licensed users should be divided into multiple narrow-band media to perform spectrum sensing. After this step, these channels will be accessed by the secondary user for opportunistic transmissions. This process works under the scheme of physical (PHY) and medium access control (MAC) layers techniques. The general functionality of the physical layer is to communicate the sensor information between two media.

In contrast, the medium access control makes the users access the information and shares to medium channels. Utmost, secondary users detect the white space by spectrum sensing, and spectrum access focuses on optimizing the transceiver design and energy efficiency in the PHY layer. As spectrum access relies on spectrum sensing, the channel encapsulates the spectrum and gives access to it to multiple users in the MAC layer by consuming energy. Thus, MAC protocol broadcasts the channel by managing the traffic, operates the activities between the media, and avoids the collision in WSNs. Traditional communication protocols in WSNs have more inconsiderable efficiency. Hence, MAC protocol interacts with cross-layer design to maximize the throughput of the network and enhance the lifetime.

This access technology allows the secondary users to share the licensed spectrum, thus improving the spectrum utilization. Many methods were proposed for the effective utilization of the available spectrum. The basic concept of all the opportunistic spectrum access methods is same: the individual secondary user accesses the vacant spectrum dynamically to improve the spectrum utilization while keeping less interference to primary users.

In this paper, we propose a novel MAC protocol named opportunistic spectrum distribution protocol for wireless sensor networks. In this OSDP, all the nodes in a neighborhood share the integrated spectrum allocated for communication. The OSDP implements the wireless network devices to sense and then dynamically access this limited spectrum, obtaining the most out of it.

OSDP provides parallel opportunities to a node to define data packet transmission to reduce end-to-end delay. The fuzzy inference of network is considered using three states (high-speed, average speed, and low speed). The fuzziness present in the wireless channel is described below. Fuzzy rules form the infrastructure of the fuzzy transformer. The fuzzy rules character has been created for WLAN to serve its specification for high efficiency and optimum management. The aim of this work is to create a novel OSDP for WSNs to achieve good throughput, lesser end-to-end delay, and energy consumption.

The remaining sections are standardized as follows: Sect. 2 reviews the related work. Section 3 details the proposed opportunistic spectrum distribution protocol. Section 4 provides the simulation results. Section 5 concludes the paper.

2 Related Works

Most of the proposed spectrum access protocols have ensured best performance in terms of spectrum utilization and energy consumption. Cognitive radio is an evolving technology with various benefits including cross-layer adaptation, spectrum distribution and cooperative networking. Many access MAC protocols based on CSMA, TDMA, hybrid, and cross-layer optimizations are popular today [4–6]. There are some schemes in which the secondary user must sense the channel continually even with no frames to transmit [7, 8]. These schemes are not suitable for applications with strict energy constraints. Some energy effective schemes are also proposed in the literature. One such nonpersistent scheme is proposed by Manuj et al. [9]. They have proposed a stochastic channel occupancy model-based channel distribution scheme. They have modeled the primary user channel state as alternating renewal process.

The secondary user tracks the elapsed time and records its next frame's probability of positive transmission. Later, they also proposed a cognitive multichannel MAC protocol [10]. This protocol supports many secondary users to transmit simultaneously in the primary network and ensures less interference probability.

Dan et al. [11] proposed a spectrum access technique that specifies when to switch off the cognitive radio. They have detailed the problems of spectrum sensing in primary and secondary users in cognitive radio networks [12–14]. The performance of the spectrum access protocols under non-stochastic channels is discussed in [15]. Best performance was achieved by using the spectrum sensing and access methods by assuming stochastic channels [16, 17]. Some prior knowledge is required to work with these channels.

Valehi [18] proposed a cognitive wireless sensor network that maximizes efficiency and boosts timespan for limited delay protocols. A cognitive wireless network

of the necessary sensors is taken into consideration, and the next set of secondary sensors uses the vacancies of the channels present in the first set of networks. The two sets of primary and secondary sensors will be transmitting the samples measured to the common sink node.

Lin [19] put forth his idea of switching to channels automatically, which will help in sharing the spectrum for the industrial wireless networks. The latest concept of equilibrium is introduced to show out the relationship between the evenness and performance of the sharing of the spectrum. A new set of rules called local equilibrium guided autonomous channel switching is introduced with all the required sensors placed. So through this method, spectrum sharing can be done equally.

Jayaprakasam [20] carried out a survey on distributed and collaborative beam-forming in wireless sensor networks. This survey has brought to our notice about the various classification, trends, and directions of research of DCBF in wireless sensor networks. The features and the various characteristics of the DCBF in the wireless sensor networks are presented.

Jain [21] has proposed a different approach for resolving the power and spectrum impacts in wireless sensor networks. A necessary programmed sensor is used to transmit the information to the base station. This system can be considered as an autonomous method for wireless technologies.

The latest technologies and the developments in the competitive world of electronics and also the wireless communications have made way for further developments in significant power and the low-cost WSNs. A routing protocol has been proposed known as grid clustering hierarchy (GCH) [22]. The proposed system of GCH divides the network into many virtual grids as needed. The parameter of the current average energy of the network is taken into consideration in this system.

Jadhav et al. [23] proposed an efficient routing protocol for wireless sensor networks known as “opportunistic routing” to maintain the lifetime of the network. The functioning of the OR is to broadcast the data through the wireless network from one node and tunes to multiple nodes. The data will be forwarded from node to intermediate nodes with reasonable probability and approaches the destination. A similar protocol named “EX-OR” is linked in the wireless network where the data is sent by attaching the batches to each packet. The packet which has a shorter distance will have higher priority and includes in the forward list. The packet with the nearest batch map will be prioritized, and nearest to the destination will be replaced further in the local batch map. Nevertheless, the transmission of data takes once by granting only one node, and data will not be updated if changed. Hence, OR has efficient performance in energy saving and sustains better throughput.

In this scheme, the authors [24] investigated the maximum coverage routing protocol for underwater wireless sensor networks to transmit the data to mobile sink in respective transmission range with minimal interval by covering the maximum network field. Transmission nodes discover the mobile sink by disseminating a message packet in the field. If the mobile sink receives the message, then it will abandon a new message to sensor nodes. Sensor nodes figure out the mobile sinks by determining the distance between them. Every sensor node of the same energy gives out a data packet to the closest mobile sink. Mobile sinks comprise the entire

region of the network field rotating in a clockwise direction. The sensor calculates the distance and sends each data packet to the mobile sink in the extent of transmission as they are moving. Thus, the MC protocol has better throughput and ensures minimum energy consumption with an efficient network lifetime.

Majid et al. [25] proposed the energy-efficient and balanced energy consumption cluster-based routing protocol for underwater wireless sensor networks to achieve network lifetime and maximum stability period. As the battery power and bandwidth are limited in underwater, loads are split up equally into nodes by balancing energy consumption. This mechanism performs by cluster-based routing by evading the depth base routing. The clustering technique reduces the transmissions to the sink and consumes energy. The transmission of data to sink from the nodes can be done in Type C, S, N, and BS. The cluster heads of each type are receiving data from the nodes and transferred to the nearest sink. The mobile sinks alter their position consistently for holding the load on sensor nodes. The packet sends a message to locate the sink and start the network initialization, and when it finds out, the data will be exchanged. Consequently, the sink moment raises the network initialization and upgrades the sensor nodes. Thus, the routing protocol EBECRP provides network lifetime, high throughput, and avoids packet drop.

Duy Tan and Dinh Viet [26] designed a routing protocol, “sleep scheduled and tree-based clustering approach routing algorithm” for a better network lifetime and energy consumption. A spanning tree is installed where the root communication occurs through cluster head from sensor nodes. Each cluster forwards the data packets into the base station. The clustering decreases a vast number of nodes and delivers the data to transmit directly by balancing energy. The SSTBC conserves energy by stopping radio and unnecessary nodes; thus, replica data would not be passed to the target. Compared with the wireless routing protocols such as LEACH and PEGASIS, the SSTBC has a high cluster lifetime, and more messages are passed to the base station. Therefore, the network life is improved in the SSTBC protocol.

De Paulo et al. [27] proposed the community detection-based routing protocol for wireless sensor networks for automatic development of clusters and delivering them to sink. This protocol works by sensor nodes and subdivided into two, namely the set-up phase (S-Phase) and the communication phase (C-Phase). The functioning of the protocol starts in the S-Phase. The WSN communities are upholding vertex label propagation. The C-phase starts after the completion of the S-phase, where the cluster heads get rotated every time. Termination of rounds involves two steps, cluster head establishment and data transmission. The CHE is the crucial step in which it drifts in each round and determines every community. The data from the communities transfer to the closest cluster heads in the data transmission. Therefore, entire data from cluster heads move into the sink. As a result, in comparison with LEACH, the designed RLP protocol has a high lifetime and network connectivity.

Lee [28] designed cross-layer routing protocol called TRIX-MAC. The functioning of TRIX-MAC is to keep the receiver wake-up without sleep and not to make the receiver for a long time. To maintain a high network lifetime and minimize energy consumption in WSN, duty cycling is initiated. The wake-up and sleep state are two states in every sensor node of duty cycling. The duty of the wake-up is to assemble

the data from nodes and dispatch them, whereas the sleep state will be inactive and saves energy. They exist two-state periods in wake-up for one duty cycle, which are scheduled and synchronous wake-up. The functioning of Sched-wakeup periods is to check whether the minute errors rupture the data, and time is scheduled occasionally and stops it. Concurrently, the Synch-wakeup sends the data frames from the node and synchronizes them to the receiver. As the protocol is multipath routing, it increases the data reliability and minimizes the traffic rate from source to destination. Additionally, a new metric is added to the protocol called Estimated Duty-Cycled Wait (EDW), which delivers the data by decreasing the waiting time, helps in saving energy, maintains low-latency, and strengthens the network throughput.

Messaoudi et al. [29] proposed a cluster-based routing protocol known as low energy adaptive clustering hierarchy protocol (LEACH), which saves the energy of nodes by keeping them alive until the data is forwarded to the required target and extends its network lifetime. An algorithm is implemented in this protocol called fuzzy logic module, employed in clustering process and selects the required cluster head. This logic produces a standard logic that lies between 0 and 1. The LEACH protocol functions by forming rounds and creates clusters for every round. The cluster head is obtained if the value of the node is beneath the threshold. The needed cluster heads will be taken according to the probability given by the fuzzy logic. Hence, this routing protocol gives a better network lifetime by transferring the aggregate data to the base station effectively.

3 Methodology

The proposed protocol uses opportunistic spectrum distribution (OSD) communication technique on the wireless sensor network. With the OSD approach, all the nodes in a neighborhood share the integrated spectrum allocated for communication. This work introduces an easy and powerful technique that concedes opportunistic channel access in wireless sensor networks in an entirely distributed form. The CSMA or TDMA is chosen based on the data priority rate. Our proposed method can achieve maximum spectrum utilization and throughput. It also lessens the interference within base stations and the licensed users. The OSD technology responds quickly and efficiently to the network parameter variations and achieves a high degree of fairness between spectrum sharing.

The system architecture used in this work is depicted in Fig. 1. The general control structure allows the nodes to login and dynamically accommodate their physical link layer, MAC, and network-level parameters. The fundamental segments to advance the control plane are reset, cross-layer protocol, discovery, and naming/addressing. When the spectrum sharing nodes boots-up for the first time, the MAC protocol action allows the detection of local links and constructs the PHY/MAC parameters. The nodes start building a discovery protocol according to the periodic reporting of local neighboring nodes link states employing a broadcast controlled hop mechanism. The discovery protocol also communicates with a cross-layer routing module that

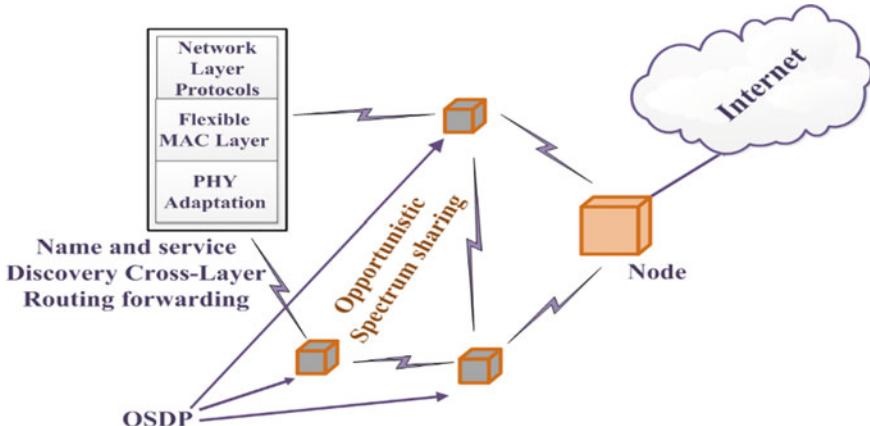


Fig. 1 System architecture

prepares end-to-end reachability and path data across parallel transaction that gives dynamically by cross-layer parameters entailing frequency, energy, low cost, and high throughput, etc. The critical basic is the support for shared call and search. To frequently examine the address of the network, the nodes permanently map their names varying with its arrangement and docility. We have extended the concept of OSD to assist as the control plane by promoting an economical control radio for spectrum sharing performing at the edge of the allotted spectrum band.

3.1 Promote Data

An OSD uses further nodes to spread packets, which means that the concession to forward a packet is made by the collecting node, rather than the broadcast node. Hence, when a node collects a packet, it stores it in a queue until its next moment to transmit. The agreement to forward a message is made based on the acknowledged packets angle metric. In this course, there are three possible behaviors to be considered based on the received packets location information.

- If the neighborhood knowledge points out that the packet is received from a node that is far away from the destination, then it should be delivered.
- If the packet is received from a node that is at an equal distance away from the destination, then it should be delivered.
- If the packet is received from a node that is very nearer to the destination, then it should be discarded.

The data may be corrupted when it is transmitted over a channel. So, there should be a mechanism to correct the errors while decoding. The inverse error detection (IED) is the approach used for correcting the errors. Fuzzy-based WLAN mechanism

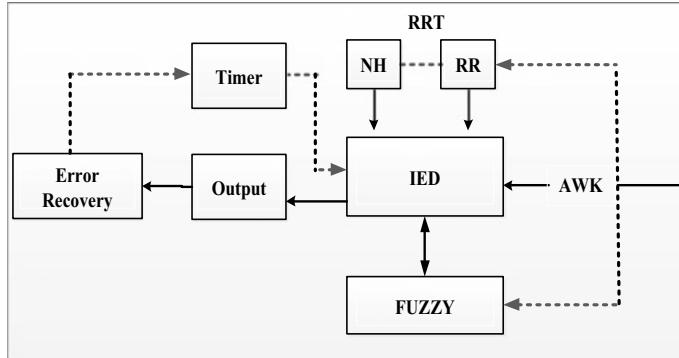


Fig. 2 Fuzzy-based WLAN mechanism

is depicted in Fig. 2. IED coding introduces a known code word onto the headway of data before the communication to observe any errors that may have developed during communication. This method of disclosure and improvement has become more and more important for data communication via wireless channel. Because of the nature of wireless communication, the typical “retransmit” function of most detection arrangements is incomplete, hence the need for alteration as well.

New standards are developed that correct this problem by pointing out the cause of packet loss. Network-oriented explanation uses the exact message from the network to detect interference. The fuzzy logic assumption is used to represent the network interference problem by using the RTT, time-to-live, and RR (rate at which RTT increases) as encouragement variables. In this arrangement, detection of interference takes priority over bit error detection. This is because, at the time of overpopulation, the TCP sender must slow the communication regardless of whether there are concurrent bit errors.

3.2 Fuzzy-Based WLAN Mechanism

The mechanism given in this fuzzy logic contains a sequence of steps after the localization of the network nodes and they are as follows,

Algorithm: Opportunistic Spectrum Distribution Protocol

```

Input – Queue, Source, High Priority Packets (HPP), Low Priority Packets (LPP)
Output – packet _to _send
// After the packets has been sent, no packets to be send
Packet _to _send ← null;
if the Queue is not empty then
    for each packet in Queue do
        if Source (packet = high data rate) then TDMA Activation
            packet _to _send = HPP; //High priority packets are sent
            end
        if Source(packet = low data rate) then CSMA Activation
        //After the delivery of high priority packets, low priority packets are forwarded
            packet _to _send = LPP;
            end
    end
return packet _to _send;

```

Step 1: The nodes of the network start their work by figuring the first hop neighboring node from its deployment. This is possible by transmitting messages occasionally. The message exchange is done every 15 s. The memory is then swapped between intermediate neighbor nodes, leading to having the nodes with two hop neighbors list.

Step 2: Through the CSMA mechanism, the nodes fetch the data to sink by detecting the relevant parameter. This proceeds till there is no maximization of traffic load.

Step 3: When the node achieves the condition to transfer the higher priority packets, the transmission of data from the neighboring nodes terminates. The nodes then relocate to TDMA imparting the primary slot in the predominant region.

Step 4: If two network nodes have remained in the corresponding predominant region, then slots are allocated successively to the network nodes. The network reciprocates to CDMA, after revamping.

Step 5: The data packets access the address of the nodes, and this information is carried along with other packets, so the data moves faster.

4 Results and Discussion

In this section, the performance of the proposed opportunistic spectrum distribution protocol is evaluated and contrasted with IEEE 802.15.4 and SMAC. The evaluation is computed under the criterion of average end-to-end delay, energy consumption, and throughput. The proposed protocol is simulated in NS-2, and the simulation parameters are given in Table 1. Network animator of the developed VANET simulation is depicted in Fig. 3. The completion of data transmission between source and destination is depicted in Fig. 4.

The simulation of the VANET model for the architecture shown in figure is performed in NS2. The simulation results show the performance of the OSDP, IEEE 802.15.4, and SMAC. Figure 5 delineates the comparative performance of the OSDP,

Table 1 Simulation parameters

Parameters	Value
Number of nodes	50
Coverage area	1000 × 1000 m ²
Simulation period	150 s
Traffic type	CBR
Agent type	UDP
MAC standard	IEEE 802.15.4
Frequency	914 MHz
Transmit power	0.282 J
Transmit antenna gain	1.0 J
Receive antenna gain	1.0 J
High priority data rate	8 Mbps
Low priority data rate	2 Mbps

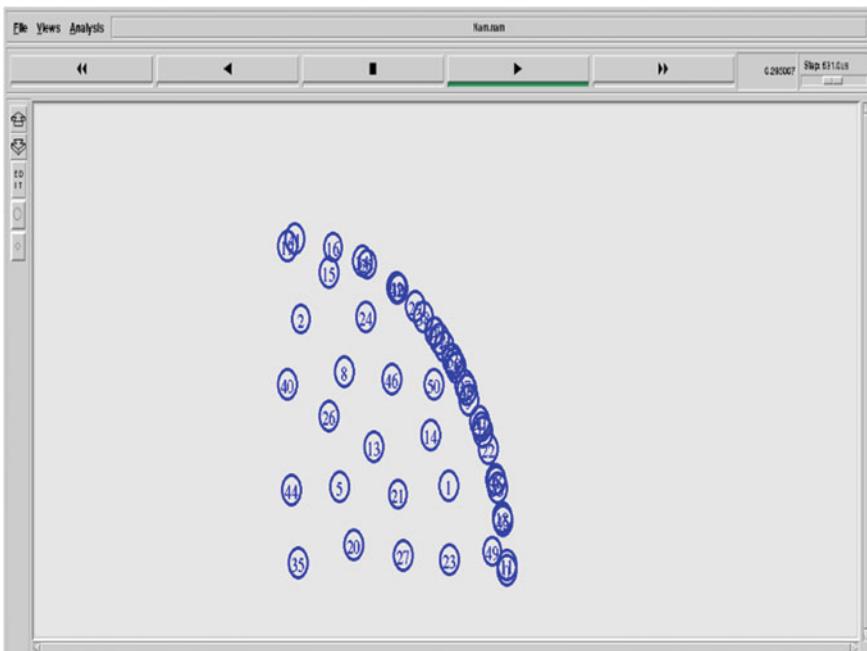


Fig. 3 Network animator of the developed VANET

IEEE 802.15.4, and SMAC in accordance with energy consumption of the nodes under distant traffic arrival rates. It is evident that the proposed MAC consumes very less energy with increasing number of nodes compared to other MAC protocols. OSDP utilizes nearly 57% and 43% less energy than IEEE 802.15.4 and SMAC.

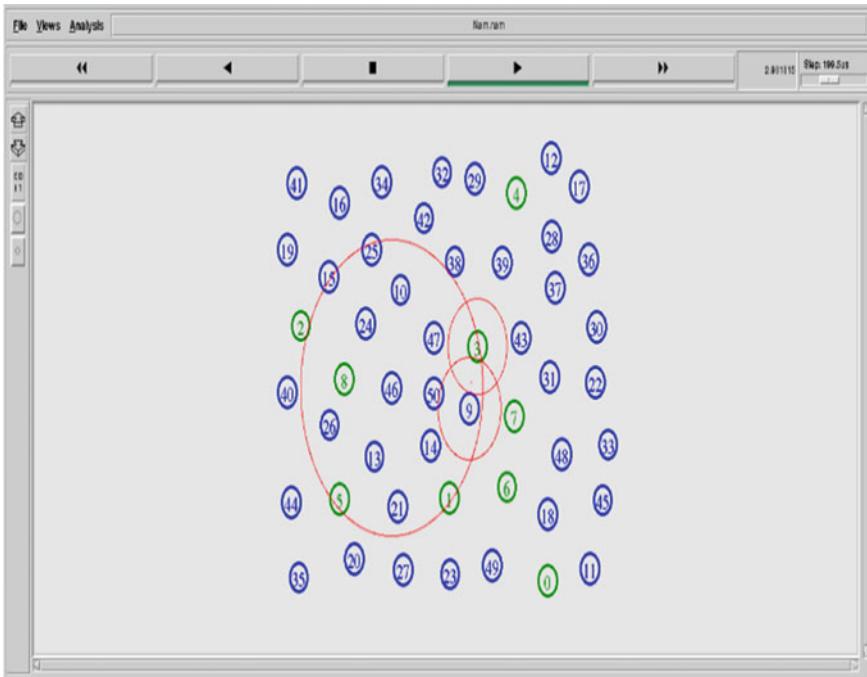


Fig. 4 Data transmission between source and destination

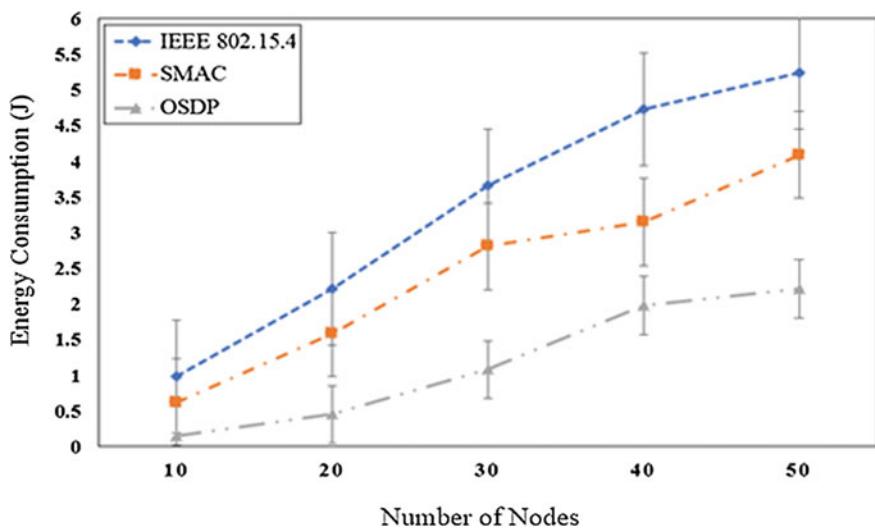


Fig. 5 Energy consumption vs number of nodes

Figure 6 depicts the comparative behavior of the OSDP, IEEE 802.15.4, and SMAC according to average end-to-end delay for several numbers of nodes. In comparison with our designed scheme, the MAC protocols has more end-to-end delay; thus, OSDP consumes energy. The average end-to-end delay of OSDP is approximately equal to 63% inferior than IEEE 802.15.4 and similarly 35% inferior than SMAC.

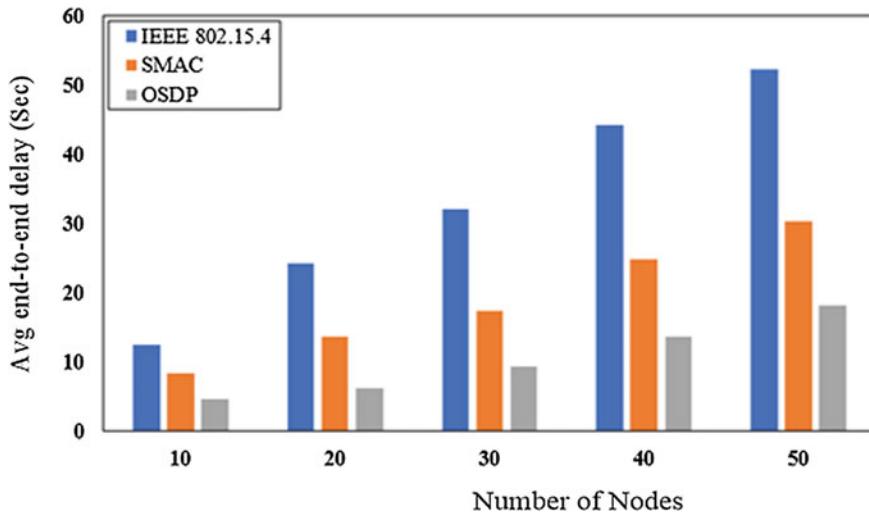


Fig. 6 Avg end-to-end delay vs number of nodes

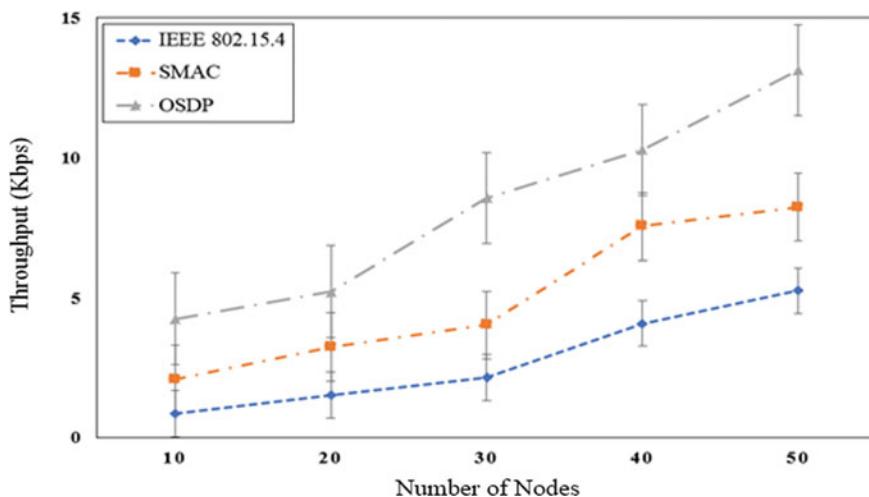


Fig. 7 Impact of nodes on throughput

Figure 7 shows the throughput performance of the OSDP, IEEE 802.15.4, and SMAC. In comparison with IEEE 802.15.4 and SMAC, OSDP attains a better throughput. The enhancement achieved here is approximately 68% over IEEE 802.15.4 and 47% over SMAC.

5 Conclusion

In this paper, we depicted a unique protocol named opportunistic spectrum distribution protocol (OSDP). The main intention of this protocol is to provide parallel communication. The proposed protocol performs spectrum distribution on considering the transmission history, fuzziness present in the channel due to different paths, path changes, and priority of data transmission. According to the type of data rate, the CSMA or TDMA is selected. The proposed OSDP uses both TDMA and CSMA together and achieves good spectrum utilization, more throughput, and less delay. Based on the proposed framework, we designed an effective algorithm in tackling and analyzing the network utility. The simulation results have shown that the proposed OSDP outperforms the IEEE 802.15.4 and SMAC in accordance of end-to-end delay, energy consumption, and throughput. This protocol eliminates all the delay issues by consuming the energy and maximizes the throughput of the network. It is concluded that the proposed OSDP has superior performances, by utilizing nearly 57% and 43% less energy than IEEE 802.15.4 and SMAC. The average end-to-end delay is 63% and 35% lesser than IEEE 802.15.4 and SMAC from our proposed protocol. Hence, its performance is better than other three protocols, and it is convenient for real-time, secure, and steadfast communication.

References

1. Pei Y, Liang YC (2015) Cooperative spectrum sharing with bidirectional secondary transmissions. *IEEE Trans Veh Technol* 64:108–117
2. Park J, Pawelczak P, Cabric D (2011) Performance of joint spectrum sensing and MAC algorithms for multichannel opportunistic spectrum access Ad Hoc networks. *IEEE Trans Mobile Comput* 10:1011–1027
3. Cheng N, Zhang N, Lu N, Shen X, Mark JW, Liu F (2014) Opportunistic spectrum access for cr-vanets: a game-theoretic approach. *IEEE Trans Veh Technol* 63:237–251
4. Kapileswar N, Vijayasanthi P, Palepu MRD, Chenchela VKR (2016) Improving the lifespan of wireless sensor network via efficient carrier sensing scheme-CSMA/SDF. *Int J Eng Sci Res Technol* 5:723–732
5. Ammar I, Awan I, Min G (2010) An improved S-MAC protocol based on parallel transmission for wireless sensor networks. In: Proceedings of IEEE international conference on network-based information systems, pp 48–54
6. Nellore K, Hancke GP (1892) Traffic management for emergency vehicle priority based on visual sensing. *Sensors* 16(11):1–22
7. Plummer A, Taghizadeh M, Biswas S (2009) Measurement based capacity scavenging via whitespace modeling in wireless networks. In: Proceedings of IEEE GLOBECOM

8. Huang S, Liu X, Ding Z (2009) Optimal sensing-transmission structure for dynamic spectrum access. In: Proceedings of IEEE INFO-COM
9. Sharma M, Sahoo A (2010) Opportunistic channel access scheme for cognitive radio system based on residual white space distribution. In: Proceedings of IEEE 21st international symposium on personal, indore and mobile radio communications
10. Sharma M, Sahoo A (2013) Residual white space distribution based opportunistic multichannel access protocol for dynamic spectrum access networks. In: Proceedings of IEEE international conference on communication systems and networks, Bangalore, India
11. Xu D, Liu X (2008) Opportunistic spectrum access in cognitive radio networks: when to turn off the spectrum sensors. In: Proceedings of fourth annual international conference wireless internet (WICON)
12. Haykin S (2005) Cognitive radio: brain-empowered wireless communications. *IEEE J Sel Areas Commun* 23:201–220
13. Thomas RW, DaSilva LA, MacKenzie AB (2005) Cognitive networks. In: Proceedings of IEEE international symposium on new frontiers in dynamic spectrum access networks, pp. 352–360
14. Kim KJ, Kwak KS, Choi BD (2013) Performance analysis of opportunistic spectrum access protocol for multi-channel cognitive radio networks. *J Commun Netw* 15:77–86
15. Ren K, Wang Q (2013) Opportunistic spectrum access: from stochastic channels to non-stochastic channels. *IEEE Wirel Commun* 20:128–135
16. Babaei M, Basar E, Aygolu U (2016) A cooperative spectrum sharing protocol using STBC-SM at secondary user. In: Proceedings of IEEE 24th international conference on telecommunications forum, Belgrade, Serbia, 22–23 Nov 2016
17. Wu X, Han X, Labeau F (2016) Cooperative spectrum sharing protocol based on transform domain processing with joint secondary selection and power allocation. In: Proceedings of IEEE international conference on wireless communications and mobile computing conference, pp 475–480
18. Valehi A, Razi A (2017) Maximizing energy efficiency of cognitive wireless sensor networks with constrained age of information. *IEEE Trans Cogn Commun Netw* 3(4):643–654
19. Lin F, Hen C, Zhang N, Guan X, Shen X (2016) Autonomous channel switching: towards efficient spectrum sharing for industrial wireless sensor networks. *IEEE Internet Things J* 3(2):231–243
20. Jayaprakasham S, Abdul Rahim SK, Leow CY (2017) Distributed and collaborative beam forming in wireless sensor networks: classifications, trends, and research directions. *IEEE Commun Surv Tutor* 19(4):2092–2116
21. Jain N, Bohara VA (2015) Energy harvesting and spectrum sharing protocol for wireless sensor networks. *IEEE Wirel Commun Lett* 4(6):697–700
22. Amsalu SB, Zegeye WK, Hailemariam D, Astatke Y, Moazzami F (2016) Energy efficient Grid Clustering Hierarchy (GCH) routing protocol for wireless sensor networks. In: IEEE 7th annual ubiquitous computing, electronics & mobile communication conference, 12 Dec 2016
23. Jadhav P, Satao R (2016) A survey on opportunistic routing protocols for wireless sensor networks. In: 7th international conference on communication, computing and virtualization, vol 79, pp 603–609
24. Sher A, Javaid N, Ahmed G, Islam S, Qasim U, Ali Khan Z (2016) MC: Maximum coverage routing protocol for underwater wireless sensor networks. In: 19th international conference on network-based information systems, pp 91–98, 19 Dec 2016
25. Masjid A, Azam I, Waheed A, Zain-ul-Abidin M, Hafeez T, Ali Khan Z, Qasim U, Javaid N (2016) An energy efficient and balanced energy consumption cluster based routing protocol for underwater wireless sensor networks. In: 30th international conference on advanced information networking and applications, pp 324–333, 23–25 Mar 2016
26. Duy Tan N, Dinh N (2015) SSTBC: sleep scheduled and tree-based clustering routing protocol for energy-efficient in wireless sensor networks. In: International conference on computing & communication technologies—research, innovation, and vision for future (RIVF), pp 180–185, 25–28 Jan 2015

27. De Paulo MA, Nascimento MCV, Rosset V (2016) RLP: a community detection-based routing protocol for wireless sensor networks. In: 13th international symposium on network computing and applications, pp 237–244, 16 Oct 2016
28. Lee H (2020) Design of a cross-layer multi-path routing protocol for duty-cycled wireless sensor networks. In: International conference on computational science and computational intelligence (CSCI), pp 1092–1096
29. Messaoudi A, Elkamel R, Helali A, Bouallegue R (2016) Distributed fuzzy logic based routing protocol for wireless sensor networks. In: 24th international conference on software, telecommunications and computer networks, 8 Dec 2016

Harmonic Reduction in a Three-Phase Voltage Source Inverter Using RLC Filter and FFT



Vivek Pratap Singh, M. A. Ansari, and Nivedita Singh

Abstract With the increasing concern for the environment, attention is shifting toward solar energy in the power industry. But the power obtained from solar panels is in DC form; thus, we need to use an inverter to convert this power from DC to AC. During power conversion harmonics are introduced in the system which affects the quality of power. Thus, it is necessary to reduce the harmonics present in the system for better power quality. This paper describes the implementation of a second-order RLC low-pass filter to a three-phase voltage source inverter with the 180° mode of conduction. Here, two systems have been simulated in the MATLAB/Simulink for harmonics analysis in fast Fourier transform (FFT). The level of harmonics present in the system is determined with the help of total harmonic distortion (THD). Results determine that the distortion in the output waveform is reduced with the implementation of a second-order RLC low-pass filter in the three-phase voltage source inverter.

Keywords Fast Fourier transform · RLC filter · Harmonics · Inverter · Total harmonic distortion

1 Introduction

Nowadays, renewable energy is getting more attention because of the increasing concern for the environment and to meet the increasing electricity demand. Photovoltaics plays an important role in renewable energy because it is considered a clean and ecological source of energy. The power obtained from solar photovoltaics is in DC form; thus, a converter device is needed. A three-phase voltage source inverter can be used to convert the power from DC to AC. While converting the power from one form to another, harmonics are introduced in the system which leads to losses and also affects the power quality [1]. Thus, it is necessary to reduce the harmonics present in the system in order to obtain a better power quality at the output.

V. P. Singh · M. A. Ansari (✉) · N. Singh (✉)

Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India
e-mail: ma.ansari@gbu.ac.in

1.1 Inverter

Three-phase inverters have applications in variable frequency drives and HVDC transmission when we need to transfer high power [2]. Inverters can be broadly classified as:

Voltage source inverter (VSI): These types of inverters have stiff-type DC source voltage, and the inverter terminal of a VSI has zero or limited impedance.

Current source inverter (CSI): These types of inverters are supplied with variable current from the DC source having a high impedance.

1.2 Harmonics

Total harmonic distortion is the ratio of RMS value of total harmonics of the waveform to the RMS value of the fundamental wave that is the ratio of all harmonic components in the numerator and the fundamental component in the denominator. Total harmonic distortion is a unitless quantity. For a perfect sine wave, the value of total harmonic distortion should be equal to zero. So, we can say that total harmonic distortion determines the level of distortion present in the waveform.

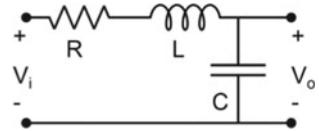
$$\text{THD} = \frac{\sqrt{\sum_{n=2}^{\infty} V_{n\text{-rms}}^2}}{V_{\text{fund_rms}}} \quad (1)$$

The presence of harmonics causes heating problems in motors, and the value of torque is also reduced because of the presence of harmonics in the system. Harmonics also increase the voltage stress and heating problems for the generator [3]. Harmonics present in the system also cause mal-operation of electronics switchgear and relaying. So, we can say that the presence of harmonics will ultimately lead to a reduction in the equipment life; thus, it is necessary to consider these harmonics while designing the equipment that is their rating should be proper in order to limit the harmonics in the power system [4].

1.3 RLC Filter

RLC filter is designed for the three-phase VSI, which contains two reactive components that are inductor and capacitor. In the RLC circuit, capacitor and inductor are elements with storing capability, where the inductor initially opposes the change of current, while the capacitor initially opposes the change of voltage. The inductor successfully blocks high frequencies and passes low frequencies, while the capacitor passes high frequencies but blocks low frequencies.

Second-order RLC
low-pass filter



After taking the Laplace transform of the R, L, and C in Fig. 1, the transfer function of the filter circuit is given by equation 2.

Comparing the above equation with the standard equation of the second-order system, we get

$$T(s) = \frac{V_o(s)}{V_i(s)} = \frac{1/Cs}{R + Ls + (\frac{1}{Cs})} \quad (2)$$

$$T(s) = \frac{1}{LCs^2 + RCs + 1} \quad (3)$$

$$T(s) = \frac{1/LC}{s^2 + \frac{R}{L}s + \frac{1}{LC}} \quad (4)$$

$$T(s) = \frac{w_n^2}{s^2 + 2\varepsilon w_n s + w_n^2} \quad (5)$$

From the above equation, we will get the following values

$$w_n^2 = \frac{1}{LC} \quad (6)$$

Thus, we will get $w_n = \frac{1}{\sqrt{LC}}$

$$2\varepsilon w_n = \frac{R}{L} \quad (7)$$

Put the value of w_n thus we will get

$$\varepsilon = \frac{R}{2} \sqrt{\frac{C}{L}} \quad (8)$$

where ε = damping ratio of the system and w_n = frequency of oscillation.

2 System Description and Modeling

In this paper, a three-phase voltage source inverter for the 180° mode of conduction is modeled in MATLAB. Six IGBT two in each branch has been used as a switching device in the inverter. The pulse generator is used to provide a pulse to the switching devices. IGBT is utilized in the system since it has low power capacity and high switching speed.

2.1 Three-Phase Voltage Source Inverter with Resistive Load

A DC source voltage source is considered here with a value of 220 V. The type of load considered in the system is resistive, and the value of the resistance is 20 ohms. The internal resistance for all the six IGBT is taken as 0.001Ω . The value of the amplitude taken in the pulse generator is 10, while the value of the pulse width is taken as 0.02. The Powergui block is taken as continuous in the system. Three-phase voltage waveforms have been obtained for the resistive-type load (Figs. 2 and 3 and Table 1).

2.2 Three-Phase VSI with RLC Filter

A second-order RLC low-pass filter is designed for a three-phase voltage source inverter with the 180° mode of conduction and with a frequency of oscillation of 50 Hz.

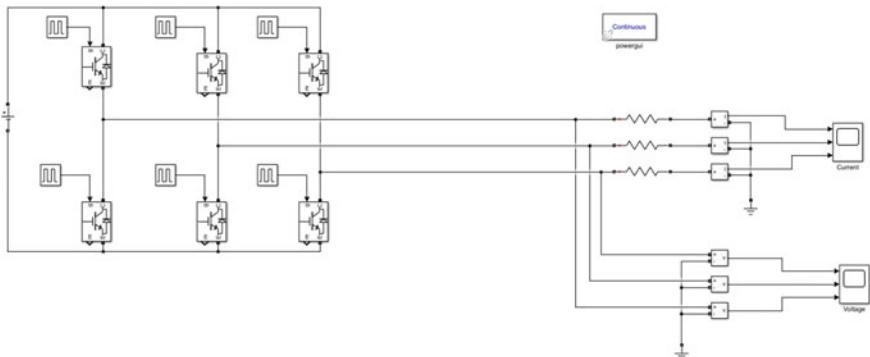


Fig. 2 Simulink model of three-phase VSI without filter

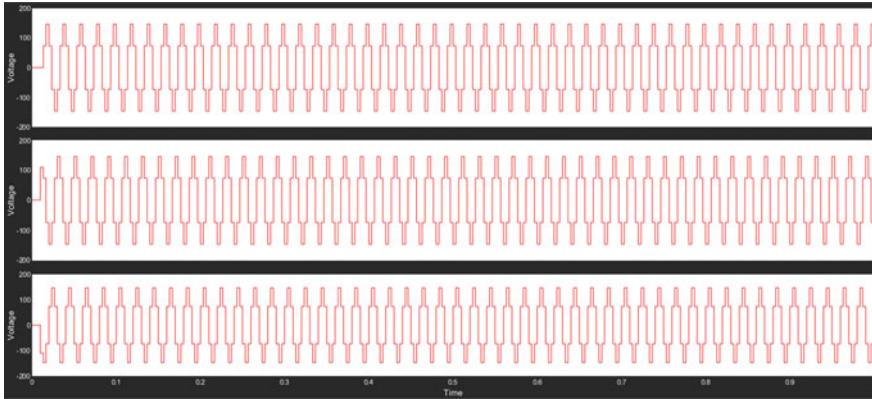


Fig. 3 Three-phase output current waveform of system without filter

Table 1 Parameters of three-phase voltage source inverter with resistive load

S. No.	Parameter name	Selected values
1	Source voltage	220 V
2	Load resistance	20Ω
3	IGBT internal resistance	0.001Ω
4	Powergui	Continuous
5	Pulse generator	Amplitude = 10 Pulse Width = 0.02

Filter designing: Cut off frequency (f_c) is given by, $f_c = \frac{1}{2\pi\sqrt{LC}}$. f_c is taken as 50 Hz; in this case, we know that $\omega_o = 2\pi f_c$

$$\omega_o = 2 \times \pi \times 50 = 314 \text{ rad/sec.}$$

Taking, $Q = 3.2$ and $C = 100 \mu\text{F}$

$$Q = \frac{\omega L}{R} \text{ also } Q = \frac{1}{\omega_0 CR}$$

Thus, the calculated values of resistance, inductance, and capacitance are 10Ω , 100 mH , and $100 \mu\text{F}$, respectively. And the values of other parameters will remain the same that is source voltage is 220 V, the load resistance is 20Ω , and IGBT internal resistance is 0.001Ω (Figs. 4 and 5 and Table 2).

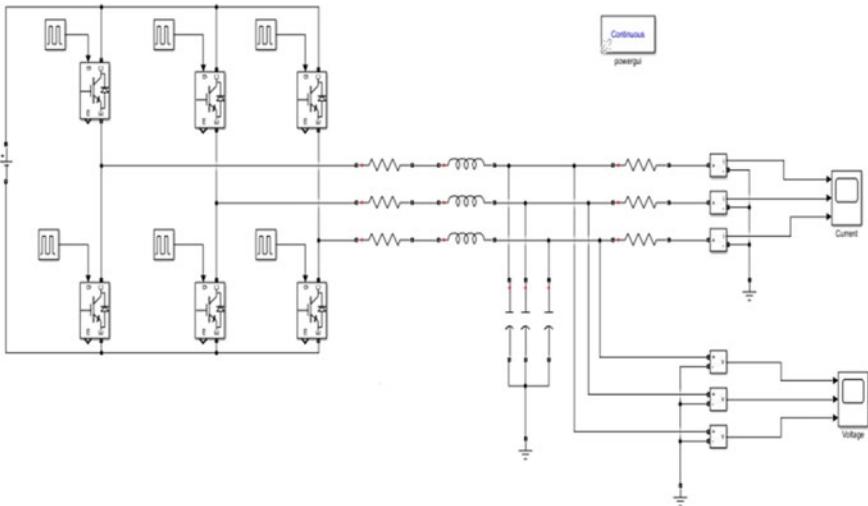


Fig. 4 Simulink model of three-phase VSI with filter

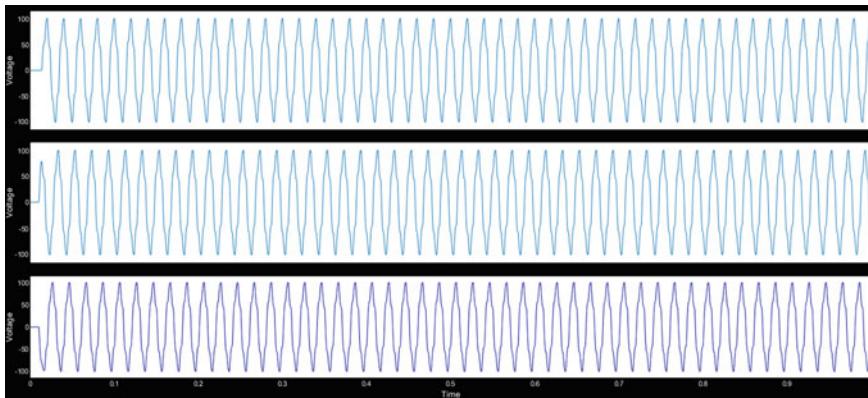


Fig. 5 Three-phase output current waveform of system with filter

Table 2 Comparison parameters of three-phase voltage source inverter with resistive load and RLC filter

S. No.	Parameter name	Selected values
1	Source voltage	220 V
2	Load resistance	20Ω
3	IGBT internal resistance	0.001Ω
4	Powergui	Continuous
5	Pulse generator	Amplitude = 10 Pulse Width = 0.02
	RLC Filter	Resistance = 10Ω Inductance = 100 mH Capacitance = $100 \mu\text{F}$

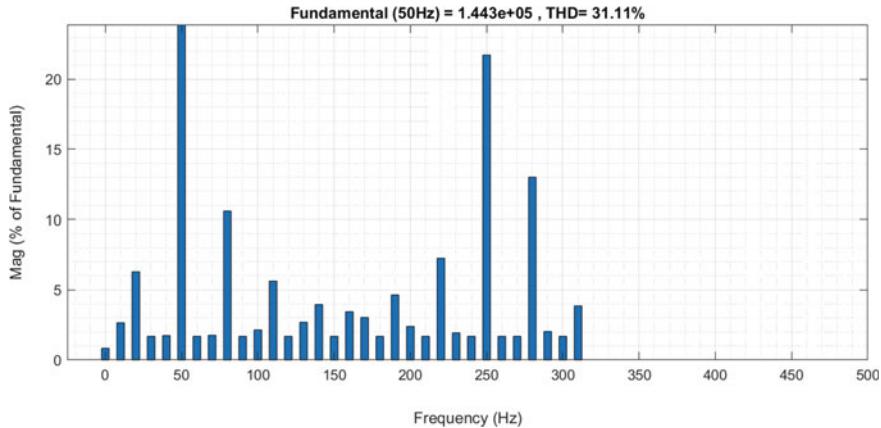


Fig. 6 FFT analysis of inverter without filter

3 Results and Discussion

Fast Fourier transform analysis of the three-phase output voltage waveform has been done in order to determine the value of total harmonic distortion present in the system.

3.1 *Three-Phase VSI with R-Load*

Figure 6 shows that the value of total harmonic distortion (THD) is 31.11% for the output three-phase voltage waveform.

3.2 *Three-Phase VSI with RLC Filter and R-Load*

The value of total harmonic distortion (THD) is 12.41% for the output three-phase voltage waveform of the inverter with RLC filter at the output side (Fig. 7).

4 Conclusion

A simulation model of three-phase voltage source inverter with resistive load and 180° mode of conduction has been developed in MATLAB/Simulink for a duration of 10 s. From the comparison table, it is seen that the value of total harmonic distortion is 31.11% for the output voltage waveform from three-phase VSI without any filter. While with the implementation of a second-order RLC low-pass filter, the value of

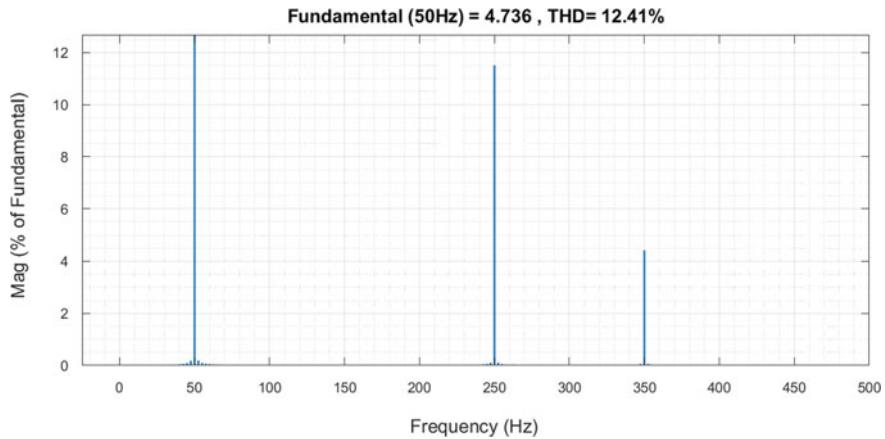


Fig. 7 FFT analysis of inverter with filter

Table 3 Comparison of total harmonic distortion

S. No.	Case studied	Total harmonic distortion (THD in %)
1	Three-phase voltage source inverter with R-load	31.11
2	Three-phase voltage source inverter with second-order RLC filter	12.41

total harmonic distortion obtained is 12.41% for the output voltage waveform. Thus, we can say that with the implementation of a second-order RLC low-pass filter to a basic circuit of three-phase voltage source inverter, the level of harmonics can be reduced which implies a better power quality (Table 3).

References

1. Mahajan V, Agarwal P, Gupta HO (2020) Power quality problems with renewable energy integration. In: Power quality in modern power systems. Academic Press, pp 105–131
2. Phukan R, Ohn S, Dong D, Burgos R, Mondal G, Nielebock S (2020) Evaluation of modular AC filter building blocks for full SiC based grid-tied three phase converters. In: 2020 IEEE energy conversion congress and exposition (ECCE). IEEE, pp 1835–1841
3. Ali AEMM, Mansy III, Ismael IM (2020) Improvement of power quality in electrical distribution system using shunt active power filter controlled by fuzzy logic controller (Dept E). MEJ. Mansoura Eng J 36(1):18–24
4. Onovakpuri O (2019) Investigation of power quality issues due to increased levels of distributed generation and possible solutions

Fault Detection and Classification Using Fuzzy Logic Controller and ANN



Rishabh Verma and M. A. Ansari

Abstract Power transmission is a major component in electrical engineering after power generation. Fault in transmission lines is common and an obvious problem to continue proper power supply and reliability. This paper illustrates a technique to detect the type of the different faults that occur on a transmission line for accurate operation using artificial intelligence technique, i.e. fuzzy logic-based control scheme. The simulation model for fault detection is developed in MATLAB using fuzzy logic controller using phase components of voltage and current as inputs and different types of fault for output as classification. The proposed fuzzy logic controller takes neutral and phase currents, i.e. current in the red phase (IA), current in the yellow phase (IB), current in blue phase (IC), and current in the neutral phase (IN), and similarly the phase voltages. Based on the simulation results, it has been found that the proposed fuzzy logic-based fault detection model detects and classifies both symmetrical and unsymmetrical shunt faults correctly voltage in the red phase (VA), the voltage in the yellow phase (VB), and voltage in the blue phase (VC) as inputs. These membership functions are used in forming the rule base for the fuzzy logic fault detection system.

Keywords Faults · Fault detection · Transmission lines · Fuzzy logic controller

1 Introduction

Generator units, transformers, transmission lines, isolation equipment, circuit breakers, connecting rods, cables, joints, voltage transformers, distribution equipment, and various sorts of loads are all part of the electrical system.

Artificial intelligence technologies that process information from alarms and protection rates in energy transmission and distribution systems are used in several research projects connected to debt detection. A dramatic decline in line resistance due to an increase in failure makes faults simple to spot. Your rating is the added

R. Verma (✉) · M. A. Ansari (✉)

Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India
e-mail: ma.ansari@gbu.ac.in

benefit. Different stages of the fault currents and voltages for fault conditions on the transmission lines are classified into degrees of membership functions such as low, normal, and high categories. The three phase (red, yellow, and blue) and neutral phase IN feeder currents and phase voltages are used as inputs to the fuzzy inference system in this fault detection approach (FIS). This is a type of failure that occurs. LLL and LLG mistakes are symmetrical and undetectable among these five faults. With failure of insulation, voltage and current levels vary abruptly, and phase balance worsens. The procedure necessitates continual monitoring of the line's resistance readings at each level. The phase with a significant decrease in resistance shows a problem with the path.

When a machine malfunctions, its characteristic properties (such as resistance) can vary to other values until the defect is fixed. There is a good chance that the electrical system will collapse, such as due to insulations, wind, falling trees on power lines, hardware failures, and so on. Electrical equipment in the electrical system functions at normal voltage and current speed under normal or safe operating conditions (Fig. 1).

The three phase (red, yellow, and blue) and neutral phase IN feeder currents and phase voltages are used as inputs to the fuzzy inference system in this fault detection approach (FIS). Different levels of fault currents and voltages for various fault conditions on transmission lines are divided into low, normal, and high degrees of membership functions. The fuzzy logic defect detection system's rule base is built using these membership functions.

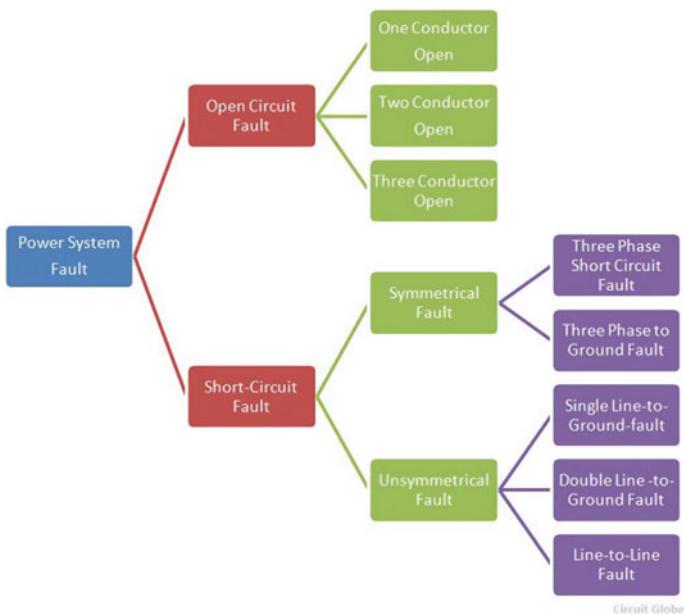


Fig. 1 Basic classification of power system fault [1]

2 Previous Related Works

Many kinds of research on fault diagnosis include artificial intelligence methods that process warning and protection rate information in energy transmission systems [1–4]. Special systems have been implemented in collaboration with SCADA to develop more efficient and accurate centralized fault diagnosis systems in transmission systems. The approach records information such as the location of the failure, the causes of the failure and identifies the unwanted use of protective equipment. Voltage and current sensors are installed on transmission lines for real-time implementation, and this leads to high costs. Artificial neural network (ANN)-based application written in MATLAB environment to diagnose electrical system failure has been developed to detect disturbances and faults [5]. The fault position and fault level of the line and bus section have been obtained using information from the warning. This technology provides effective information to decision makers, but most transmission systems are not fully alerted. A combination of ANN and hairy logic has been used to process alarm and protection rate information for the purpose of identifying defective components and line components. ANN approach with waves is developed to diagnose and fault failures. Reference [6] uses oscillation data from automotive devices and therefore requires a communication network between a remote power system and digital recorders. Systems for detecting substations have been developed using the Petri network theory. In this method, information from circuit breakers and defective protection devices is adjusted based on mathematical combinations to calculate the exact fault component. Two Petri network terms, namely the neural Petri network and the hairy neural Petri network are used to locate failures on the lines or sections. Though, these methods are not appropriate for fault analysis and detection in transmission systems due to the lack of notifying information about failure or fault occurrence. The proposed method uses voltages and currents as input data to diagnose the fault presented an electric protection allocation framework to analyse and classify the type of faults in the electricity system. Reference [17] work uses a reading of the phase value of current using only the first (1/4) part of the cycle in an integrated method that combines symmetrical component technology with the principal component diagnosis (PCA) to declare, identify and classify failure. Fault analysis of multiphase power transfer systems with symmetrical components was proposed. After taking account of these previous works, we concluded that the fuzzy logic-based scheme is an easier, economical, and reliable method for fault detection.

3 Adopted Fuzzy Algorithm

1. Start the fuzzy logic designer
2. Select the Input Variables (i.e. 3 phase Fault currents and fault voltages) for input

3. Define input Membership Function (triangular membership function adopted for Fuzzification)
4. Set input range of the variables
5. Set input variable (different types of fault) for output i.e. symmetrical and unsymmetrical faults
6. Define output Membership Function
7. Set output range of faults
8. Define controller
9. Set rules for the controller
10. If rules are satisfied and then go to step 11
11. Else go to the step 9
12. Stop.

The use of fuzzy logic is to detect and classify all symmetrical and unsymmetrical defects. The fundamental component of each phase's voltage signals is employed as input to the fuzzy inference system for this purpose. The benefit of adopting fuzzy is that it is a simple and flexible solution to any situation. Fuzzy logic can also be applied to challenges involving partial or inaccurate data. The controller's design begins with the selection of the Mamdani in the fuzzy logic designer. The fault classifier controller will then be fed seven input variables: $I(A)$, $I(B)$, $I(C)$, $I(N)$, $V(A)$, $V(B)$, $V(C)$. Following that, three (3) members of the triangle type function are chosen, namely low, normal, and high. The parameters for the membership functions are:

Low— $[-0.5, 0, 0.5]$.

Normal— $[0, 0.5, 1]$.

High— $[0.5, 1, 1.5]$.

The membership function plots are 181 (Figs. 2 and 3).

4 Modelling and Simulation Work

The suggested fault detection Simulink model includes a 200 km long three-phase pi section transmission line with a fault occurring at 30 km along the line with a three-phase AC source. Additional information can be found in Table 1 (Fig. 4).

5 Results and Discussions

Below are the results obtained after simulating the proposed model under the given specifications/ratings. Each case of fault detection is discussed with their simulated waveforms and the rule base formulation (Figs. 5 and 6).

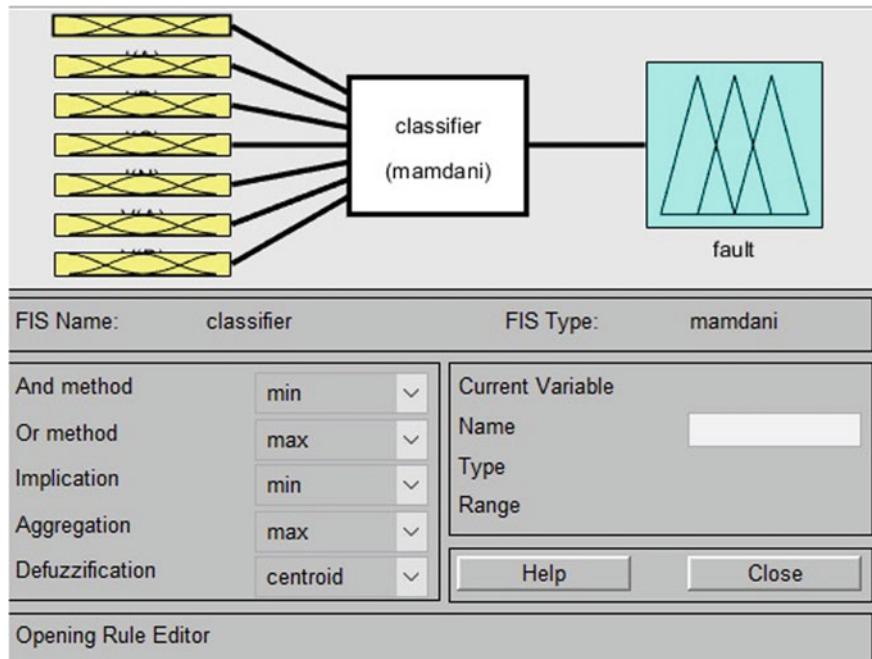
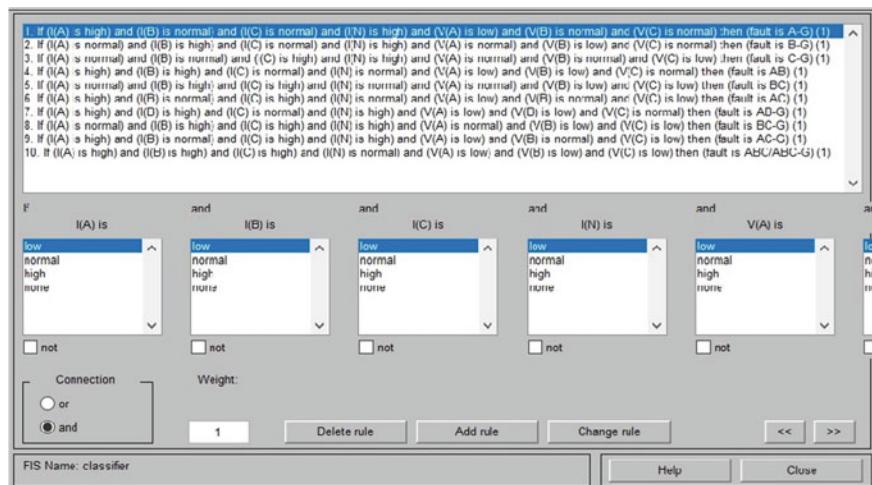
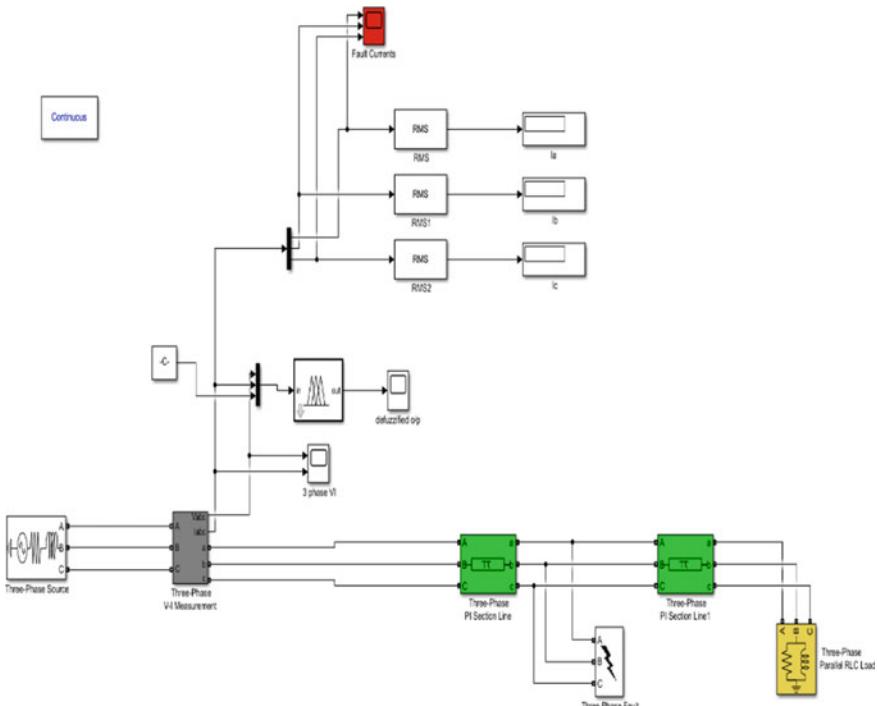
**Fig. 2** Fuzzy logic designer for fault designer**Fig. 3** Rule editor with fault classifier logic

Table 1 Simulink block parameters with ratings

Equipments	Specifications/ratings
1. Three-phase AC source	Phase-to-phase voltage (Vrms): 25e3 Frequency (Hz): 50 Short-circuit level at base voltage (MVA): 100e3
2. Three-phase parallel RLC load	Resistance: 20 ohms Active power P (W): 10e3 Inductive reactive Power QL (positive var): 100 Capacitive reactive power QC (negative var): 0
3. Three-phase fault	Switching times (s): 5/50 Fault resistance Ron (Ohm): 0.001; Ssubber capacitance: infinite; snubber resistance: 1e6 Time Interval (sec): 3e-1
4. Three-phase PI section	Frequency used for RLC specification (Hz): 50 Positive- and zero-sequence resistances (Ohms/km) [r1 r0]: [0.01273 0.3864] Positive- and zero-sequence inductances (H/km) [l1 l0]: [0.9337e-3 4.1264e-3] Length: 30 km, 170 km (can be varied acc. to our need)

**Fig. 4** Fault detection Simulink model

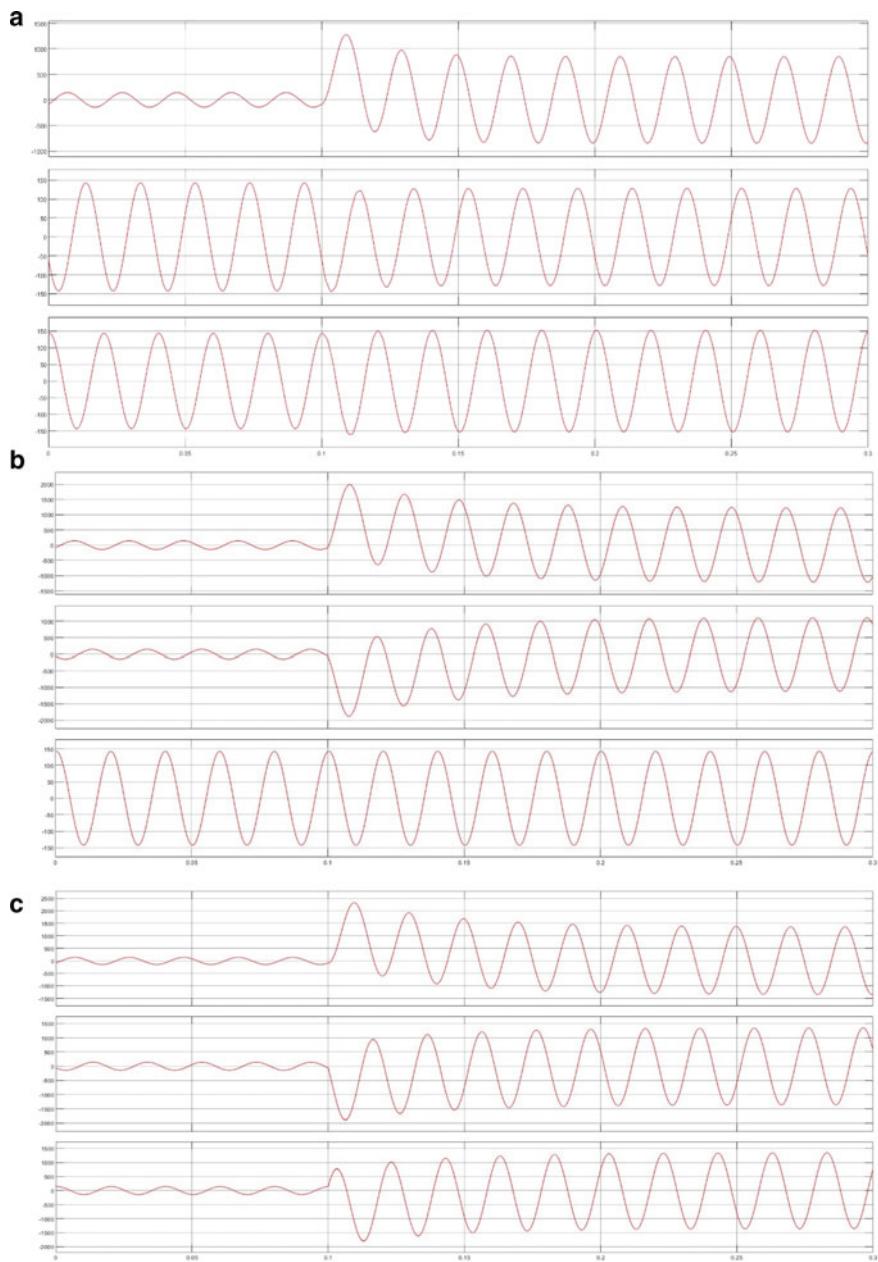


Fig. 5 **a** System output current waveform of single line to ground fault. **b** System output current waveform of double line to ground fault. **c** System output current waveform of three-phase fault.

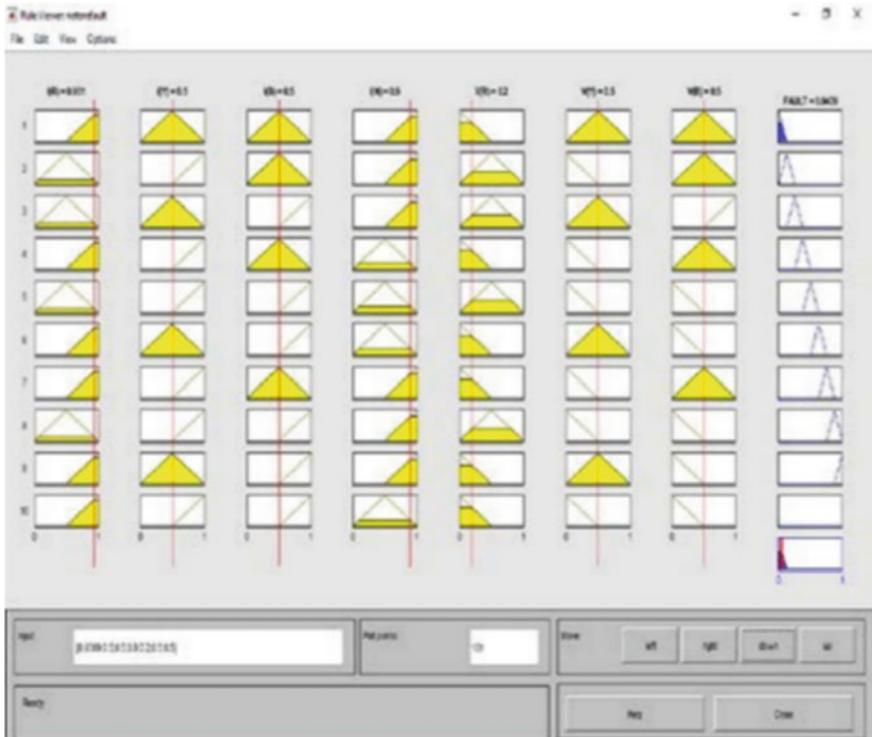


Fig. 6 Rule viewer with of single line to ground fault (S-L-G) detection

The first seven columns from the left are the inputs of the controller, while the last and the eighth column from the left is the output. The de-fuzzified output is displayed as a bold vertical line (red) on this plot. The rule for this fault condition is read as:

- If (I(A) is 0.931) and (I(B) is 0.5) and (I(C) is 0.5) and (I(N) is 0.9) and (V(A) is 0.2) and (V(B) is 0.5) and (V(C) is 0.5) then (FAULT AG/BG(CG is 0.0439)).
- If (I(A) is 0.5) and (I(B) is 0.9) and (I(C) is 0.9) and (I(N) is 0.9) and (V(A) is 0.5) and (V(B) is 0.2) and (V(C) is 0.2) then (FAULT, A-B-G/B-C-G/C-A-G is 0.616). The de-fuzzified crisp output is Fault = 0.616 and is displayed as a bold red vertical line on this plot.
- If (I(A) is 0.9) and (I(B) is 0.9) and (I(C) is 0.5) and (I(N) is 0.5) and (V(A) is 0.2) and (V(B) is 0.2) and (V(C) is 0.5) then (FAULT, AB/BC/CA is 0.375). The de-fuzzified crisp output is Fault = 0.616 and is displayed as a bold red vertical line on this plot (Figs. 7 and 8).
- If (I(A) is 0.9) and (I(B) is 0.9) and (I(C) is 0.9) and (I(N) is 0.5) and (V(A) is 0.2) and (V(B) is 0.2) and (V(C) is 0.2) then (FAULT 3phase, i.e. LLL is 0.5). The de-fuzzified crisp output is Fault = 0.5 and is displayed as a bold red vertical line on this plot.

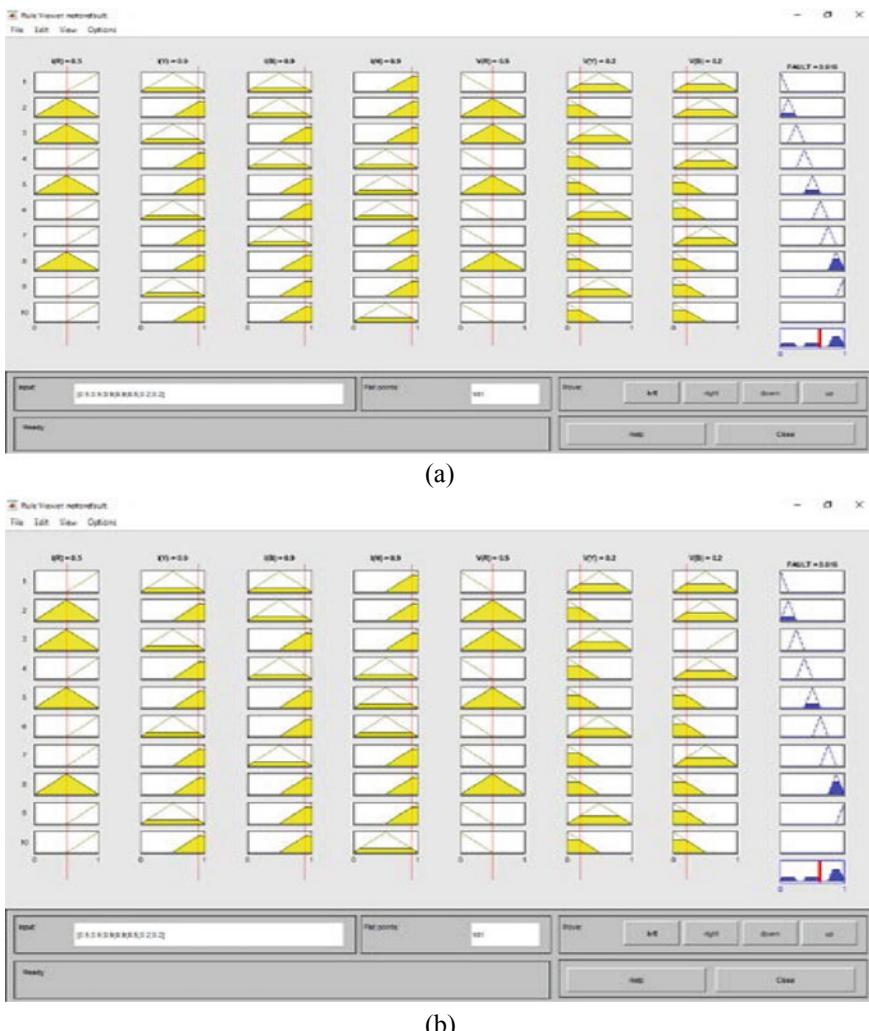


Fig. 7 **a** Rule viewer with double line to ground fault detection. **b** Rule viewer with line to line fault detection

6 Conclusion

The proposed approach is used to create a fuzzy logic controller for a fault detection system utilizing a fuzzy algorithm applied to a three-phase fault Simulink model. The fault classifier designed for the system improved the awareness regarding the different types of faults, i.e. both the symmetrical and unsymmetrical faults. For all three situations, the de-fuzzified output is 0.0439, 0.616, 0.375, and 0.5, respectively. These values are inside the controller's operating range of zero to one (0–1). The proposed

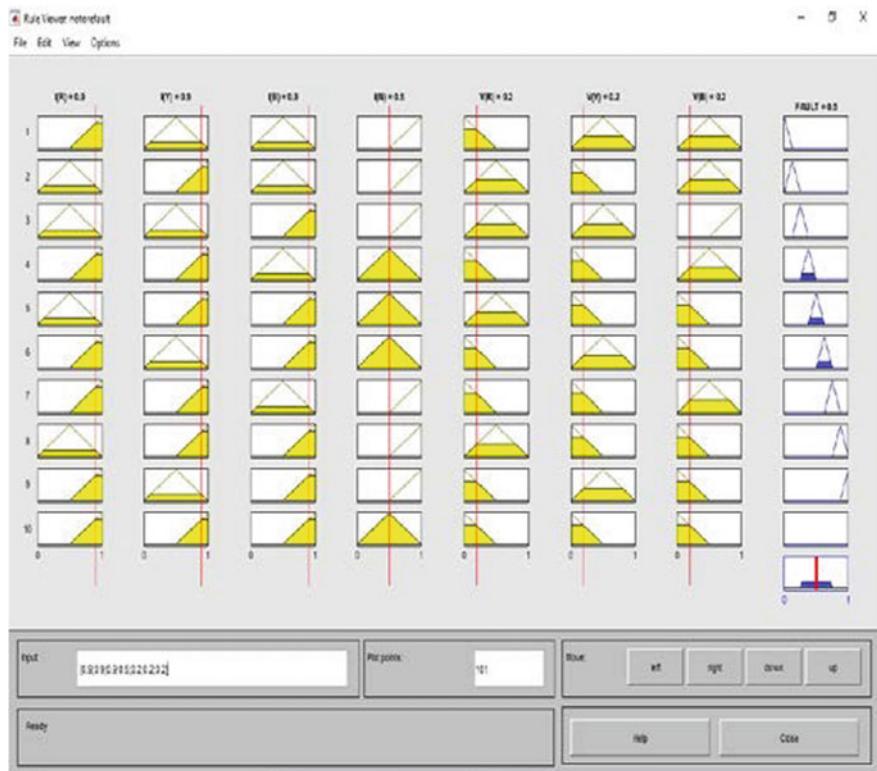


Fig. 8 Rule viewer with three-phase fault detection

fuzzy logic produces the best results. In addition to detection and classification, the relay operation time is critical in eliminating the fault as quickly as possible to avoid any transmission line damage so fuzzy controller can be integrated with the relays. A fuzzy logic-based approach for fault detection and classification for transmission lines has proven to be very efficient and successful in a variety of fault conditions. This technology can not only detect and classify errors, but it can also give real-time automatic protection. The defect detection technique provides a fast and secure response. Because it only requires a few language principles, the proposed logic is simple.

References

1. Koley E, Kumar R, Ghosh S (2016) Low cost microcontroller based fault detector, classifier, zone identifier and locator for transmission lines using artificial neural network: a hardware co-simulation approach. Int J Electr Power Energy Syst 81
2. Saradarzadeh M, Sanaye-Pasand M (2018) An accurate fuzzy logic-based fault classification

- algorithm using voltage and current phase sequence components. *Int Trans Electr Energy Syst* 25(10):2275–2288
- 3. Faig J, Melendez J, Herraiz S, Sánchez J (2010) Analysis of faults in power distribution systems with distributed generation. In: International conference on renewable energies and power quality (ICREPQ'10) Granada (Spain)
 - 4. Jingbo H, Longhua M (2006) Fault diagnosis of substation based on petri nets technology. <https://doi.org/10.1109/ICPST.2006.321717>. Zhisheng Zhang and Yarning Sun, 2007
 - 5. Zhang Z, Zhao J (2017) A deep belief network based fault diagnosis model for complex chemical processes. *Comput Chem Eng* 107:395–407
 - 6. Evans R, Grefenstette E (2018) Learning explanatory rules from noisydata. *J Artif Intell Res* 61(1):1–64
 - 7. Anthapol et al (2017) Behaviour analysis of winding to ground fault in transformer using high and low frequency components from discrete wavelet transform. In: 2017 Int conference on applied system Innovation (ICASI) (Sapporo), pp 1102–1105
 - 8. Mokhlis H (2018) High Impedance fault detection and identification based on pattern Design Optimization for Electric power distribution systems, Ph.D Dissertation, University of Washington, Seatle, WA
 - 9. Fathabadi H (2016) Novel filter base ANN approach for short circuit faults detections classification and location power transmission lines. *Int J Electr Power Energy Syst*
 - 10. Abeid M, El-Ghany HAA, Azmy AM (2017) An advanced traveling-wave fault-location algorithm for simultaneous faults. In: Nineteenth international middle east power systems conference (MEPCON), Cairo, 2017, pp 747–752
 - 11. Singh S, Mamatha KR, Thejaswini S (2019) Intelligent fault identification system for transmission lines using artificial neural network. *IOSR J Comput Eng* 16(1):23–31
 - 12. Jamil M, Sharma SK, Singh R (2020) Fault recognition and classification in electrical power transmission system using artificial neural network. Springerplus, vol 4, no 1
 - 13. Upadhyay S, Kapoor SR, Choudhary R (2021) Fault classification and detection in transmission lines using ANN. In: 2021 International Conference on Inventive Research Computing Applications, no Icirca, pp 1029–1034

Comparative Analysis of Phasor-Phasor and Detailed-Phasor Models of Regulating Transformer



Nitin Karnwal, Mukul Singh , Nidhi Singh, and M. A. Ansari

Abstract The research paper presents the comparative analysis of MATLAB model of a regulating transformer (RT). Two MATLAB models are built for analysis. Each model contains two type of blocks of regulating transformer. One block is phasor type, and the other is detailed type Simulink block. Difference between both types of blocks used and integrated in both models are explained in the research paper. On-load tap changers are used for voltage regulation by altering the turn ratio of the transformer after connecting a tapped winding on each phase. In first model, model 1 is detailed type integrated with model 2 which is phasor type. In second model, both model 1 and model 2 are phasor type. Both models are connected in parallel with each other through a distribution network rated 25 kV. Therefore, Simulink models are developed two times with different types of regulating transformer blocks. The results are obtained after simulation of both Simulink MATLAB models, and four traces of tap positions, superposition of voltage, active power, and reactive power are observed to analyze the impact of voltage regulation because of the tap change regulating transformer. The research paper also focuses on observing and comparing the speed of simulation when we switch from first model to second model of distribution network connected with regulating transformers.

Keywords Regulating transformer (RT) · On-load tap changers (OLTC) · Phasor type model · Detailed type model · Distribution network · Voltage regulation · MATLAB Simulink

1 Introduction

In today's world, stabilization of voltage in distribution network is a major concern specifically during loading conditions. Non-stable voltage conditions may lead to frequent faults, wear, and tear of equipment and devices. To overcome this shortcoming, voltage regulation is performed by the transformer using various methods.

N. Karnwal · M. Singh · N. Singh · M. A. Ansari

Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

One of the method for regulating the voltage is by using tap-change regulating transformer during on-load conditions. Using thyristor for auxiliary switch solved the problem of arcing during tap changing. Recently, based on the concept of bidirectional thyristor, solid-state relays are used because of their non-contact feature while tap changing [1]. It enables the voltage regulation even during the voltage sag or swell associated with the distribution network of the microgrid [2]. The regulating transformer consists of two units, i.e., series and parallel units. The parallel units are given input by series units so as to have an injection of voltage into distribution transmission system [3]. Strict regulation of voltage is necessary because of the attached sensitive electronics. A properly regulated voltage permits for higher voltage drop in output of distribution network and can enhance the reduction of weights of cables connecting the network. It can be achieved by using (R-TRU) regulated transformer rectifier units [4]. High efficiency and improved power density is can also be achieved by partial energy regulation but adding the auxiliary transformer in isolated topology is not preferential as it do not give any advantage while improving power density [5]. It is also observed that polarity of any associated CT is should be reversed manually because the voltage regulation is severely influenced by protection calculation logic [6].

In first model, the detailed type model 1 represents the tap changing switches and characteristics of the transformer, while the phasor type model 2 uses current source for simulation. Model 1 is can be simulated in continuous or discrete mode of simulation, but model 2 is can be simulated in phasor mode of powergui only. This makes the model 2 much faster in simulation speed as compared to model 1. Model 2 is preferable to be used for studies of transient stability. Changes in phasor voltage and current can be observed using model 1. So in order to make system much faster, model 1 of first model is deleted and replaced by the model 2. After replacement, we receive a whole new different second model which makes the simulation of the model, even much faster.

2 Literature Review

Voltage regulation is very important to maintain stability of the buses of distribution network as frequent unregulated changes may lead to severe faults. The voltage regulation is given by Eqs. (1) and (2) as shown.

$$V = (V_s \cdot R) / (R^2 + X^2)^{1/2} \quad (1)$$

The efficiency of the voltage regulation is 71% which very low and undesired by the customers. The impedance magnitude is related to R by the coefficient α , i.e., $\alpha = X/R$.

$$V = (V_s \cdot N) / (1 + \alpha^2 N^4)^{1/2} \quad (2)$$

For real solution of N , $0 < \alpha < 0.5$ corresponding to N positive solutions by $1 < N < \sqrt{2}$. So, maximum regulation is possible if the regulation transformer ratio is maximum with value of 1.414. Also, to compensate the losses, capacitor bank is also used as static compensator connected in parallel with the load [7]. The regulating capacity is enhanced by using new product, i.e., capacity regulating transformer during on-load conditions as is energy saving a specifically used by industrial entities and mining enterprises or at the places where there is a variation in seasonal load [8]. Different operating conditions are studied to understand the changing nature of power network by modeling of transformers. The types of regulating transformers include phase shifting (PST) and under-load tap changing (ULTC) transformer as they regulate the voltage without interfering with the load [9]. The structure of a UHV transformer is highly complex as compared to EHV transformers because of the complex interaction between the parts of transformer leading to complex differential current affecting the working of the differential relay [10]. With the developing economies, the demand of distribution transformers are increasing heavily as they are important equipment in distribution network [11]. So it is very important to build smart, intelligent, and practical strategies to develop the new types of transformer for different new applications with proper safety and speed of operation [12]. The mechanical contacts lead to shocks while connection and arcing during disconnection. The capacitor banks generates more loss due to fluctuations and deviation in voltage [13]. Strong internal resonances are produced due to the regulating winding while choosing the tap position leading to a building up of high resonant voltage in regulating winding thus leading to initiation of ground faults in feeding cable [14]. During fixed tap position, less on-state resistance is provided by mechanical contacts leading to less power loss; however, wear-less commutation is provided by the semiconductor devices while performing the tap change [15]. The generated arc during switching of the mechanical taps increases with increase in the capacity and load current of the regulating transformer. This leads to vaporization and decomposition of insulating oil to generate gas leading to a surge in insulating oil which is not a favorable circumstance under current required conditions [16]. These limitations should be overcome while using tap-change regulating transformer in near future by using IEEE draft standards for general requirements and test code for regulating transformers [17–22].

3 Simulation Model

Two MATLAB Simulink models are developed each containing model 1 and model 2, with the help of Gilbert Sybille (Hydro-Quebec) model. Both model 1, 2, and 3- Φ regulating transformer with rating 47 MVA, 120 kV/25 kV Wye/Delta. The first case of model containing model 1 and model 2 is shown in Fig. 1.

The tap changer is connected in the HV side, i.e., 120 kV side, and transformer regulates the voltage associated with the buses B2 and B4 at 25 kV side. Tap changer has nine switches from zero to eight. Tap zero is nominal position initially with

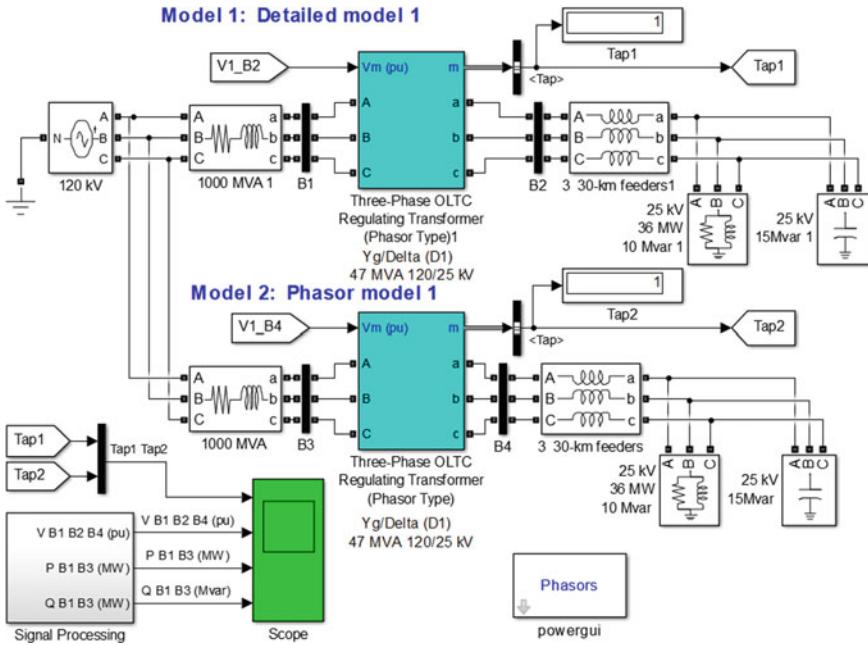


Fig. 1 Detailed-Phasor model of regulating transformer (First case of model)

120 kV/25 kV ratio, and there are total eight different positions for changing the tap from one to eight number. To have positive/additive or negative/subtractive position of tap, reversing switch is added with the regulation winding to allow reverse connections. The voltage correction is possible for ± 0.01875 p.u. of 25 kV at secondary side voltage or $\pm 1.875\%$ of 120 kV of nominal voltage. Hence, range of voltage variation allowed is 102–138 kV or 0.85–1.15 p.u. by tap change of 2.25 kV or 0.01875 p.u. per step. Voltage regulators receive input from buses B2 and B4 in the form of positive sequence voltage with reference voltage at 1.04 p.u. The factor by which transformers boost the voltage with initial position of tap set at -4 is 1.081. Depending upon the need of voltage regulation, the ‘Up’ or ‘Down’ outputs are generated by pulses using the tap changer to move the tap in upward or downward direction for effective voltage regulation. The range of voltage is should be (1.021 < V < 1.059) p.u. at buses B2 and B4 until the tap position does not reach +8 or -8 position. So the permissible error range is 0.01875 p.u. The second case of model containing model 1 and model 2 is shown in Fig. 2.

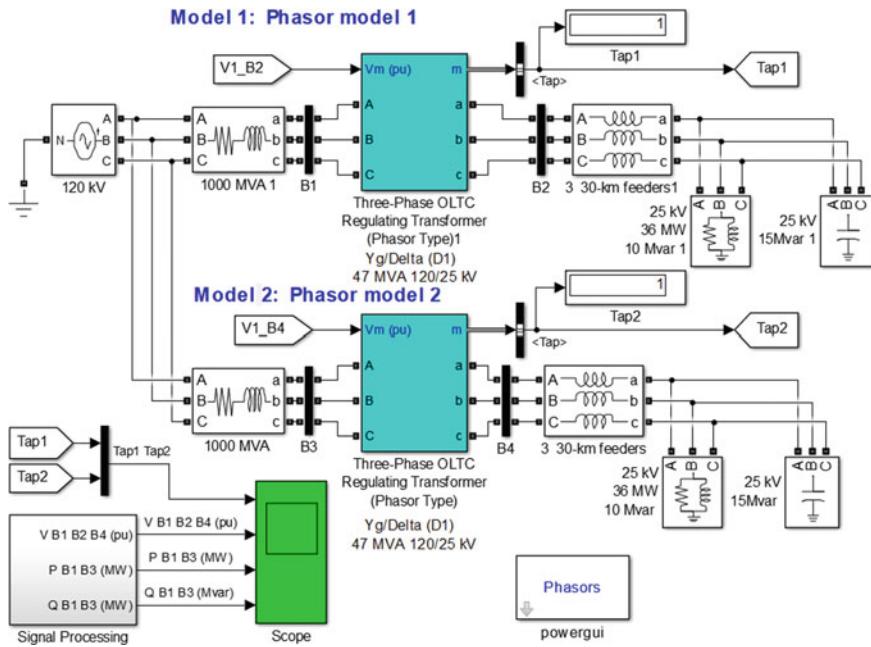


Fig. 2 Phasor-Phasor model of regulating transformer (Second case of model)

4 Results and Discussion

The changing of tap is a quite slow process as it is a mechanical process. It almost takes four seconds to change the tap position. The simulation time is set to 120 s (2 min). At starting, the nominal voltage is being generated by the source, and programmable voltage source is implemented to change the voltage of the system to see the effect on the performance of the tap changer during loading condition of 120 kV system. Later, the voltage is decreased and increased progressively to 0.95 p.u. at time $t = 10$ s and 1.10 at time $t = 50$ s. Various curves/traces were observed and plotted as can be seen in the results above. Figures 3 and 4. display the position of tap of the

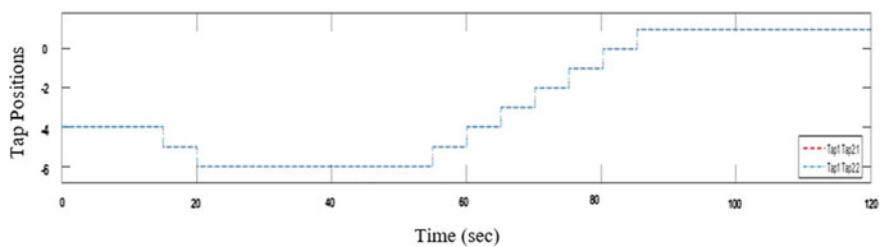


Fig. 3 Tap positions (first model) versus time

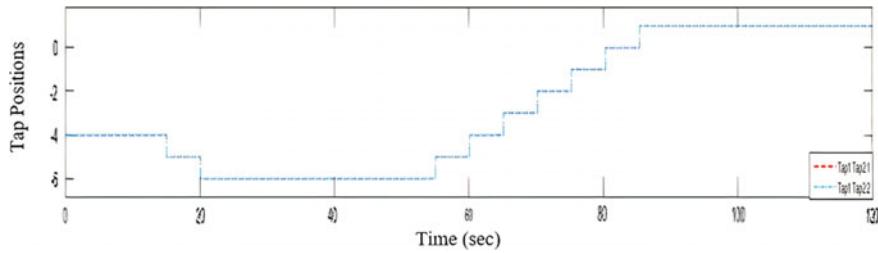


Fig. 4 Tap positions (second model) versus time

tap changers during entire simulation of 120 s. Figures 5 and 6 display the positive sequence voltage superposition at bus B1 (120 kV), bus B2 (2 kV), and bus B4. Figures 7 and 8 display the reactive power, while Figs. 9 and 10 display the active powers of buses B1 and B3 on 120 kV voltage side. A total of 15 MVAR capacitor banks are used for voltage compensation. The voltage regulator regulated the voltage successfully using the tap changers during loading conditions.

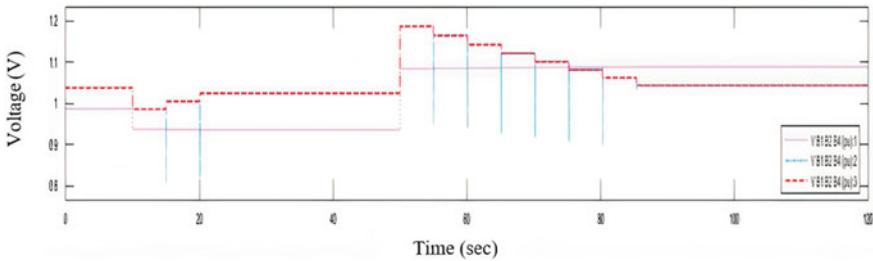


Fig. 5 Voltage variation at B1, B2, and B4 (first model) versus time

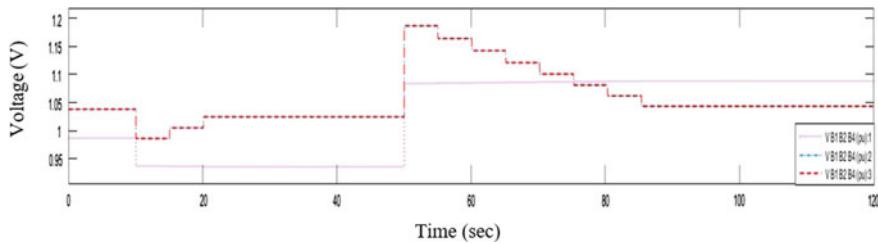


Fig. 6 Voltage variation at B1, B2, and B4 (second model) versus time

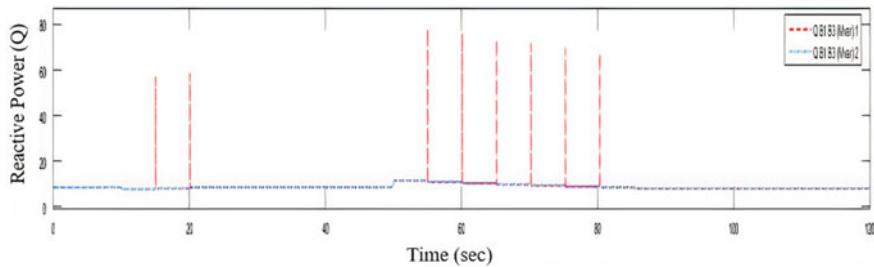


Fig. 7 Reactive power at B1 and B3 (first model) versus time

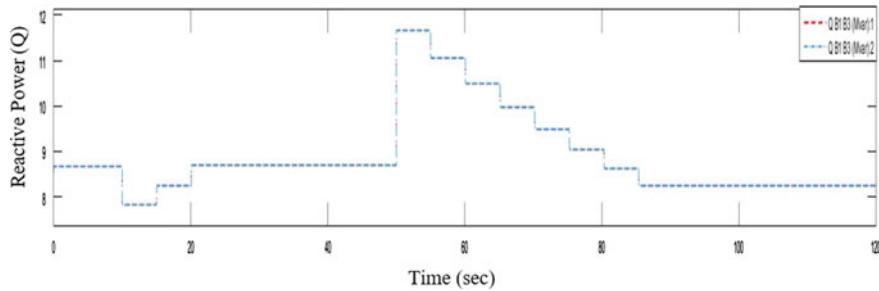


Fig. 8 Reactive power at B1 and B3 (second model) versus time

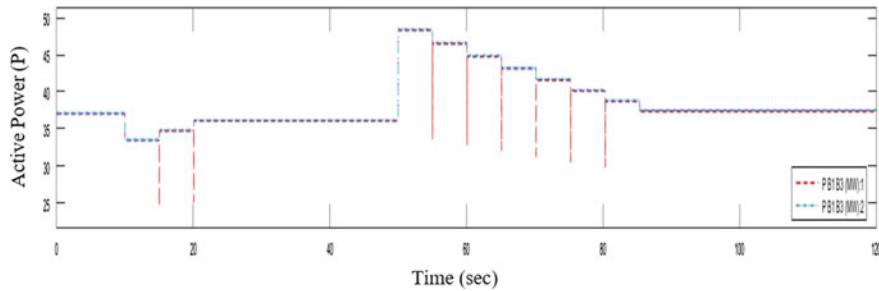


Fig. 9 Active power at B1 and B3 (first model) versus time

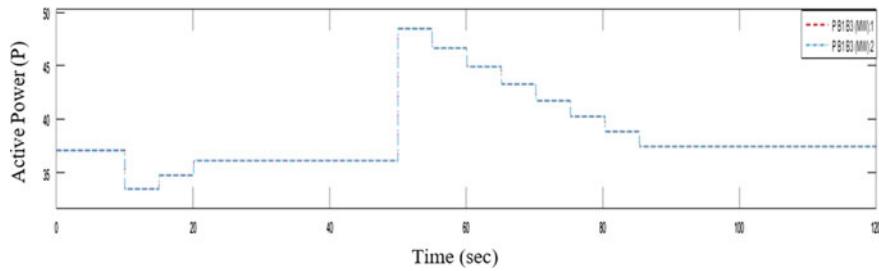


Fig. 10 Active power at B1 and B3 (second model) versus time

5 Conclusion and Future Scope

In this paper, two models of distribution network are built among each of which, two regulating transformers are integrated through parallel connection using MATLAB Simulink blocks. After achieving the desired results and analyzing them, it was observed that second model with phasor-phasor type integrated model runs two and half times faster as compared to detailed-phasor type integrated model connected parallel to the distribution system. Algebraic loops are broken by using first-order transfer function due to which some glitches are observed in voltage during step down at time $t = 10$ s and during step up at time $t = 50$ s which are can be ignored. The tap positions of the tap changer are varied to stabilize the voltage at the buses in the permissible limit of voltage range in p.u. ($1.021 < V < 1.059$). When the internal voltage of the source decreases or increases, tap positions are varied in lower or upward direction to attain stability of voltage at the buses. Hence, voltage is regulated by regulating transformer at the different buses using on-load tap changer, and finally, the system stabilizes at voltage at 1.043 p.u. achieved at tap position +1. In future, on-load tap change phase-shifting transformer is can be built using phasor-phasor and detailed-phasor type MATLAB Simulink model.

References

1. Yulin Z, Shoutian D, Jiahui L, Xin Y, Na Z, Xueli L (2006) Study on non-contact automatic on-load voltage regulating distributing transformer based on solid state relay. In: CES/IEEE 5th international power electronics and motion control conference, pp 1–5. <https://doi.org/10.1109/IPEMC.2006.4778031>
2. Pimenta A, Costa PBC, Paraíso GM, Pinto SF, Silva JF (2020) Active voltage regulation transformer for AC microgrids. In: 2020 IEEE 9th international power electronics and motion control conference (IPEMC2020-ECCE Asia), pp 2012–2017
3. Hu L (1998) A new model for HVDC systems based on model of a regulating transformer. In: 1998 seventh international conference on power electronics and variable speed drives (IEE Conf Publ No 456), pp 105–110. <https://doi.org/10.1049/cp:19980508>
4. Wambsganss WJ (2020) Regulating transformer rectifier unit (R-TRU) for more electric aircraft (MEA). In: 2020 IEEE applied power electronics conference and exposition (APEC), pp 1673–1678. <https://doi.org/10.1109/APEC39645.2020.9124007>
5. Liu T, Han Y, Wu X, Yang S, Xie G (2018) A MHz regulated DC transformer with wide voltage range. In: 2018 IEEE international power electronics and application conference and exposition (PEAC), pp 1–4. <https://doi.org/10.1109/PEAC.2018.8590494>
6. Xiaowei F, Zhendong L, Xun W, Jinbo L, Xiaofei Z, Liming Y (2018) Method for debugging the polarity of the mutual inductor for 1000 kV EHV regulating and compensating transformer. In: 2018 2nd IEEE conference on energy internet and energy system integration (EI2), pp 1–4. <https://doi.org/10.1109/EI2.2018.8582034>
7. Swift GW, Menzies RW, Gole AM (1991) Sensitivity analysis for a regulating transformer connected to a high impedance source. IEEE Trans Circ Syst 38(2):227–229. <https://doi.org/10.1109/31.68303>

8. Weibin L, Weibo F, Shengfeng L, Rongbo P, Meiyun L, Changjiang L (2015) Fault oriented safety technology of on-load capacity regulating transformer. In: 2015 8th international conference on intelligent computation technology and automation (ICICTA), pp 423–425. <https://doi.org/10.1109/ICICTA.2015.112>
9. Murad MAA, Gómez FJ, Vanfretti L (2015) Equation-based modeling of three-winding and regulating transformers using Modelica. In: 2015 IEEE Eindhoven PowerTech, pp 1–6
10. Zheng T, Chen PL, Qi Z, Terzija V (2014) A novel algorithm to avoid the maloperation of UHV voltage-regulating transformers. *IEEE Trans Power Deliv* 29(5):2146–2153. <https://doi.org/10.1109/TPWRD.2014.2301452>
11. Wang J, Sheng W, Wang L, Yang H (2014) Study on technical and economical efficiency of amorphous alloy transformer and on-load capacity regulating transformer in distribution network application. In: 2014 China international conference on electricity distribution (CICED), pp 30–34. <https://doi.org/10.1109/CICED.2014.6991657>
12. Jianjun L, Jinlei H, Xiaoping L, Jingtao Y, Zhongyu Z, Shulin L (2015) Control strategy study and discussion of on-load capacity regulating transformer. In: 2015 8th international conference on intelligent computation technology and automation, pp 337–340
13. Klimash VS, Tabarov BD (2018) The method and structure of switching on and off, and regulating the voltage of a transformer substation. In: 2018 international multi-conference on industrial engineering and modern technologies (FarEastCon), pp 1–4
14. Gustavsen B, Portillo A, Ronchi R, Mjelde A (2018) High-frequency resonant overvoltages in transformer regulating winding caused by ground fault initiation on feeding cable. *IEEE Trans Power Deliv* 33(2):699–708
15. Rogers DJ, Green TC (2013) An active-shunt diverter for on-load tap changers. *IEEE Trans Power Deliv* 28(2):649–657
16. Si J, Hao Z, Zhang Y, Yao S, Ding G, Wu X (2019) Mathematical modeling and simulation of flow field of switching process of on-load tap changer for large capacity transformer. In: 2019 IEEE 8th international conference on advanced power system automation and protection (APAP), pp 1427–1431. <https://doi.org/10.1109/APAP47170.2019.9225051>
17. Singh M, Ansari MA, Tripathi P, Wadhwania A (2018) VSC-HVDC transmission system and its dynamic stability analysis. In: International conference on computational and characterization techniques in engineering & sciences (CCTES), pp 177–182. <https://doi.org/10.1109/CCTES.2018.8674095>
18. Singh M, Singh O, Kumar A (2019) Renewable energy sources integration in micro-grid including load patterns. In: 3rd international conference on recent developments in control, automation & power engineering (RDCAPE), pp 88–93. <https://doi.org/10.1109/RDCAPE47089.2019.8979036>
19. IEEE draft standard for general requirements for liquid-immersed distribution, power, and regulating transformers, pp 1–70, 23 Mar 2021. IEEE PC57.12.00/D1.0
20. IEEE draft standard test code for liquid-immersed distribution, power, and regulating transformers, pp 1–115, 23 Mar 2021. IEEE PC57.12.90/D4
21. IEEE draft standard test code for liquid-immersed distribution, power, and regulating transformers, pp 1–118, 14 Jun 2021. IEEE PC57.12.90/D5
22. Singh O, Singh M (2020) A comparative analysis of economic load dispatch problem using soft computing techniques. *Int J Softw Sci Comput Intell IGI Global* 12(2). ISSN 1942-9045

Intelligent Condition Monitoring of Electrical Assets Using Infrared Thermography and Image Processing Techniques



Rishabh Anand and M. A. Ansari

Abstract Infrared thermography has been frequently utilized by industries to predict, prevent, and maintain the electrical anomalies in the equipment's. A user-friendly thermal image analysis program has been created to detect the hotspot region. The image defect processing techniques, K-means and fuzzy C-mean, are employed to boost the program's detection and analytical capabilities. These image segmentation techniques have been proposed for measuring the faulty region of the apparatus and can also detect abnormalities by analyzing the thermal images.

Keywords Infrared thermography · k -means image segmentation · Anomaly detection · Condition monitoring

1 Introduction

Acoustic monitoring, vibration analysis, and thermography are just a few examples of condition monitoring methods that can be used. It has been described as a non-deleterious test method that detects the anomalies related to temperature in building mechanical installation, short-circuit detection, and loose connection [1]. It is used in accessible areas where there is risk of high voltages and high temperature and constraints for a person to go in, due to which IRT techniques are utilized to calculate high precision and off the grid temperature measurement. IRT- and condition-based monitoring techniques are being used in electricity generation plants because they can encounter regular problems in transmitting and distributing systems. These plants can have inconsequential complication such as loose joining, corrosion, and unbalanced load [2]. Because of the increase in impedance, which increases current, these issues might cause the equipment to overheat [3]. This work uses thermogram-based temperature measuring techniques, which offer a number of benefits for electrical component monitoring, including high spatial resolution, high reliability, wide temperature ranges, and fast response time. By analyzing the thermal pictures and

R. Anand (✉) · M. A. Ansari

Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India

employing image processing techniques, IR thermography can be used as well for fault diagnostics.

2 Related Work

- A. **Zainul Abdin Jaffery, Ashwani Kumar Dubey (2014)** They introduced a noninvasive monitoring and controlling system which detects early faults in electrical equipment's. A new estimation algorithm was used that was used to estimate and monitor the variations in the red color section of the thermogram when there was a rapid temperature spike [4–7]. There are various factors in the speed of the system which depends in such as processing speed, cabling types, distance between the host terminal and camera, protocol types, and medium of transferring data. They have also provided and compared the system which they created invasive type and noninvasive type systems [8–10]. They have also included various factor which may perhaps be responsible for these automatic guiding and monitoring system because of its high speed of operation and accuracy [11]. These factor which required the machine intervention are fatigue, vision level.
- B. **Deepak Kumar, Ansari (2017)** They have used infrared thermography as their means which monitors and diagnose faults in the power electronics equipment's. The noninvasive techniques have been utilized; it uses a thermogram image that regulates and supervises and monitors the power equipment. The region that is impacted by the increment of the temperature is detected by watershed segmentation. The threshold value of the impacted region can be found by segmentation. They can also track the image's temperature and intensity in that area. This can avoid unnecessary damage to the equipment.
- C. **Chandira Sekaranb (2019)** In this, they have proposed a method IRT image segmentation to attain abnormalities in an electrical equipment. They have utilized an innovative method known as MALO algorithm. This optimization technique depends on the basis of hunting mechanism of lion ants. They have separated a malfunctioning component of the electrical system by segmenting the faulty portion of the image [6]. Then, image which has faulty portions can be edited out by the region prop's function which cut out the portion of image that has size less than 50 pixels. They have implemented a method for automated segmentation, and many thresholding methods are used to identify the hot spot on the infrared thermal image with hybrid optimization technique.

3 Infrared Thermography

It has been a long time; thermography inspections have been utilized to perform preventative and predictive maintenance on equipment. When an infrared imaging system is used to sense, display, and record the thermal patterns and temperature

across a given surface, it is done to record the patterns to dish out the probable machine breakdown in the future. The notion of thermography is simple; all objects with a value greater than zero discharge infrared radiation; although infrared radiation is indistinguishable to the human eye, it can be detected with a thermal imager known as infrared camera [12]. Thermography becomes an important part of the company as there is an increase demand which results in increase competition. These company should be able to maintain and manage high-quality electrical products by utilizing less manpower and efficient approach. This will also help them remain more profitable, due to which we can use preventive maintenance such as thermal imaging [7]. This technology is newest nondestructive approach while giving them substantial improvement in performance as well as tremendous cost saving. Infrared thermography can be used to locate loose or deteriorating connections, overloads imbalance loads, and open circuits. It may detect excessive friction, misalignment, and a variety of other issues.

- Electrical distribution system
- Mechanical system
- Structural system
- Research and development.

3.1 Operation Performed for the Detection of Fault

- (a) **Image Acquisition:** Firstly, the medical image is acquired, and then, these images are given as input to the preprocessing stage [6].
- (b) **Preprocessing:** Image preprocessing comes with the parts such as read image, resize as well as removing of noise to enhance image quality to perform further operations. It converts image in digital form and also performs some other operations to get an enhanced image and to extract some useful information [13].
- (c) **Segmentation:** Segmentation is the process of separating the tumor from normal brain tissues. This process helps in dividing an image into multiple segments.
- (d) **Feature Extraction:** Basically, it is the transformation of the original image to a data set with a decreased number of variables that contains discriminated information [13].

4 Thermal Image Analysis

The thermal image analysis is done with the help of MATLAB 2020a, a graphical user interface model, which is made in which the thermal images are analyzed by the color map of the area.

First Proposed Methodology

- The analysis of the thermal images is done via MATLAB 2020a software. A GUI model Fig. 1 is made which is very user interactive and easy to use (Fig. 2).
- The thermal images can be easily loaded for temperature inspection.
- The first image is of a transmission line, and the second one is that of an insulator
- Transmission line image is loaded in the GUI, and then, we detect the temperature at any given point in order to analyze the image in Fig. 3. The detected temperature at the point is 125.25 °F
- Then, the program is run, and the image is analyzed through which the pattern of the affected hotspot region is detected. The highest temperature detected is 183.78 °F. These detected regions are easily marked by green region; this is displayed in Fig. 4.
- Here, we have taken the image of the insulator which is analyzed, and highest temperature is detected in the image shown in Fig. 5. The highest temperature is detected as 105.60 °F.
- Then, we run the program to see and analyze the hotspot region which has abnormally high temperature as compared to its normal running temperature. This is shown in Fig. 6.
- This program can further be improved to have better accuracy and analyzing capability by adding more functions to the codes (Fig. 7).

k-Means Image Clustering

These can be used to separate two different pixels to different classes. Initially, the algorithm will assign different seeds, and then, we will calculate the distance from

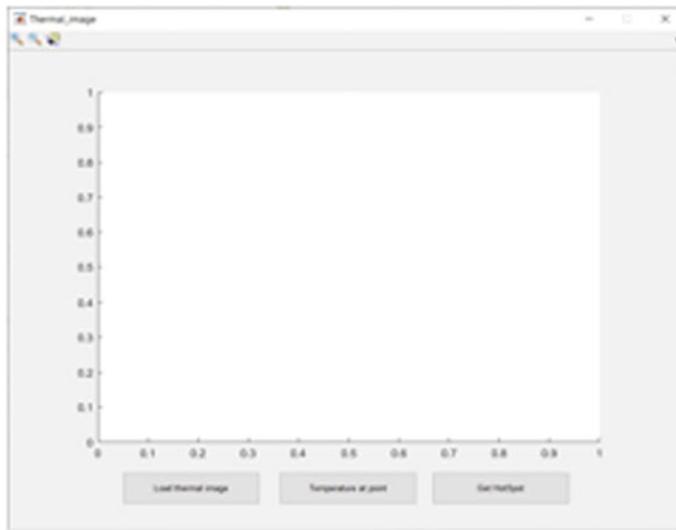


Fig. 1 GUI interface of the model

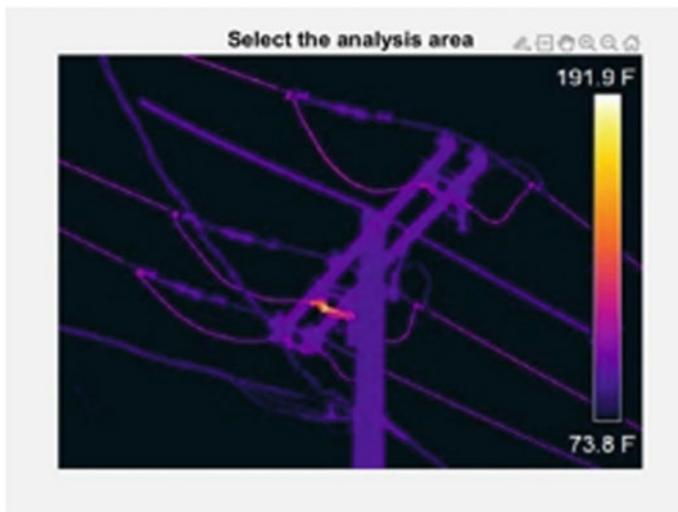


Fig. 2 Original thermal image



Fig. 3 Temperature of the transmission line at the given point

each of the seeds from every data point that the original pixels of the image is shown in Fig. 8. Then, the K-Means algorithm will synchronize the data points and find the closest data set from the seed, it will also assign labels to the data. Then, the clusters of different classes will be found after a few iterations. Then, we will find the seed

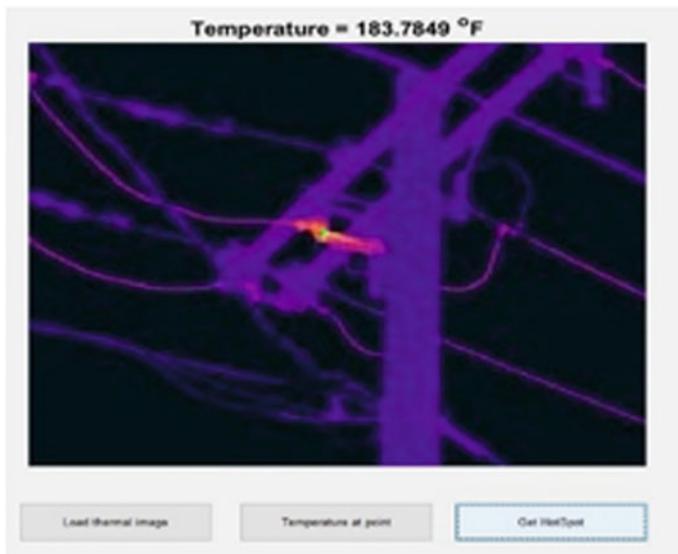


Fig. 4 Highest temperature (hotspot region) detected of the transmission line



Fig. 5 Temperature of the insulator at the selected point

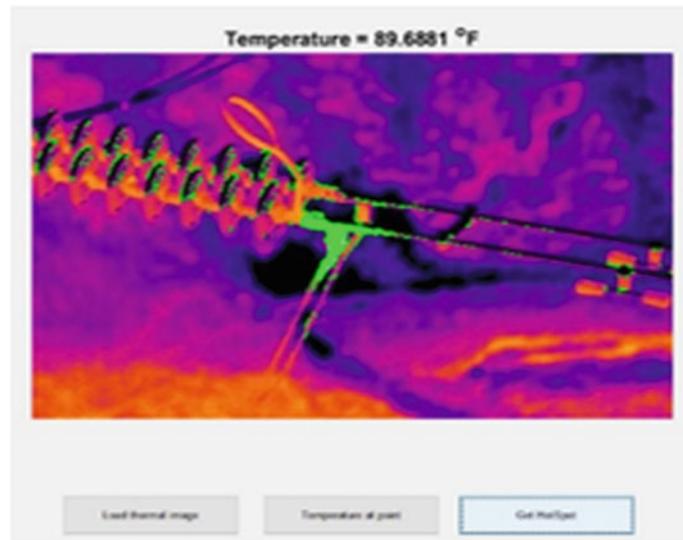


Fig. 6 Hotspot region detected at the insulator

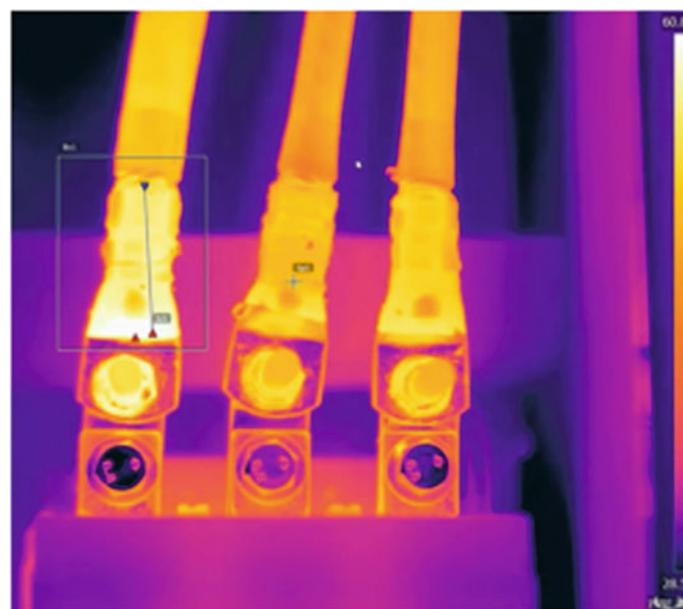


Fig. 7 Original image of the MCB

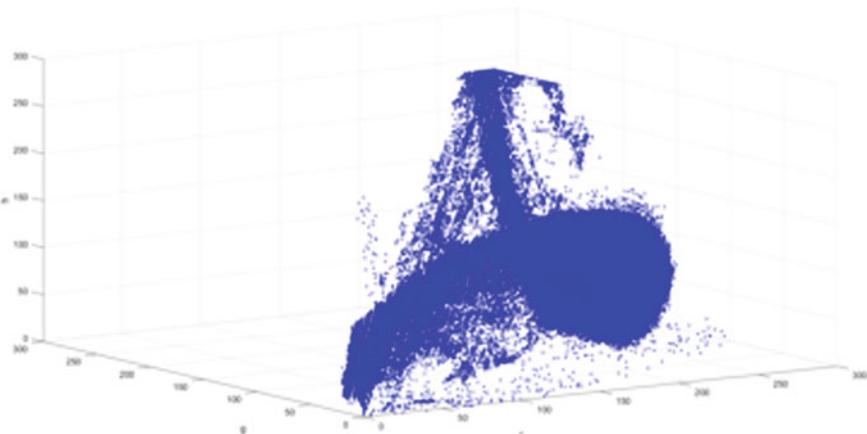


Fig. 8 Pixels of the GB color bands in the 3D space

at the centroid of the clusters where it will not be moving any further after certain iteration shown in Fig. 9. In this, the clusters are the hot part of the thermal images which can be clustered based on the rise of the temperature. It works by choosing the number of clusters you want to find which is k . After which, it randomly assigns the data points to any of the k clusters. Then, the center of the clusters is calculated. The distance of the data point from the centers of each of the clusters is calculated. The distance of each cluster is analyzed and reassigned to the nearest clusters. Then, these steps are repeated until the data point have do not change the clusters or has reached the iterations (Fig. 10).

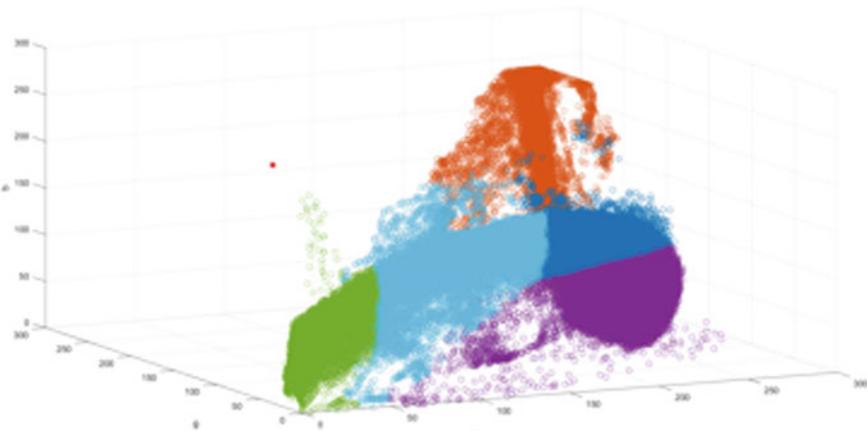


Fig. 9 After some iteration the red dot gets centralized within their respective clusters of RGB bands



Fig. 10 Final segmented image

Fuzzy C-Mean Image Segmentation Technique

It is an unsupervised clustering algorithm and is used for its easy implementation for clustering purposes [14]. The images are segmented according by grouping the pixel that contains or have similar values in a cluster. FCM technique can also further reduce the noise level in the digital image. It is also called soft clustering or soft k -means as its design is centered around K-means. Now, the main difference between K-means and FCM is that it assigns membership for every node of its cluster which in turns allows the node to be partially in more than one clusters [1]. It is necessary to determine an initial cluster for each IRT pixel value in order to perform the FCM algorithm.

Pseudocode for FCM

- Unless the user specifies it, it randomly chooses the c clusters center. A set of starting cluster and membership values are selected.
- It repeats the process and calculate the fuzzy membership U_{ij} for each data in the cluster.
- Then, it computes the centroid v_j for each cluster.
- It will compute the clusters unless the centroid in each cluster does not make any significant changes in the distance.
- The original image used is of a three-phase line from which the heated part will be segmented shown in Fig. 11.
- Noise removal is done with the help of median filters Fig. 12 (Figs. 13 and 14).

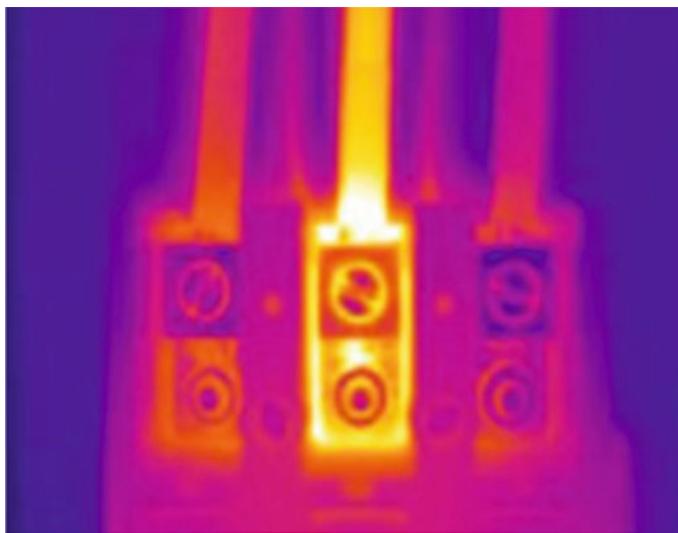


Fig. 11 Original image of the three-phase line

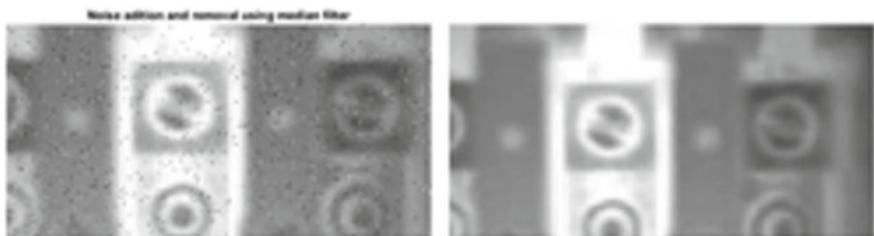


Fig. 12 Removal of noise by using median filter

- The differences between the original and final image is shown via histogram Figs. 15 and 16.

5 Conclusion

In this paper, we have used certain image processing techniques to further improve the process of detecting and analyzing faults in electrical equipment's. The infrared thermal images are segmented to attain their thermal abnormalities condition if found by the system. This also creates an intelligent barrier between the fault prognosis and preventive monitoring. The image processing is done to rectify the dullness present in the digital infrared thermal images. This can also improve the quality of the image. The suggested FCM method was used to segment the high-temperature zone of the

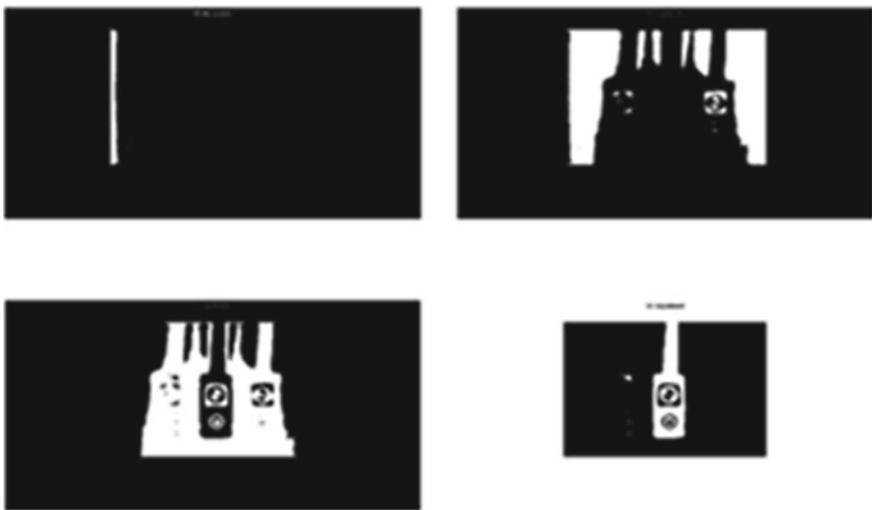


Fig. 13 Intermediate stages of fuzzy C-mean image processing



Fig. 14 Final processed image

switch pictures during the preprocessing step. Threshold-based segmentation is used to detect the high thermal impact regions. The proposed system is simulated for a large number of thermogram images. We can also find the variation of temperature within a given time. Much more information can be accessed. The K-mean algorithm can find the clusters of the temperature variation. Whereas to further update the process, we can do the fuzzy C-mean segmentation to find out more points in the clusters. This helps in preprocessing stage. The high-temperature zone is then identified and segmented. This segmentation technique has given maximum accuracy as compared

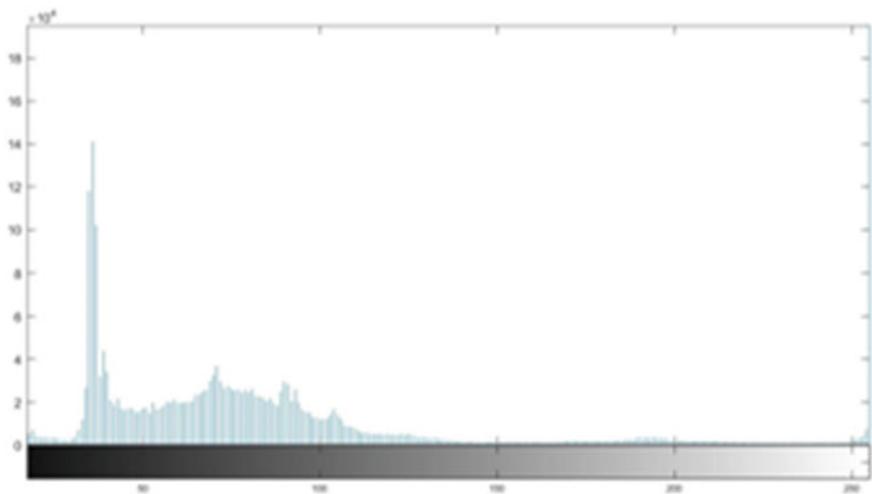


Fig. 15 Histogram of the original image

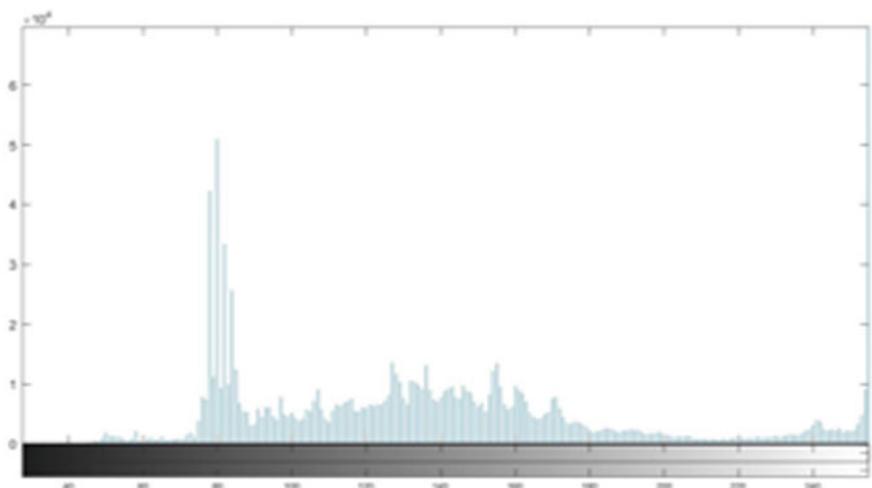


Fig. 16 Histogram of the final segmented image

to other methods. The table is shown below explaining the differences between the two used image processing techniques (Table 1).

Table 1 Difference between the K-means and fuzzy C-mean techniques

Parameter	K-mean	Fuzzy C-mean
1. Execution time	Less	More
2. Accuracy	89.52%	91.65%
3. Number of cluster used	Large no. of clusters are used	Both large and small clusters are used
4. Segmented object	Eight	Four

References

- Heidary K, Caulfield HJ (2017) Pre smoothing effects in artificial color image segmentation. *Comput Vis Image Underst*
- Kacmaz S, Ercelebi E, Zengin S, Cindoruk S (2016) The use of infrared thermal imaging in the diagnosis of deep vein thrombosis. *Infrared Phys Technol* 86:120–129
- Kumar D (2017) Journal article on, condition monitoring of electrical assets using digital IRT and AI technique
- Zidat F, Lecointe JP, Morganti F, Brudny JF, Jacq T, Streiff F (2010) Noninvasive sensors for monitoring the efficiency of AC electrical rotating machines. *Sensors* 10(8):7874–7895
- Ciulavu C, Helerea E (2008) Power transformer incipient faults monitoring. *Ann Univ Craiova Electr Eng Ser* 32:72–77
- Newport R (2002) Improving motor reliability with infrared condition monitoring. *Electr Line* 8(1):36–44
- Christ MCJ, Parvathi RMS (2011) Fuzzy c-means algorithm for medical image segmentation. In: 2011 3rd international conference on electronics computer technology
- Wu Y, Wang Y, Jia Y (2013) Adaptive diffusion flow active contours for image segmentation. *Comput Vis Image Underst* 117(10):1421–1435
- Jadin MS, Taib S (2012) Recent progress in diagnosing the reliability of electrical equipment by using infrared thermography. *Infrared Phys Technol* 55(4):236–245
- Chou YC, Yao L (2009) Automatic diagnosis system of electrical equipment using infrared thermography. In: International conference on soft computing and pattern recognition, pp 155–160
- Asiegbu GO, Haidar AMA, Hawari K (2013) Thermal defect analysis on transformer using a RLC network and thermography. *Circ Syst* 4:52–60
- Jaffery ZA, Dubey AK (2018) Journal article on fault detection technique for electrical assets using infrared thermograms. In: IEEE international conference on recent advances in intelligent computational system
- Chtihrakkannan R, Kavitha P, Mangayarkarasi T, Karthikeyan R (2019) International Journal paper on Breast Cancer Detection using Machine
- Xu C, Xie J, Chen G, Huang W (2014) An infrared thermal image processing framework based on superpixel algorithm to detect cracks on metal surface. *Infrared Phys Technol* 67:266–272
- Jun L, Xinyu L (2011) Heating defect detection system scheme design based on infrared image processing for high voltage plant of substation. In: Advance in control engineering and information science. Elsevier, pp 699–703

Elderly Care Using Generative Adversarial Networks (GANs) on Deep Video Analysis



S. Rajasekaran and G. Kousalya

Abstract Senior treatment at home is a connection that helps hold a loved one aged—or remain at home—as long as necessary. Home care providers hire home health helpers or contract them to attend the home on a fixed schedule. In most organizations, aid personnel or accompanying workers provide help in everyday tasks rather than offering emergency services (ADLs). Senior citizens need support and comfort to lead a safe life from stress and terror. Level of comprehension about the changing activities of senior citizens in their homes leads to their children being exploited by them. This article addresses the questions which influence the lives of the elderly and complicates significant issues in physiology and psychology.

Keywords Facial image analysis · Deep learning · CNN · Neural network

1 Introduction

One of the toughest choices a family can have to undertake is to decide how older parents (e.g., elders) are cared for because they can no longer live autonomously. Families solve a number of challenges in many dynamic and sensitive forms. Any families find opportunities to support aged people in such a manner that they can stay at home comfortably [1]. Some families are moving their parents to take care of them. And other families feel that the only choice for those concerned is to put their aged in a care center.

It is impossible to provide sufficient and reliable elderly treatment and resources. The time-intensive process, often involving consulting with medical and elderly experts, is to decide just what kind of treatment will ideally suited for elderly persons

S. Rajasekaran ()

Department of Information Technology, KGSL Institute of Technology, Coimbatore, Tamilnadu, India

G. Kousalya

Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore, Tamilnadu, India

e-mail: kousalya.g@cit.edu.in

[2]. It also takes time to find cost-effective and relevant quality treatment alternatives. In different countries, different forms of treatment are available, and prices and efficiency vary greatly. Identifying and seeking sufficient and accessible local services for elderly people can be a full-time, demanding task for those concerned [3]. Also, employers and employees can struggle if the elderly scheme burden makes cares less productive at work.

This report is organized as a reference document that helps families overcome the different difficulties associated with the collection and venue. The paper starts by explaining numerous things that an elder might need to be taken care of. The recommendations to select the required treatment standard best for a single elderly individual are explored further. The below are listed different kinds of commonly accessible elder care, followed by a discussion of the features of care facilities or home care services. In conclusion, the paper discusses problems that could emerge in the move to aged treatment for families and elderly persons.

2 Related Works

Al-khafajiy analysis found that Ambient Adaptive Care has been the focus point both for industry and academics as a result of an increasingly rising ageing demographic and the related medical and well-being problems [1]. High political agendas need to control or even lower healthcare prices while enhancing service efficiency. While technology plays a significant role in achieving such ambitions, a plan involves the required domain expertise to be developed, applied and verified. Consequently, remote real-time surveillances of an individual's health may be used to detect repetitions of circumstances that allow for early diagnosis to resolve these problems [4]. The study mentioned in this article focuses thus on the implementation of a smart health management device, which can distantly track the elderly. The research presented in this study is focused around the potential to identify particular disturbances that may assist in early detection practices by monitoring clinical details of an individual. The collection and interpretation of the collected sensory data during the detection of the disease is performed correctly. The result indicates that the suggested framework will strengthen encouragement for treatment practice while promoting procedures for early diagnosis. Our detailed simulation results show the network performance: low latency (96% of transmissions of less than 0.1 s obtained) and small packet losses (1 ms) (only 2.2% of total packets are dropped). Thus, in terms of data collection and maintenance, the system is efficient and cost-effective.

Jenpoomjai research attempts to minimize losses for vulnerable people in real emergencies in community homes. They build a dropping detection framework that can recognize a fall of older people through the TensorFlow API [5]. In order to further evaluate dropping, the suggested VA special algorithm takes account of time, speed and speed variables in humans' flight. It will achieve a more reliable posing calculation. The studies were performed to analyze the device intended to attest to fundamental requirements of violations of traces of actual evidence from human

movements. The findings demonstrate that human activity acceleration can have a relative effect on characterization of behavior. An 88% accuracy on test results for dropping identification is obtained with the suggested technique.

3 Research Methodology

Generative Adversarial Networks (GANs) on Deep Video Analysis for elderly care

The generative opponent networks shown in Fig. 1 are a conceptual simulation technique for deep learning models including convolution neural nets. Generative modeling is a non-monitored learning task in computer vision, which includes automatically recognizes and learns observable behaviors or patterns in inputs so that new models, feasibly derived from the initial dataset, can be generated or produced from them [6]. GANs are the sophisticated way to train a training algorithm by approaching the problem as a managed learning issue in following main: the model generator that we train in order to create new examples and the model discrimination that tries to identify examples as true (from the area) or as false (generated) [7]. Both models are trained together in a 0 Summary game, which is combative, before the discrimination model is fooled for about half the time. As an input, the predictor model produces a sample throughout the field, taking a randomized corrected vector. The parameter is taken from a normal kernel randomly as well as the vector is used to seed the joint distribution [8]. After preparation, points in this multifaceted vector space will be paired by points in the impaired quality, creating a compact distributed processing model. The latent field, or vector space consisting of latent variables, is called this vector space. Innate variables are independent factors that are significant but not explicitly visible to a domain. Find a single-dimension case of what occurs. Let X be a complicated random variable from which we want to test and let U be a

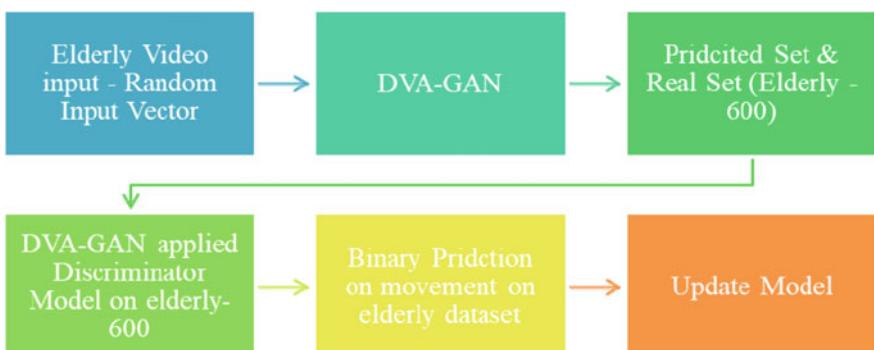


Fig. 1 Flow diagram of generative adversarial networks (GANs) on deep video analysis for elderly care

consistent random variable above [0, 1] from which we can test. We remember that the probability density function of the stochastic process is fully defined (CDF). The CDF of the probability distribution is a function that is described in one dimensional from the specification domain of the probability distribution to the range (0, 1) [9].

$$\text{CDFX}(x) = \mathbb{P}(X \leq x) \in [0, 1] \quad (1)$$

In this specific case, certain different distributions differ among different.

$$\text{CDFU}(u) = \mathbb{P}(U \leq u) = u \quad \forall u \in [0, 1] \quad (2)$$

To be simple, the CDF X feature is expected to be commutative, and its opposite is indicated by using the generalized inverse function. The procedure may simply be expanded to the non-invertible case, but it really is not the key point that we want here in focus. Then can set

$$Y = \text{CDFX} - 1(U) \quad (3)$$

$$\text{CDFY}(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\text{CDFX} - 1(U) \leq y = \mathbb{P}U \leq \text{CDFX}(y) = \text{CDFX}(y) \quad (4)$$

The object of the convert factor is conceptually to change shape the original nonlinear function: the incorporate component uses where the major factor is way too high as the target and places it too low. The simulation of machine learning practically involves specifying two things: a model and a transfer functions [10]. The design of neural network models has already been identified. Two networks are involved: $G(\cdot)$ generated network taking random input z of density p_z and returning the output $x_g = G(z)$ to the intended normal distribution (after training). A exclusionary network $D(\cdot)$ which takes the input x which is either one “true” (x_t where intensity is p_t) or one “produced” (x_g , which p_g volume is the p_z density causing G) which returns the likelihood that $D(x)$ of x will be a “true” data. The generator’s purpose is to delude the discriminator, whose objective is to discriminate between real and produced data. As a result, when attempting to mitigate this error for the discriminator, it is preferable to optimise this error while practising the generator.

$$\begin{aligned} E(G, D) &= 12\mathbb{E}x \sim pt[1 - D(x)] + 12\mathbb{E}z \sim pz[D(G(z))] \\ &= 12\mathbb{E}x \sim pt[1 - D(x)] + \mathbb{E}x \sim pg[D(x)] \end{aligned} \quad (5)$$

$$\max G \min D \quad E(G, D) \quad (6)$$

The absolute best classifier is one that reduces the mediated likelihood density p_g in either given converter G .

$$\mathbb{E}x \sim pt[1 - D(x)] + \mathbb{E}x \sim pg[D(x)] = \mathbb{R}(1 - D(x))pt(x) + D(x)pg(x)dx \quad (7)$$

Optimize the function within the integral for any value of x in order to increase (regarding G) the element. Since the p_t density is distinct from the generator G , we cannot do more than set G to

$$(1/2)\mathbb{R} \min pt(x), pg(x)dx = \mathbb{R} \min pt(x), pg(x)pt(x) + pg(x)pt(x) + pg(x)2dx \quad (8)$$

So it has shown that perfect point for the observing the process is that the generator generates the same volume as the real density in an ideal scenario, and the classifier cannot do better than be true to one out of every two cases as the intuition tells us. Ultimately, note that G also optimizes.

4 Experiments and Results and Discussions

Deep video processing applies DVA-GAN (not “digital versatile disc” but “dual video classifier”) for video production to massive data. DVA-GAN will generate videos up to 256×256 quality and up to 48 frame length. The proposed work has obtained state-of-the-art outcomes elderly move on Distance Prediction Challenge for Elderly-600, and the state-of-the-art Iteration Ranking for Senior Citizens Synthesis-101 datasets. Here, Figs. 2 and 3 are compilations of four-second synthesized short videos trained on 12 128 \times 128 objects from Senior Citizens-600, a large 10-s, high-resolution short video set initially intended for the detection of human behavior. At first sight, the clips appear to show familiar acts such as walking, skating and swimming. But a closer look reveals that much of the created streaming video is distorted, indecipherable or even unreal. Below is a more randomized batch of shortened DVA-GAN samples educated on 48 blocks of 64×64 elderly-600 [11]. The synthesized face of the elders looks pretty realistic. Despite several visual distortions, the DVA-GAN paradigm has increased the efficiency of video generation. While large-scale, elevated data is the fuel that drives model—based efficiency, research have struggled to effectively train earlier video important advance on large datasets due to high data memory and mathematical specifications.

Deep video analysis GAN has addressed in Fig. 2 this obstacle by applying its residence face detection model control discriminator GAN to video and incorporating new strategies for speeding up training, such as a double design consisting of a geographical discriminator and a time softmax, and separable self-interest implemented in a row around the height, width and time axes.

DVA-GAN was tested on elderly-600, a smaller dataset of 33,110 images of human behavior, and a model generated samples with a state-of-the-art implementation rating of 45.24. In yet another test, the DVA-GAN model, with certain changes, surpassed previous frame-conditional estimation work for Kinetics-600. DeepMind

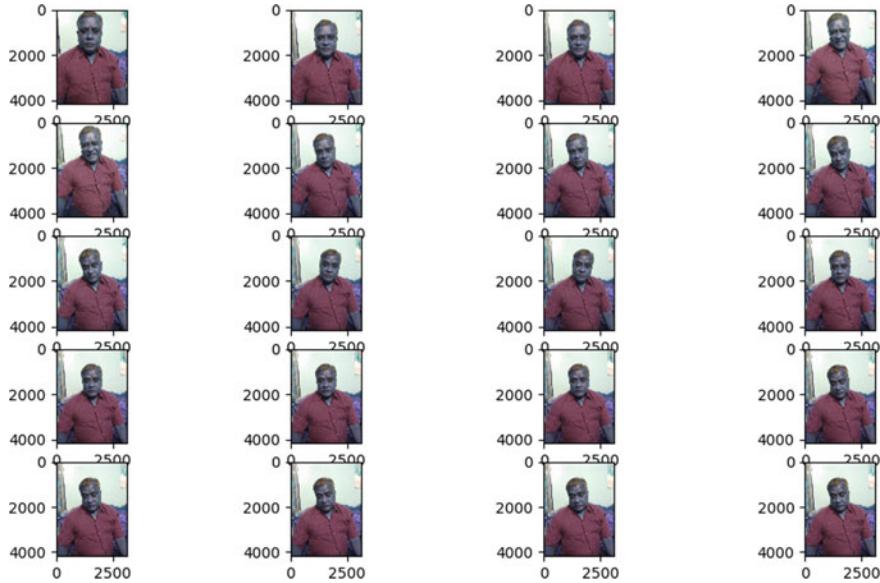


Fig. 2 Different camera live movements (hand, leg, eyesight) of persons with their behaviors (moving, sitting, sleeping) analysis by DAV-GAN

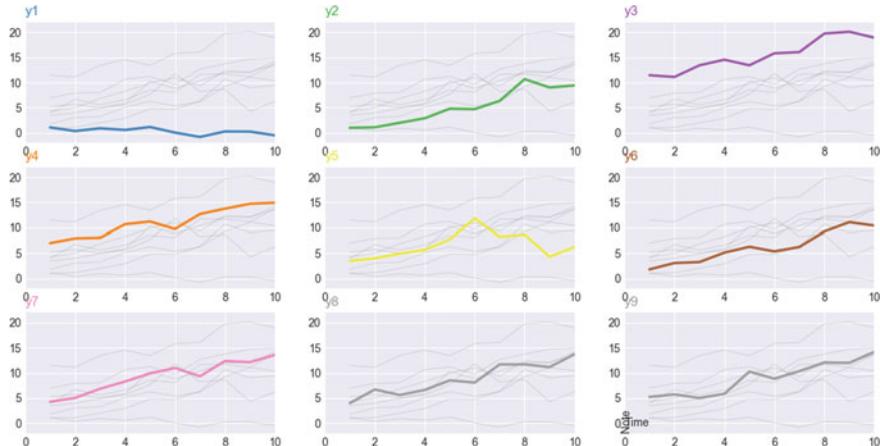


Fig. 3 Detailed review of all camera live variables in the different room's movement and sound-based classification human operation by DAV-GAN

has also developed a new reference measure for generative video modeling for senior citizens-700, with the findings of DVA-GAN as a clear reference. DVA-GAN was tested on elderly-600, a limited set of 33,110 images of human behavior, and a model generated samples with a particular state implementation rating of 45.24 in Table 1.

Table 1 Compared with existing elderly care video analysis with various video duration using the proposed video DVA-GAN

Methods	Video file taken for analysis				Classification true positive ratio (%)			Classification false positive detected			False positive ratio (%)	
	823	745	632	91.34	91.41	89.5	78	70	74	0.08	0.08	0.10
Al-khafajy	823	745	632	91.34	91.41	89.5	78	70	74	0.08	0.08	0.10
P. Jenpoomjai	790	713	609	87.68	87.48	86.2	111	102	97	0.12	0.12	0.13
Proposed DVA-GAN	876	789	684	97.22	96.80	96.8	25	26	22	0.02	0.03	0.03

The number of video files captured in 09 minute's playable duration: 901

The number of video files captured in 12 minute's playable duration: 815

The number of video files captured in 17 minute's playable duration: 706

In yet another test, the DVA-GAN model, with plenty of changes, surpassed previous frame-conditional estimation work for Kinetics-600. DeepMind has also developed a new reference measure for conceptual video modeling for senior citizens-700, with the findings of DVA-GAN as a clear reference in Figs. 2 and 3.

5 Conclusion

The entire article proposes an elderly surveillance system based on multi-information fusion technologies such as video, noise, infrared and pulse to prevent family members from wasting time and resources caring for the old. Specific guidance is provided to help residents at home, as well as their ageing parents and seniors. Here is an algorithmic network that employs two neural networks. In the creation of elderly analysis, picture, video and audio are frequently used. It is possible to train a GAN to create random noise images. The higher the visual quality of each video frame, the better the GAN can be trained to make senior films out of digital images that appear to be enlarged digital images from the elderly database. A GAN has two parts: an image generator and a disc driver that categorizes pictures and movement. Other stages are identified and classified with more than 97.22% precision and 0.02% false identification using the proposed DVA-GAN method of falling classification.

References

1. Al-khafajiy M, Baker T, Chalmers C et al (2019) Remote health monitoring of elderly through wearable sensors. *Multimed Tools Appl* 78:24681–24706. <https://doi.org/10.1007/s11042-018-7134-7>
2. Schrader L, Vargas Toro A, Konietzny S et al (2020) Advanced sensing and human activity recognition in early intervention and rehabilitation of elderly people. *Popul Ageing* 13:139–165. <https://doi.org/10.1007/s12062-020-09260-z>
3. Dubey R, Ni B, Moulin P (2012) A depth camera based fall recognition system for the elderly. In: Campilho A, Kamel M (eds) Image analysis and recognition. ICIAR 2012. Lecture Notes in Computer Science, vol 7325. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31298-4_13
4. <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>
5. <https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29>
6. Jenpoomjai P, Wosri P, Ruengittinun S, Hu C, Chootong C (2019) VA algorithm for elderly's falling detection with 2D-pose-estimation. In: 2019 Twelfth international conference on Ubimedia computing (Ubi-Media), Bali, Indonesia, pp 236–240. <https://doi.org/10.1109/Ubi-Media.2019.00053>
7. <https://syncedreview.com/2019/07/16/deepmind-dvd-gan-impressive-step-toward-realistic-video-synthesis/>
8. <https://venturebeat.com/2019/12/26/gan-generative-adversarial-network-explainer-ai-machine-learning/>
9. Leite G, Silva G, Pedrini H (2019) Fall detection in video sequences based on a three-stream convolutional neural network. In: 2019 18th IEEE international conference on machine

- learning and applications (ICMLA), Boca Raton, FL, USA, pp 191-195.<https://doi.org/10.1109/ICMLA.2019.00037>
10. Han H, Jain AK, Wang F, Shan S, Chen X (2018) Heterogeneous face attribute estimation: a deep multi-task learning approach. In: IEEE transactions on pattern analysis and machine intelligence, vol 40, no 11, pp 2597–2609, 1 Nov 2018. <https://doi.org/10.1109/TPAMI.2017.2738004>
 11. <https://medium.com/deep-math-machine-learning-ai/ch-14-general-adversarial-networks-gans-with-math-1318faf46b43>

Performance Analysis of kNN-Based Image Demosaicing for Variable Window Sizes



Gurjot Kaur Walia and Jagroop Singh Sidhu

Abstract Image demosaicing plays a prominent role when it comes to healthcare, forensics, and other imaging fields. The representation of an image in terms of numbers which can be understood and dealt with computers is termed as digital image. The process of recreating an entire colored image from incomplete color sample outputs linked to an image sensor is termed as demosaicing. The proposed work highlights the application of kNN algorithm for different window sizes. The comparison deduced is based on the evaluation metrics, viz. SNR, PSNR to detect the best window size for a set of images as they help in deciding the visualization of an image.

Keywords Digital image · Demosaicing · CFA · kNN · PSNR

1 Introduction

The recreation of full shading pictures from a CFA-based indicator requires a cycle of controlling the estimations of some other shading partitions at every pixel. The strategies for these sorts are normally alluded as shading interpolation. The demosaicing algorithms should be constructed in a way so that the image's resolutions are preserved to a large extent, artifacts should not be introduced, and fast processing with low computational complexity should be achieved. Due to the rapid growth of technology and social media, sharing of such content has increased. The quality of image is the focus behind the improvement of the existing technology as digital devices are affordable by all. Image quality generally depends on the various characteristics of devices like its resolution, sensitivity, sensor's dynamic range etc., but color filter array interpolation plays an important part to obtain a full image via CFA

G. K. Walia (✉)

Electronics and Communication Engineering, I. K. Gujral Punjab Technical University, Kapurthala, Punjab 144603, India

J. S. Sidhu

Department of ECE, DAVIET, Jalandhar, Punjab 144021, India

data [1–3] whose requirement is that there should be minimum three color samples at each pixel location and before each sensor, a color filter is accommodated in order to obtain three color channel images. It turned out to be quite challenging in alignment as well as cost, as there is a necessity of three charge-coupled devices (CCD) sensors with accurate alignment. So, the alternative to this is to use filter array in front of the sensor which will acquire a color component at a pixel and consequently retrieving the rest two color components [4, 5]. And this whole phase of reconstructing an entire color image from 1 color component is termed as demosaicing. There is an availability of number of CFA patterns, but the preferred one is Bayer pattern and it is also a recommended one [3]. Furthermore, there are various Bayer CFA algorithms based on interpolation, reconstruction, frequency domain etc. The widely used methods for the image demosaicing are based on the interpolation which generally involves estimation of missing color components by interpolation of neighboring pixels. But, it sometimes leads to certain errors such as blurring and artifacts. In [6], residual interpolation-based approach in combination with fuzzy edge strength is suggested where estimation of green channel is done by considering an edge filter. Further, the computation of other two channels is done by considering green channel as a guiding image. Second category is reconstruction methods that provide solution like optimization problems with proper regularization. Third category is frequency domain methods that deals with frequency components that are to be corrected [7]. Fourth category involved learning-based approaches dependent on CNN in [8] and [9] that aid to improve the performance of demosaicing particularly in noisy images. Still there is a need to achieve better image quality that is done by employing generative adversarial network using U-net in parallel with dense residual network which also eliminated the image artifacts [10]. However, these algorithms are complex and consume a lot of time, thereby resulting in increased computational cost of the methods. It is observed from the plethora of image demosaicing algorithms that the trade-off has to exist between accuracy as well as computational cost. The gaps that were found in literature were that different configurations of CFA Bayer pattern were not used. Also, the problem of illuminate normalization was ignored. Although the above-mentioned techniques provided a detailed overview of the methods for image demosaicing, it is summarized in Table 1.

Organization of papers is as follows: Sect. 2 contains detailed information of the methodology followed. Section 3 discusses the results of parameters for demosaicing algorithm, and conclusion is available in Sect. 4.

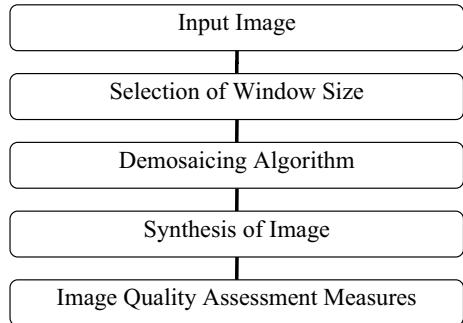
2 Methodology

The first step in the proposed algorithm takes an input image and generates interpolated arrays with these input images. After these arrays are generated, then a window size out of 3×3 , 5×3 , 5×5 or 7×5 is selected followed by the process of demosaicing using kNN for the synthesis of the input image. In order to yield better quality, the need to make use of equitable proportion of the classification window

Table 1 Summary of image demosaicing techniques

Reference	Methodology used	Outcomes
[11]	Jacobian matrix, neighborhood voting	Computation required to produce effective results is reduced
[12]	Spectral and spatial correlation-based adaptive demosaicing method	Adaptive median filter suppressed demosaicing artifacts
[13]	Edge-sensing-based spectral model is proposed for universal demosaicing	Sharpness of image is retained
[14]	Median filter, spatial deinterlacing	Provided better accuracy along with suppression of artifacts
[15]	Cubic spline, Taylor series in combination with weighted median filter to produce the output	Possess capability to preserve edges
[16]	Correlation of pixels in eight directions is considered	Reduction in artifacts
[17]	Green channel information along with sub-sampled R, B images	Aids in reduction of distortion due to aliasing
[18]	Frequency domain-based CFA involving reformulated constrained optimization	Improved speed for searching parameters
[19]	Overlapped frequency components were considered	Reduction of visual artifacts
[20]	Wavelet transform	Resolution of spatial component is enhanced
[21]	Both frequency domain and wavelet decomposition-based hybrid demosaicing algorithm is provided	Reduction of color artifacts near edges
[22]	GSM approach based on wavelets	Suppressed zippering artifacts
[23]	Posteriori estimation in order to minimize multi-term cost function	Provided effective results for both applications
[24]	Minimization of variational function was taken into consideration using restriction of consistency	Effective outcome even in noisy images
[25]	Considering self-similarity in case of natural images for calculating missing elements	Artifact are avoided for better results
[26]	Error reduction using residual interpolation	Reduced color artifacts and error in reconstruction of edges
[27]	Adaptive residual U-net	Reduction of textural artifacts
[28]	Deep learning method based on regularization and convex optimization for image demosaicing	Improvement in quality of reconstructed image

Fig. 1 Flowchart for process of image demosaicing



is must as it will contain sufficient data to estimate the missing pixels, but the thing to be kept in mind is that also that the window size should not be huge as it will become expensive. After the synthesized images are processed, the quality of the synthesized image is validated using SNR, PSNR. Two interpolation techniques, linear and nearest neighbor are used here. Few images from Kodak dataset [29] and other random images were used for training kNN, and afterward five images were used for testing. The kNN works in following manner: Firstly, training and testing data is loaded, then value of k is selected, which is 5 in this case. Further, based on distance between testing data and training data, sorting is done. Eventually, the top k rows from the sorted array are considered and assign a class to the test point based on most frequent class of these rows (Fig. 1).

3 Results

In this study, the kNN technique for image demosaicing was tested on different window sizes. Table 2 shows the PSNR values for R, G, B pixel for five test images taken from Kodak dataset [29].

From the numbers obtained after simulations, it can be deduced the overall impact of 7×5 is better than other window sizes for all the images in terms of R, G, B pixel values. In image1, for green pixel, it is 36.583 for 7×5 when compared to 34.183 for 3×3 , i.e., improvement of 6.56%. Similarly for image5, the red component PSNR of 7×5 shows improvement by 6.27 % with respect to 3×3 . For few components, although,, there is slight variation in PSNR values; in some, 5×3 shows better results than 5×5 , but overall 7×5 has remarkable impact.

Figure 2a shows one of the intermediate processes used while implementing the demosaicing, Fig. 2b, c and d displays the output images. Figures 3 and 4 depict the comparison of SNR and PSNR for various window sizes

From the results obtained, it is evident that for a particular image, SNR and PSNR values for 7×5 window size is coming out to be best. In particular, the SNR of

Table 2 Comparison of PSNR values for R, G, B pixel

Image No.	Image	R, G, B pixel values	Window Size			
			3 × 3	5 × 3	5 × 5	7 × 5
1		G	34.183	35.631	36.552	36.583
		R	33.474	33.499	34.205	33.899
		B	33.733	32.262	32.771	33.141
2		G	32.302	32.576	32.663	32.571
		R	31.197	30.74	31.177	31.945
		B	30.322	30.359	31.093	33.341
3		G	33.676	35.469	35.03	35.089
		R	33.418	33.671	33.631	33.867
		B	34.166	33.281	33.964	36.254
4		G	33.235	33.537	33.528	33.606
		R	31.845	32.121	32.521	32.446
		B	31.945	30.485	30.91	32.028
5		G	33.321	32.841	32.94	32.961
		R	31.714	32	32.591	33.836
		B	31.483	30.802	31.254	32.556

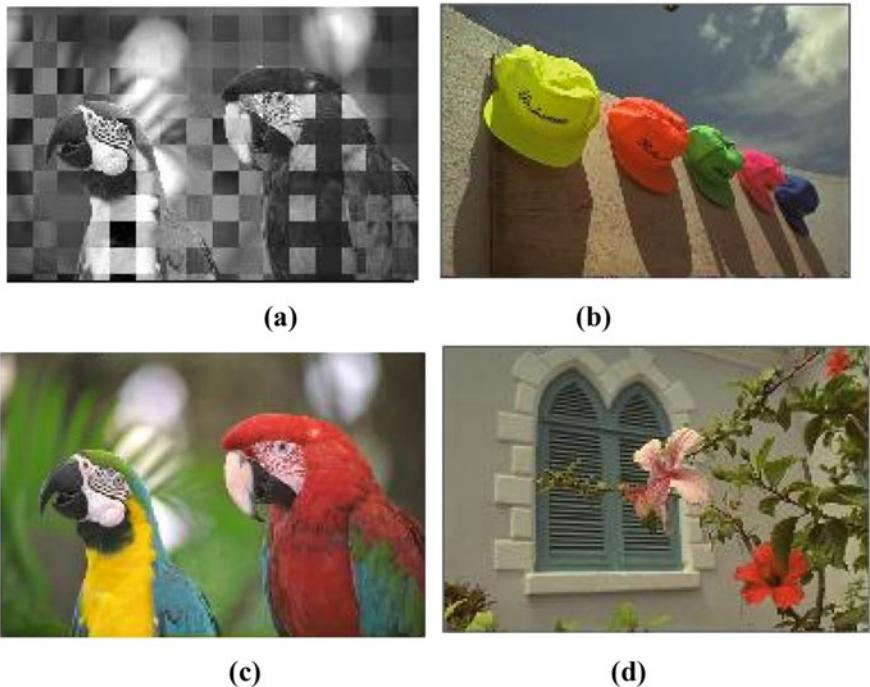


Fig. 2 **a** Shows the GBRG pattern of image3; **b, c, d** Shows the output images

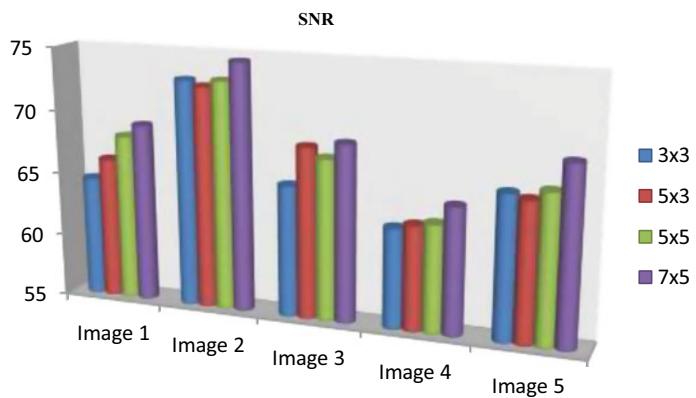


Fig. 3 SNR values for various window sizes

image1 for 7×5 window size shows improvement of 6.62% as compared to 3×3 . Also, for image3, improved SNR by 5.15% is achieved when related to 3×3 .

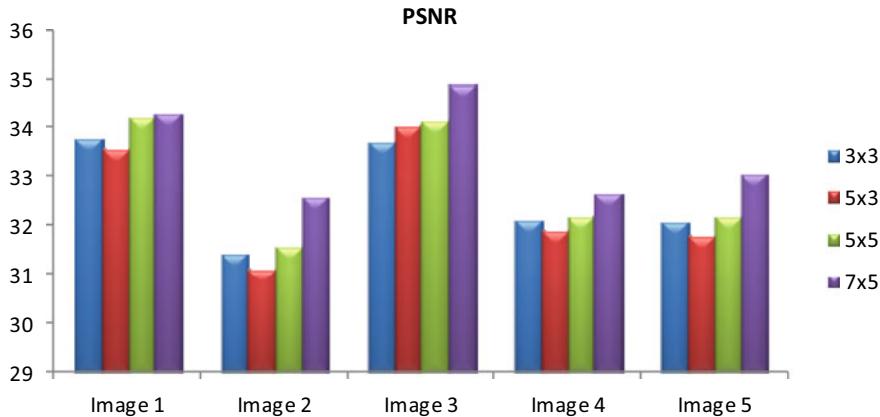


Fig. 4 PSNR values for image1–5

Similarly, 3.48% improvement in PSNR is witnessed for 7×5 window size for image3 as compared to 3×3 . For image5, PSNR of 3.88% is attained for 7×5 in contrast to 5×3 .

4 Conclusion

Different images have been used for analyzing the effect of using kNN on image demosaicing. The images have been taken from Kodak database, and after applying the concept of kNN using different window sizes of 3×3 , 5×3 , 5×5 , and 7×5 , performance metrics, SNR and PSNR have been measured. After extensive simulations, it is deduced that performance of 7×5 is superior. Since better the metrics, more efficient the algorithm is, thus, window size 7×5 has outperformed significantly in terms of SNR and PSNR.

Acknowledgements Dean RIC, I. K. Gujral Punjab Technical University is highly acknowledged for facilitating this work.

References

- Prakash VNVS, Prasad KS, Prasad TJC (2017) Color image demosaicing using sparse based radial basis function network. *Alexandria Eng J* 56(4):477–483
- Boccuto A et al (2019) A fast algorithm for the demosaicing problem concerning the bayer pattern. *Open Signal Process J* 6(1)
- Bayer BE (1976) Color imaging array. U.S. Patent No. 3,971,065

4. Gunturk BK, Glotzbach J, Altunbasak Y, Schafer RW, Mersereau RM (2005) Demosaicing: color filter array interpolation. *IEEE Signal Process Mag* 22(1):44–54
5. Li X, Gunturk B, Zhang L (2008) Image demosaicing: a systematic survey. In: Visual communications and image processing 2008, vol 6822. International Society for Optics and Photonics, p 68221J
6. Sun BY, Yuan N, Zhao Z (2019) A hybrid demosaicing algorithm for area scan industrial camera based on fuzzy edge strength and residual interpolation. *IEEE Trans Ind Inf*
7. Menon D, Calvagno G (2011) Color image demosaicing: an overview. *Signal Process Image Commun* 26(8–9):518–533
8. Ni Z, Ma KK, Zeng H, Zhong B (2020) Color image demosaicing using progressive collaborative representation. *IEEE Trans Image Process* 29:4952–4964
9. Tang J, Li J, Tan P (2021) Demosaicing by differentiable deep restoration. *Appl Sci* 11(4):1649
10. Luo J, Wang J (2020) Image demosaicing based on generative adversarial network. *Math Prob Eng*
11. Kakarala R, Baharav Z (2002) Adaptive demosaicing with the principal vector method. *IEEE Trans Consum Electron* 48(4):932–937
12. Lu W, Tan YP (2003) Color filter array demosaicing: new method and performance measures. *IEEE Trans Image Process* 12(10):1194–1210
13. Lukac R, Plataniotis KN (2005) Universal demosaicing for imaging pipelines with an RGB color filter array. *Pattern Recogn* 38(11):2208–2212
14. Huang WT, Chen WJ, Tai SC (2010) Color demosaicing using deinterlacing and median-based filtering techniques. *J Electr Imaging* 19(4):043018
15. Li JSJ, Randhawa S (2009) Color filter array demosaicing using high-order interpolation techniques with a weighted median filter for sharp color edge preservation. *IEEE Trans Image Process* 18(9):1946–1957
16. Chen X, He L, Jeon G, Jeong J (2014) Multidirectional weighted interpolation and refinement method for Bayer pattern CFA demosaicing. *IEEE Trans Circuits Syst Video Technol* 25(8):1271–1282
17. Glotzbach JW, Schafer RW, Illgner K (2001) A method of color filter array interpolation with alias cancellation properties. In: Proceedings 2001 international conference on image processing (Cat. No. 01CH37205), vol 1. IEEE, pp 141–144
18. Hao P, Li Y, Lin Z, Dubois E (2010) A geometric method for optimal design of color filter arrays. *IEEE Trans Image Process* 20(3):709–722
19. Bai C, Li J, Lin Z, Yu J (2016) Automatic design of color filter arrays in the frequency domain. *IEEE Trans Image Process* 25(4):1793–1807
20. Driesen J, Scheunders P (2004) Wavelet-based color filter array demosaicing. In: 2004 international conference on image processing. ICIP'04, vol 5. IEEE, pp 3311–3314
21. Kolta RWB, Aly HA, Fakhr W (2011) A hybrid demosaicing algorithm using frequency domain and wavelet methods. In: 2011 international conference on image information processing. IEEE, pp 1–6
22. Zhang J, Sheng A, Hirakawa K (2018) A wavelet-GSM approach to demosaicing. *IEEE Signal Process Lett* 25(6):778–782
23. Farsiu S, Elad M, Milanfar P (2005) Multiframe demosaicing and super-resolution of color images. *IEEE Trans Image Process* 15(1):141–159
24. Condat L (2009) A generic variational approach for demosaicing from an arbitrary color filter array. In: 2009 16th IEEE international conference on image processing (ICIP). IEEE, pp 1625–1628
25. Buades A, Coll B, Morel JM, Sbert C (2009) Self-similarity driven color demosaicing. *IEEE Trans Image Process* 18(6):1192–1202
26. Asiq MS, Emmanuel WS (2019) Efficient colour filter array demosaicing with prior error reduction. *J King Saud Univ Comput Inf Sci*
27. Shopovska I, Jovanov L, Philips W (2018) RGB-NIR demosaicing using deep residual U-Net. In: 2018 26th telecommunications forum (TELFOR). IEEE, pp 1–4

28. Kokkinos F, Lefkimiatis S (2018) Deep image demosaicking using a cascade of convolutional residual denoising networks. In: Proceedings of the European conference on computer vision (ECCV), pp 303–319
29. Kodak Lossless True Color Image Suite (Online). Available from: <http://r0k.us/graphics/kodak/>

A Novel Surface Crack Detection and Dimension Estimation Using Image Processing Technique



K. Shreyank, K. Yukta, N. Sowmya, M. Komal, V. S. Saroja, and S. Suhas

Abstract The authors present a novel crack detection system based on image processing. The authors have proposed a novel algorithm for detecting the crack on the surface and estimating the dimension of the crack. Monitoring the health of any character is essential, and the cracks developed on the surfaces often lead to a reduction in the strength of the material. Manual inspection is one of the most common methods for detection of crack. This approach is very time consuming and also depends upon knowledge as well as experience. It also lacks objectivity in the quantitative analysis. The proposed crack detection and dimension estimation using an image processing' system focus on automating the process using established digital image processing techniques. It is a standalone embedded device that detects the surface cracks through images captured via camera and estimates the size of the crack with proper calibration. The experimental results show that the mean error is 8.12%, and the accuracy achieved in dimension estimation of the crack is 91.88%.

Keywords Crack detection · Accuracy · Image processing · Dimension estimation

1 Introduction

Crack on the surface is common which show degeneration of the surface. These, when unnoticed, lead to greater damage. The most common method used is human inspection. This manual inspection is time consuming and inaccurate. In present days, automatic method of crack detection and estimation is becoming popular. In [1], the authors proposed an image processing approach for crack detecting on surface. The authors have used an image enhancement algorithm called M2 GLD over Otsu method. The proposed algorithm analyzes the crack characteristics like area, width, and parameter. They claim that the proposed method gives improved results than Otsu method. In developing countries, the crack detection in building surface is done manually. This takes time and human effort to measure the dimension and depth of

K. Shreyank (✉) · K. Yukta · N. Sowmya · M. Komal · V. S. Saroja · S. Suhas
KLE Technological University, Vidyanagar, Hubballi, India

the crack [2]. The human inspection, as it depends on the skill of the inspectors, it is not efficient both in terms of accuracy and cost [3]. The present-day automatic method of crack detection is replacing the manual method. Image processing methods are being applied for increasing the productivity of crack detection [4–6]. These automatic methods are becoming a vital part of crack diagnosis which can decide the prevention method and avoid failures [7]. In [8], the authors present a systematic view on the research works in the field of image-based detection and their performance. Two types of methods are described, viz image processing algorithms and percolation model. The paper reviews the famous work by Ito et al. which uses a CCD camera for a high-resolution scan. The paper also describes a percolation model where the brightness of the neighboring pixels is considered to detect a crack. The authors also describe targeting algorithms that are used to estimate the thickness of the crack. The authors in [9] present a collective survey of crack detection done using different techniques. A UAV consisting of a camera makes use of the principle of electro luminescence and time of flight diffraction for video image processing. The surface is the main objective in parameter analysis and uses real data set for crack detection. An improved process is used in post-image processing. A critical examination is done on accuracy level and error level. The paper [10] represents automatic crack detection based on image processing techniques. It includes anisotropic diffusion, wavelet de-noising, SF filtering, and projection integrals. The paper suggests using SVM and machine learning algorithms for classification. This paper also includes some edge-based crack detection techniques like stroke width transform (SWT) and crack width transform (CWT). They also describe recognizing the image texture and width, length, and depth of the crack. The rest of the paper is organized as follows. In Sect. 2, novel crack detection algorithm is discussed. Section 3 deals with the system design for the algorithm. Section 4 describes the experimental results obtained and finally conclusion in Sect. 5.

2 Proposed Novel Crack Detection System

For monitoring the cracks, image processing systems have benefits over conventional approach. The image-based approach monitors the propagation of crack on the surface. There are many remote sensing techniques which allow to measure the crack and can store the observation for any period of time for further analysis. In this paper, the authors use imaging system for identifying the cracks and to estimate the dimension of the cracks. The work focuses to automate the process of crack detection. The structure designer's user should be alerted remotely whenever a crack is detected. The remotely connected mobile displays the result of the crack detector. It also displays the width and height of the crack detected, so that the structure designers can distinguish the seriousness of the crack. The designer's phone can remotely connect using the real VNC application. There is no limitation in the range of the display. The detection of the cracks makes use of edge detection algorithm as the cracks are darker than the background. In image space, they look like edge. The

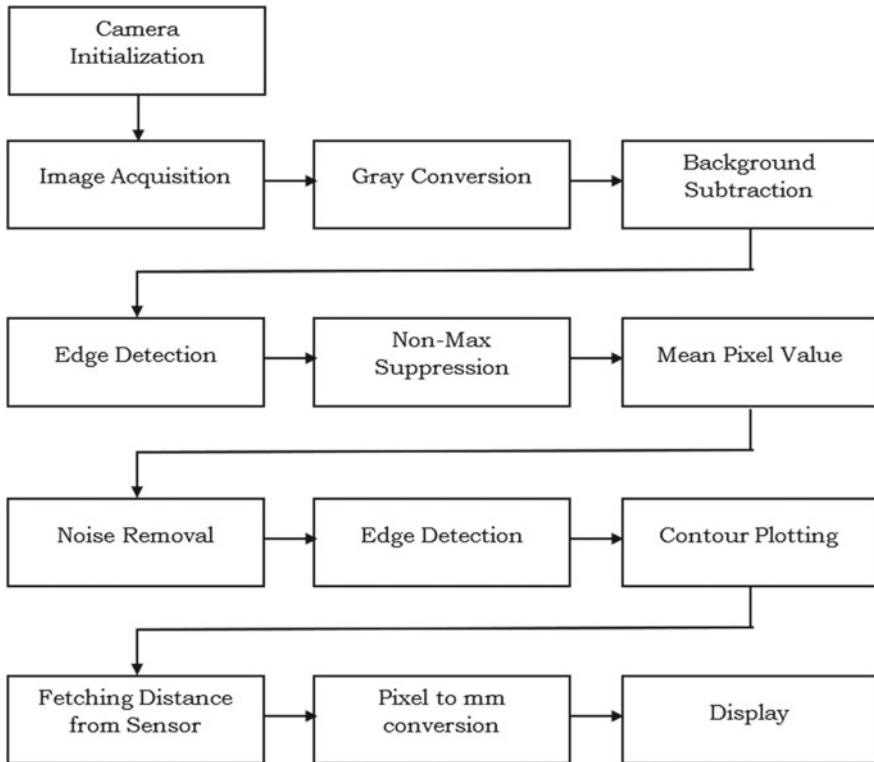


Fig. 1 Functional block diagram crack detection and estimation

algorithm used for crack detection is shown below. Figure 1 shows flowchart for the crack detection.

Algorithm 1: Algorithm for crack detection.

Input: Frame with resolution step.

Output: Display crack with dimension on original frame.

1. Capture a frame by specifying the resolution
2. Color image → Gray image
3. Enhance the feature by Gaussian blurring

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

4. Find the edges-Sobel detection

$$|G| = \sqrt{G_x^2 + G_y^2}$$

$$\hat{G} = \arctan\left(\frac{G_y}{G_x}\right)$$

5. Threshold the image

$$g(x, y) = \begin{cases} 0, & f(x, y) < T \\ 1, & f(x, y) \geq T \end{cases}$$

6. Non-max suppression to enhance the edge detection.
 7. If ($T > 1$) then step 8; else loop to step 1
 8. Contour plot-highlight crack.

3 System Design

Each functional block pertains to the sequence of operations that provides the required functionality. Figure 1 shows the functional block of crack detection and dimension estimation. The works aims to automate the detection of crack and intimate the responsible person on the device the crack detected and its depth, area, etc. There are three main functional blocks in the design namely.

- Initialization block
- Detection block
- Crack processing block.

3.1 Initialization Block

The hardware components off board used in the design need to be initialized for the parameters. The camera module is initialized with the resolution and number of frames per-second. The HC-SR04 sensor module is initialized with the PINs to which it is connected. This block is functional only once at the beginning of the execution of the program. All the modules are initialized as shown below.

- Set resolution = (480×288)
- Set number of frames = 30
- Capture the frame
- Import sensor from GPIOZERO
- Declare pins 23, 24
- Fetching distance by echo.

3.2 Detection Block

When hovered over the surface, the frames are captured and processed to check for the crack. The detection block is responsible to decide the existence of crack. This also acts as a decision block. The captured image is converted to grayscale. The crack feature is enhanced by the background removal process. Gaussian blur is the technique used to enhance the crack feature. After this, the image is subjected to Sobel-edge detection. To get sharper edges, the results of edge detection are filtered using non-max suppression. The average pixel value is computed which is compared with the threshold value. If the average pixel value is greater than the threshold, implies a crack has been found.

3.3 Crack Processing Block

If the implication of the decision block is positive (crack found), then the control is given to this block. The noise present along with the crack is to be removed to get accurate results in the further steps. Thus, a blob detector is parameterized with the values that can detect a noise. The detected blobs are removed by filling the blob with the background color. To highlight the crack, the contour is plotted along the crack outline. Minimum area rectangles are fitted across the span of the crack. It could be assumed that the dimensions are nearly equal to the crack's dimension. The pixel dimensions of these rectangles are converted to mm/cm using the conversion model.

3.4 Conversion Modeling

The conversion model requires the distance as input to compute, and thus, the reading from the HC-SR04 sensor is fetched. A camera has properties like focal length and pixel pitch that can directly influence the size of image formation. Experimentally, for a particular pixel dimension, the focal length is inversely proportional to the actual size, whereas pixel pitch and distance of the object from the camera are directly proportional. At this stage, the equation for actual size is:

$$\text{Actual Size} = (p \times d \times pp)/2.1 \quad (1)$$

where p -pixel coverage or pixel dimension of a bounding rectangle, d -distance of the surface from camera pp -pixel pitch of the camera, and f -focal length of the camera. It can be noted that pixel pitch and focal length are fixed constants for a given camera lens. From the datasheet of the Raspberry Pi camera module $pp = 0.00112$ mm and $f = 3.04$ mm, using these values and distance obtained from the sensor, pixels are converted to mm or cm.

4 Experimental Results and Discussion

The experimental setup is as shown in Fig. 2. Experiment is conducted on the live crack on the wall surface. A snapshot of the live crack is considered as a sample image. Table 1 shows the design specification. Based on the specification given in Table 1, the PoC module is designed. The PoC was tested for various cracks on the wall surface. Table 2 gives the comparison of experimental setup for different data. As discussed, a real crack is taken a sample image. The steps are shown in Fig. 3. This sample image is subjected to thresholding as shown in Fig. 3a. The threshold value is not fixed for the application.

It is considered as 6.4 times average of magnitude. Fig. 3b shows the result of inversion. The dark connected are blobs, these are filtered using OpenCV. The blob detector is parameterized. The detected blobs are removed by filling the blob with the background color, and noise is removed. The crack feature is enhanced by the background removal process. Gaussian blur is the technique used to enhance the crack feature. After this, the image is subjected to Sobel-edge detection. To get sharper edges, the results of edge detection are filtered using non-max suppression as in Fig. 3c. The contour plot and dimension are displayed in Fig. 3d. Table 2 shows the

Fig. 2 Experimental setup
for crack detection



Table 1 Design specification

1	Raspberry Pi 3b+ platform	Raw voltage input	5 V, 2A
		Max current through I/O pins	16 mA
		Flash memory	16 GB SSD
		Internal RAM	1 GB DDR2
		Processor	Broadcom BCM2837 64bit quad core processor
		Operating temperature	-40–85 degree centigrade
2	Raspberry Pi camera module	Sensor	Sony IMX219
		Maximum resolution	3280 × 2464
		Focal length	3.04 mm
3	HC-SR04 sensor	HC-SR04 sensor	5 V
		Measuring distance	2–80 cm
		Current	15 mA
		Frequency	40 Hz
4	Design specifications	Execution time	Average = 0.9653 s Maximum = 1.2130 s Minimum = 0.4952 s
		Minimum width of the crack	3 mm
		Distance from crack	10–60 cm Ideally 25 cm
		Angle of vision	45°
		Accuracy of estimations	95%

Table 2 Comparison experimental data with real data

Trial	Actual width (mm)	Estimated width (mm)	Error (%)
1	5	6	10
2	10	9.9	1
3	7	7.5	7.14
4	6	6.4	6.67
5	8	8.5	6.25

comparison of experiment data with real data. The average error obtained is 8.12%. This shows that the experimental results show 91.88% accuracy.

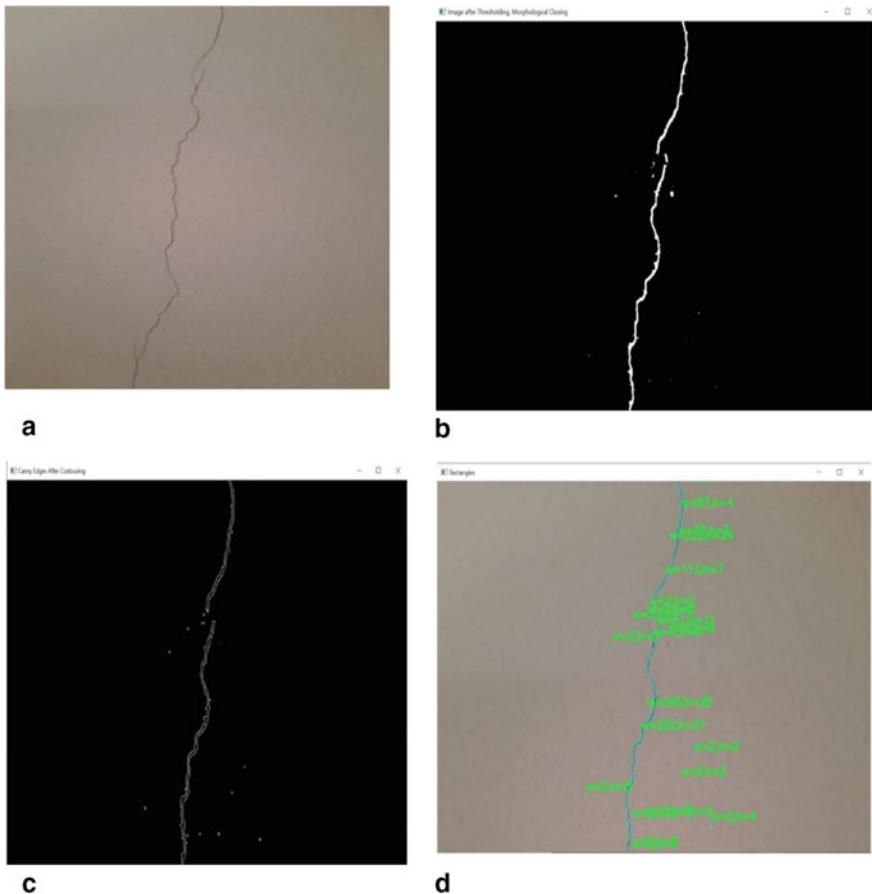


Fig. 3 Steps for detecting crack. **a** Thresholding. **b** Result of inversion. **c** Sharper edges. **d** Contour plot

5 Conclusion

The authors have proposed an algorithm for crack detection and dimension estimation. Experiments are done on the live crack on the wall surface to validate the algorithm. Proof of concept is designed based on the specification. The system focuses on automating the process using established digital image processing techniques. It is a standalone embedded device that detects the surface cracks through images captured via camera and estimates the size of the crack with proper calibration. The experimental results show that the mean error is 8.12%, and the accuracy achieved in dimension estimation of the crack is 91.88%.

6 Future Scope

In the future, the experiment on various other surfaces like glass and metal. This design can be easily re-implemented for other surfaces by changing certain design parameters and thresholds. Many refinements and improvements can also be done on the initial design.

References

1. Hoang N-D (2018) Detection of surface crack in building structures using image processing technique with an improved Otsu method for image thresholding. *Adv Civil Eng* 2018:1–10
2. Koch C, Georgieva K, Kasireddy V, Akinci B, Fieguth P (2015) A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv Eng Inform* 29(2):196–210
3. Lee BY, Kim YY, Yi S-T, Kim J-K (2013) Automated image processing technique for detecting and analysing concrete surface cracks. *Struct Infrastruct Eng* 9(6):567–577
4. Kim H, Ahn E, Cho S, Shin M, Sim S-H (2017) Comparative analysis of image binarization methods for crack identification in concrete structures. *Cem Concr Res* 99:53–61
5. Zakeri H, Nejad FM, Fahimifar A (2016) Image based techniques for crack detection, classification and quantification in asphalt pavement: a review. *Arch Comput Methods Eng* 24(4):935–977
6. Mohan, Poobal S (2017) Crack detection using image processing: a critical review and analysis. *Alexandria Eng J*
7. Rabah M, Elhattab A, Fayad A (2013) Automatic concrete cracks detection and mapping of terrestrial laser scan data. *NRIAG J Astron Geophys* 2(2):250–255
8. Shahrokhinasab E, Hosseinzadeh N, Monir Abbasi A, Torkaman S (2020) Performance of image-based crack detection systems in concrete structures. *J Soft Comput Civil Eng* 1(1):127–129
9. Mohan A, Poobal S (2018) Crack detection using image processing: a critical review and analysis. *Alexandria Eng J* 57(2):787, 798. ISSN 1110-0168. <https://doi.org/10.1016/j.aej.2017.01.020>
10. Gupta D, Goel G, Singla N (2020) Pixel-level crack detection using image processing techniques and machine learning algorithms. *Jcr* 7(17):2649–2651. <https://doi.org/10.31838/jcr.07.17.329>

FDTD Method-Based One-Dimensional Interaction of Electromagnetic Wave with Skin Tissue



Danvir Mandal and Indu Bala

Abstract In this work, the interaction of electromagnetic (EM) wave with skin tissue using FDTD method is presented. The frequency of the electromagnetic wave is kept at 900 MHz and 1800 MHz. The interaction is presented for one-dimensional case. Absolute boundary conditions have been used to minimize the reflections from the edges of one-dimensional space. The amount of the electric field part of the EM wave in the free space and inside the skin tissue has also been illustrated. The interaction revealed that the electric field part of the EM wave is ultimately absorbed inside the skin tissue after traveling some distance. This interaction may be further carried out to reduce the specific absorption rate of any antenna.

Keywords FDTD method · Electromagnetic wave · Skin · One dimensional · Interaction

1 Introduction

In recent few decades, finite difference time domain (FDTD) scheme had been tried by many researchers and scholars for the analysis of antennas, waveguides, and other circuits. The vast ability of the FDTD method enables the researchers to use it for different applications [1–4]. One of the applications of FDTD method is that it can be used to analyze electromagnetic wave interaction with free space and or any other media [5, 6].

D. Mandal · I. Bala
Lovely Professional University, Phagwara, Punjab, India

2 Literature Survey

The basic idea of finite difference equations in time domain and Maxwell equations are introduced in [1, 2] and [7]. The actual application of these equations for electromagnetic simulation of different circuits are illustrated and well explained in [3]. Further, the FDTD scheme is inclusively employed to analyze the antenna performance in [3] and [4]. In [5], various examples of finite difference equations which are used for simulation of electromagnetic waves interaction with different media are presented. The use of boundary conditions in the implementation of the FDTD equations to minimize the reflections is elaborated and illustrated in [6].

In this work, we have presented the EM wave interaction with the skin tissue at 900 MHz and 1800 MHz, respectively. Only one-dimensional scheme has been implemented, and results are presented. The electromagnetic wave is eventually absorbed in the skin tissue after traveling some distance within it. The maximum value of the E_x in the skin tissue and the amount of distance traveled by electromagnetic wave inside skin tissue till complete absorption is also computed and presented in this work. This paper has been organized in six sections. The FDTD expressions for one-dimensional problem space are illustrated in Sect 3. The results of the simulation are presented and discussed in Sect. 4. The conclusion of the simulation and the future scope along with summary of the proposed work are illustrated in Sects. 5 and 6, respectively.

3 Important Expressions for Interaction

The expressions for electric and magnetic fields in free space for one-dimensional case are expressed in Eqs. (1) and (2) as [5],

$$\frac{\partial E_x}{\partial t} = -\frac{1}{\varepsilon_0} \frac{\partial H_y}{\partial z} \quad (1)$$

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0} \frac{\partial E_x}{\partial z} \quad (2)$$

The FDTD expressions for the above one-dimensional cases [5] are illustrated in the following Eqs. (3) and (4),

$$\frac{E_x^{n+0.5}(k) - E_x^{n-0.5}(k)}{\Delta t} = -\frac{1}{\varepsilon_0} \frac{H_y^n(k + 0.5) - H_y^n(k - 0.5)}{\Delta x} \quad (3)$$

$$\frac{H_y^{n+1}(k + 0.5) - H_y^n(k + 0.5)}{\Delta t} = -\frac{1}{\mu_0} \frac{E_x^{n+0.5}(k + 1) - E_x^{n+0.5}(k)}{\Delta x} \quad (4)$$

Table 1 Specific values of cell size, time step, permittivity [8], and conductivity [8]

Frequency	Cell size	Time step	Permittivity (skin)	Conductivity (skin)
900 MHz	0.5180	8.6333e-012	41.405334	0.866780
1800 MHz	0.2673	4.4550e-012	38.871857	1.184768

Now, if we consider a dielectric medium with permittivity (ϵ_r) and conductivity (σ) values, the modified equation for FDTD method [5] is expressed as equation given in (5) and (6),

$$E_x^{n+0.5}(k) = \frac{\left(1 - \frac{\Delta t \cdot \sigma}{2\epsilon_r \epsilon_0}\right)}{\left(1 + \frac{\Delta t \cdot \sigma}{2\epsilon_r \epsilon_0}\right)} E_x^{n-0.5}(k) - \frac{0.5}{\epsilon_r \cdot \left(1 + \frac{\Delta t \cdot \sigma}{2\epsilon_r \epsilon_0}\right)} [H_y^n(k + 0.5) - H_y^n(k - 0.5)] \quad (5)$$

$$H_y^{n+1}(k + 0.5) = H_y^n(k + 0.5) - \frac{1}{2} [E_x^{n+0.5}(k + 1) - E_x^{n+0.5}(k)] \quad (6)$$

All the above expressions with complete explanation are expressed in [5]. The cell size in centimeters, time step in seconds, permittivity, and conductivity values in Siemens/meter for dry skin at 900 and 1800 MHz are presented in Table 1.

4 Results of Simulation

To present the interaction of the electromagnetic wave with skin at 900 and 1800 MHz, 250 FDTD cells are considered in z-direction. The electromagnetic waveforms in free space for the two above-mentioned frequencies after 1000-time steps are illustrated in Fig. 1. Since the cell size is half for 1800 MHz as compared to the cell size for 900 MHz, the waveform frequency is looking same in Fig. 1, but it is not. The source in the simulation is inserted at cell number 10.

For the interaction of EM wave with skin tissue, first 125 cells are considered in free space, whereas next 125 cells are considered in skin tissue. The interaction of electromagnetic wave with skin tissue at 900 and 1800 MHz is presented in Figs. 2 and 3, respectively.

The complete electromagnetic waves during their travel in skin tissue cells at 900 and 1800 MHz are illustrated in Figs. 4 and 5, respectively.

It is observed that at 1800 MHz, after traveling 110 cells, the electric field is completely absorbed in the skin tissue.

However, at 900 MHz, the electric field was completely absorbed after 104 cells. The distance traveled by the electric field in skin tissue before complete absorption

Fig. 1 Simulation results of EM waves in free space

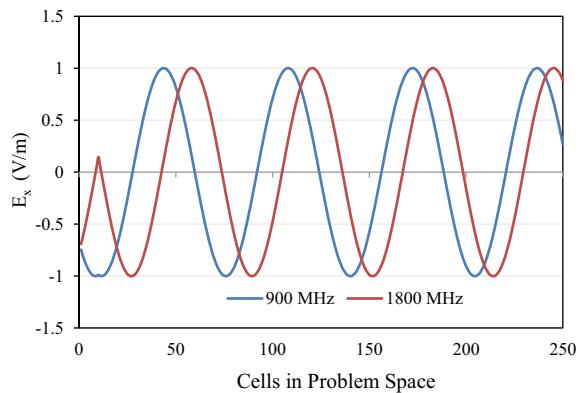


Fig. 2 Simulation results of EM wave interaction with skin tissue at 900 MHz

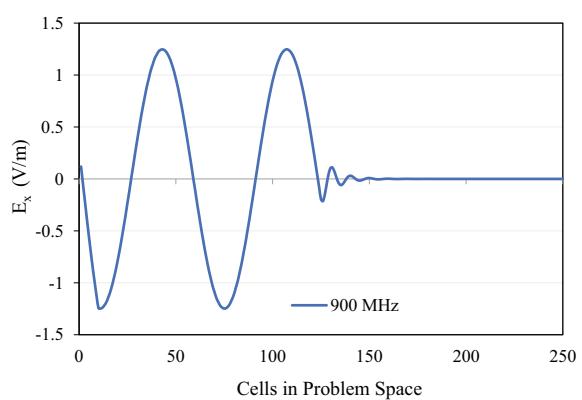
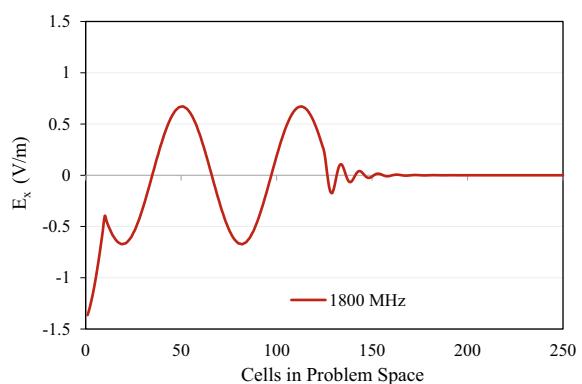


Fig. 3 Simulation results of EM wave interaction with skin tissue at 1800 MHz



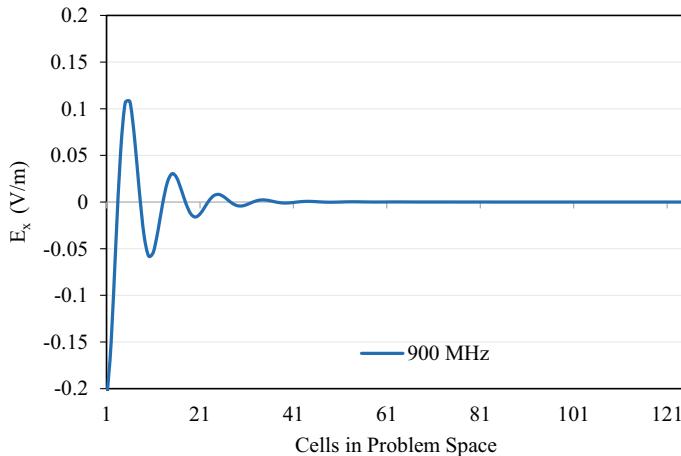


Fig. 4 Electromagnetic waves inside skin tissue at 900 MHz

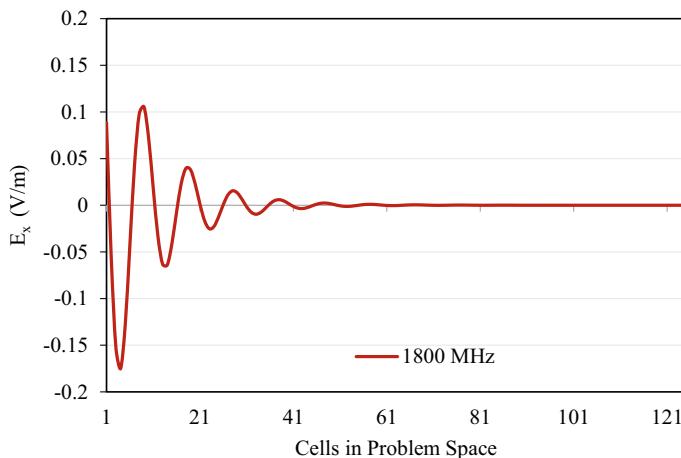


Fig. 5 Electromagnetic waves inside skin tissue at 1800 MHz

is about 53.87 cm for 900 MHz. Similarly, at 1800 MHz, the electric field traveled a distance of about 26.73 cm before the complete absorption by skin tissue.

5 Conclusion

The interaction of the electromagnetic wave with skin tissue is presented in this work. It is observed that the electromagnetic wave is eventually absorbed by the skin tissue layer. The distance traveled by the electric field inside the skin tissue is

greater for 900 MHz as compared to 1800 MHz. This study revealed that if some kind of material with high permittivity value and low conductivity is used as cover for protection from electromagnetic waves, then its use will further reduce the specific absorption rate of any antenna.

6 Future Scope and Summary of the Paper

The proposed work revealed that if an electromagnetic wave is traveling with in a material with higher value of permittivity and correspondingly lower value of conductivity, then the material can absorb the wave, and effective specific absorption rate value will be reduced. In future work, intense research can be done to design such wearable materials which can absorb the electromagnetic waves and reduce specific absorption rate values.

References

1. Yee KS (1966) Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans Antennas Propag* 17:585–589
2. Taflove A, Brodin M (1975) Numerical solution of steady state electromagnetic scattering problems using the time-dependent Maxwell's equations. *IEEE Trans Microw Theory Tech* 23:623–730
3. Sheen DM, Ali SM, Abouzahra MD, Kong JA (1990) Application of the three-dimensional finite-difference time-domain method to the analysis of planar microstrip circuits. *IEEE Trans Microw Theory Tech* 38(7):849–857
4. Reineix A, Jecko B (1989) Analysis of microstrip patch antennas using finite difference time domain method. *IEEE Trans Antenna Propag* 37(11):1361–1369
5. Sullivan DM (2000) Electromagnetic simulation using the FDTD method. IEEE Press
6. Mur G (1981) Absorbing boundary conditions for the finite difference approximation of the time domain electromagnetic field equations. *IEEE Trans Electromagn Compat* 23:377–384
7. Cheng DK (2006) Field and wave electromagnetics. Pearson Education Asia Limited and Tsinghua University Press
8. FCC: Body tissue dielectric parameters. <http://www.fcc.gov/oet/rfsafety/dielectric.html>. Last accessed 27 June 2021

Performance Enhancement of UAV-Based Cognitive Radio Network



Indu Bala , Danvir Mandal , and Ankur Singh

Abstract To conquer the existing spectrum shortage issues for the successful deployment of new wireless applications, cognitive radio has evolved as a burgeoning strategy for wireless networks. However, the technology still has some bottlenecks such as fading, heterogeneous operating conditions, and sensing errors, etc., due to which its full potential cannot be exploited. Recently, unmanned aerial vehicles (UAVs) are also gaining momentum in many communication paradigms due to their high mobility and flexibility. The ability to form a flying network makes UAV technology the most suitable candidate to address the challenges like coverage and on-demand network deployment, posed by beyond 5G (B5G) and 6G networks. In this paper, the accomplishment of a UAV-based cognitive radio network system is investigated. The proposed system considered line-of-sight conditions between the licensed primary user and UAV secondary users to sense the channel, and the transmission mode diversity is used to enhance the throughput of the secondary user. Simulation results are presented to corroborate the proposed scheme. Moreover, the comparison results are also presented to corroborate the effectiveness of the proposed scheme.

Keywords Cognitive radio · Sensing radian · Unmanned aerial vehicles · Throughput · Diversity

1 Introduction

Unmanned aerial vehicles (UAVs) are gaining momentum in many communication paradigms due to their high mobility and flexibility [1]. The ability to form a flying network makes UAV technology the most suitable candidate to address the challenges like coverage and on-demand network deployment, posed by beyond 5G

I. Bala · D. Mandal
Lovely Professional University, Phagwara, Punjab, India

A. Singh
Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

(B5G) and 6G networks. Earlier, the military was using UAVs for surveillance and tracking operations [2, 3], but now with the advancement in sensor technology, they are equipped with a variety of sensors and communication capabilities to perform complicated tasks where human intervention could be dangerous. In comparison to the other terrestrial communication paradigms, UAV communication is immune to path loss during transmission and thus has the potential to achieve high throughput [4]. However, there are many challenges to UAV communication that needs an open discussion. UAVs, in general, operate on IEEE S-band (2–4 GHz), L-band (1–2 GHz), and industrial, scientific, and medical (ISM) frequency bands (2.4 GHz) [5]. With the rapid deployment of 5G networks and associated applications in the market, these bands are becoming overcrowded, and therefore, there exists a spectrum shortage to deploy these UAV-based communication systems [6].

Recently, cognitive radio (CR) technology has materialized as a potential 5G technology to overcome the spectrum crunching problem while meeting the critical spectrum requirements of 5G networks and future IoT applications [1]. It defines an intelligent radio that constantly monitors the surrounding RF environment with the capability to adapt its transmission characteristic accordingly to meet end user data rate requirements. Considering the benefits of both, UAV and CR technologies, the combination can overcome spectrum scarcity issues while improving the overall channel capacity [7]. The technology allows dynamic spectrum access by allowing unlicensed secondary users (SUs) to transmit their data over the licensed channels of incumbent primary users (PU) on finding them vacant. Various communication paradigms have been investigated by various researchers for the throughput enhancement of CR networks through dynamic spectrum access which mainly falls into three main categories, namely underlay, overlay, and interweave communication paradigm [8]. The underlay communication permits licensed and unlicensed users to share the same channel for their data transmission without exceeding the interference limit of PU. In the overlay scheme, two main approaches are used [4, 9], namely (i) channel coding which divides SUs transmission power PU as well as SU data without exceeding the interference constraints at the PU receiver side [4], and (ii) network coding that encodes SU data packet onto the PU packet and relay onto the licensed channel [12]. The interweave approach, on the other hand, allows SU to sense the licensed spectrum and transmission on detecting PU absence in the channel [4, 10–12]. The effectiveness of the CR network depends upon the correctness of the sensing outcomes which is measured in terms of probability of detection and probability of false alarm. For efficient spectrum utilization, the probability of detection must be high and the probability of false alarm should be low. Significant research work is available to enhance the sensing accuracy such as the cooperative spectrum sensing schemes with different fusion rules proposed in [13–15]. Nowadays, machine learning approaches are used by many researchers to enhance sensing accuracy [16]. Since the wireless channels are highly unpredictable, it is impossible to have perfect sensing in realistic wireless fading scenarios. The performance of the sensing method depends on the sensing duration. For a given frame, the long sensing durations reduce the transmission duration time and therefore SU throughput [4, 12]. Thus, the sensing time is an important parameter that decides the throughput

of the secondary user. Since UAVs have better signal reception under severe fading and shadowing also, UAVs can yield better sensing results as compared to ground spectrum sensing due to line-of-sight (LOS) communication.

In most of the literature reviewed above, what they fail to take into consideration is the diversity of the transmission modes in CRNs. Thus, while addressing the spectrum shortage problem, a UAV-based CRN is considered in this paper which optimizes the unlicensed user's throughput by exploiting transmission mode diversity.

In this paper, the analytical model for UAV-based CR is developed for hybrid spectrum access, and the analytical model is developed for the throughput of UAV-based CR considering various scenarios of PU availability in a channel. The rest of the paper is structured as follows: The proposed UAV-based communication scenario is explained and analyzed in Sect. 2. In Sect. 3, numerically simulated results are presented, and then, the conclusion and future scope are discussed in Sect. 4.

2 Proposed Communication System Model

A typical UAV-based spectrum sharing communication system is considered, in which the UAV is acting as a secondary user (SU) and making continuous periodic circular flight around a primary user (PU) with flight speed v as given in [17]. The whole flight duration is alienated into (i) sensing radian and (ii) transmission radian. The energy detector is mounted on a UAV to sense the PU availability in a channel [18]. Unlike the conventional scheme, in which PU transmits the data on detecting PU absent from the channel, in the proposed scheme, the UAV transmits the data when PU is detected absent from the channel in overlay mode. However, on detecting PU present in a channel, the data is transmitted over the channel using interweave approach, as shown in Fig. 1.

In a conventional scheme, SU transmits the data on detecting PU absent from the channel. Based on the sensing outcomes, the data transmission scenarios and their respective secondary throughputs are as follows:

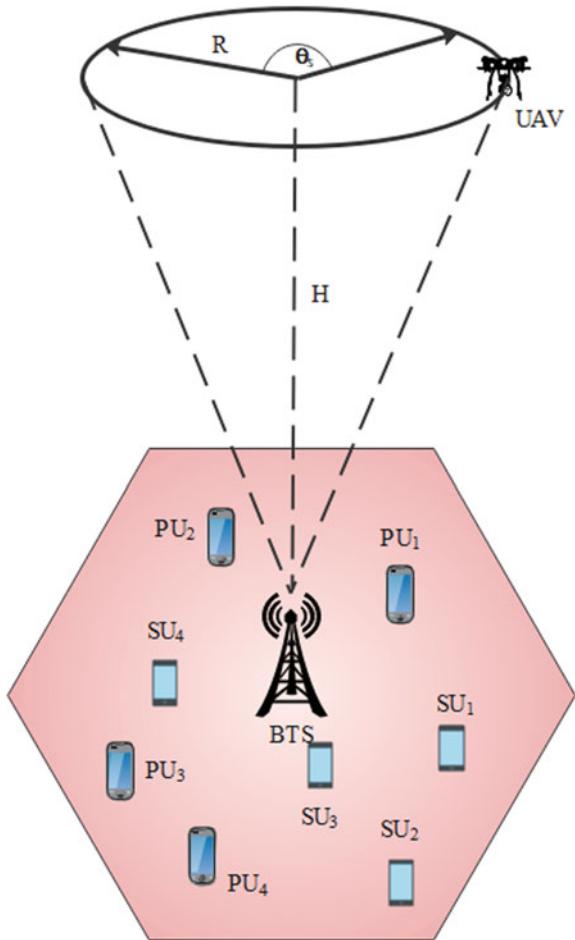
Scenario I: $P(H0/H0)$: In this scenario, CR accurately detect the PU absence from the channel with detection probability $P(H0)(1 - P_f)$. The SU throughput, in this case, is given by

$$R_{H0/H0}(\theta_s) = \frac{2\pi - \theta_s}{2\pi} P(H0)(1 - P_f) \log\left(1 + \frac{P_1 h_{pu}^2}{\sigma_m^2}\right) \quad (1)$$

Scenario II: $P(H0/H1)$: In this scenario, CR missed the detection of PU with probability $P(H1)(1 - P_d)$. The SU throughput, in this case, is given by

$$R_{H0/H1}(\theta_s) = \frac{2\pi - \theta_s}{2\pi} P(H0)(1 - P_d) \log\left(1 + \frac{P_1 h_{pu}^2}{\sigma_m^2 (1 + \text{SNR}_p)}\right) \quad (2)$$

Fig. 1 Proposed UAV-based secondary communication system



where P_1 represents the UAV transmission power.

In the proposed communication scenario, UAV senses the PU channel and transmits with transmission power P_1 using an overlay scheme on detecting channel absent from PU. However, it transmits in interweave mode with transmission power P_2 , such that $P_2 < P_1$, on detecting PU present in a channel. Thus, the scheme exploits every transmission opportunity under these scenarios:

Scenario I: $P(H1/H1)$: In this scenario, UAV detects PU accurately under the hypothesis $H1$ and starts transmitting data in interweave mode with transmission power P_2 with probability $P(H1)(P_d)$. The SU throughput, in this case, is given by

$$R_{H1/H1}(\theta_s) = \frac{2\pi - \theta_s}{2\pi} P(H1)(P_d) \log \left(1 + \frac{P_2 h_{p-u}^2}{\sigma_m^2 (1 + \text{SNR}_p)} \right) \quad (3)$$

Scenario II: (H_0/H_1): In this scenario, UAV is unable to detect PU in a channel under H_1 , and thus, starts transmitting data with transmission power P_1 in overlay mode with probability $P(H_1)(1 - P_d)$. The SU throughput, in this case, is given by

$$R_{H_0/H_1}(\theta_s) = \frac{2\pi - \theta_s}{2\pi} P(H_1)(1 - P_d) \log\left(1 + \frac{P_1 h_{p_u}^2}{\sigma_m^2 (1 + \text{SNR}_p)}\right) \quad (4)$$

Scenario III: (H_1/H_0): In this scenario, UAV detects PU in a channel under H_0 , and thus, starts transmitting data in with transmission power P_2 with probability $P(H_0)(P_f)$. The SU throughput, in this case, is given by

$$R_{H_1/H_0}(\theta_s) = \frac{2\pi - \theta_s}{2\pi} P(H_0)(P_f) \log\left(1 + \frac{P_2 h_{p_u}^2}{\sigma_m^2}\right) \quad (5)$$

Scenario IV: (H_0/H_0): In this scenario, UAV generates no false alarm under H_0 , and thus, starts transmitting data in with transmission power P_1 . Thus, with probability $P(H_0)(1 - P_f)$, SU throughput, in this case, is given by

$$R_{H_1/H_0}(\theta_s) = \frac{2\pi - \theta_s}{2\pi} P(H_0)(1 - P_f) \log\left(1 + \frac{P_1 h_{p_u}^2}{\sigma_m^2}\right) \quad (6)$$

Thus, the goal is to get the most out of the UAV-based SU's throughput by optimizing sensing radian and by exploiting the transmission mode diversity. Thus, the throughput optimization problem can be articulated as

$$\max_{\theta_s} R(\theta_s) \quad (7a)$$

$$\text{s. t. } 0 \leq \theta_s \leq 2\pi \quad (7b)$$

3 Results and Discussion

In this segment, the numerical results are presented for the proposed secondary communication system considered in this paper. The simulation parameters considered in this paper are the same as that of the conventional scheme proposed in [17].

The transmission power profile of UAV with respect to the sensing radian for various values of UAV speed is shown in Fig. 2. It can be inferred that transmission power increases when the sensing radian is small and reaches the maximum values and then start decreasing as the sensing radian increases further. The impact

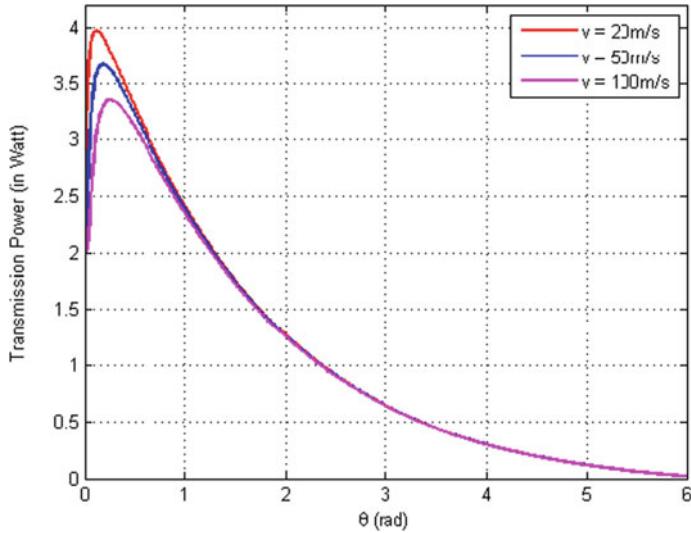


Fig. 2 Transmission power profile with respect to the sensing radian for different UAV velocities

of the UAV speed can also be observed here. As the UAV velocity increases, the communication power decreases.

In this section, we investigated the performance of the proposed hybrid spectrum access scheme. Figure 3 shows the UAV throughput with respect to the sensing

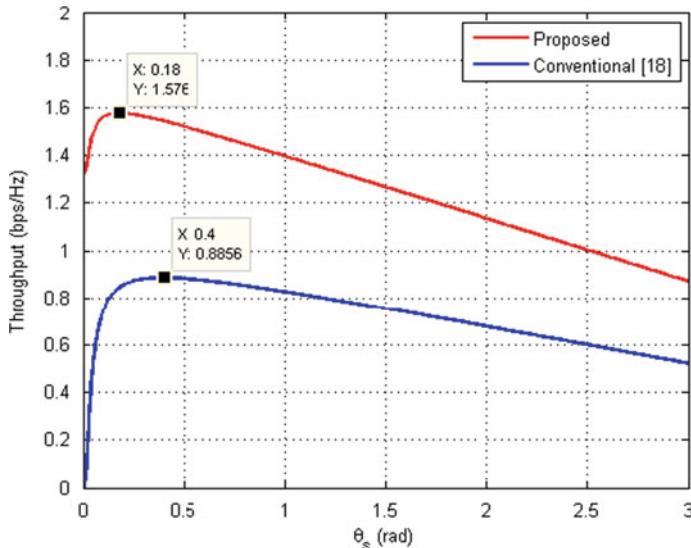


Fig. 3 Comparison of throughput versus sensing radian (θ_s)

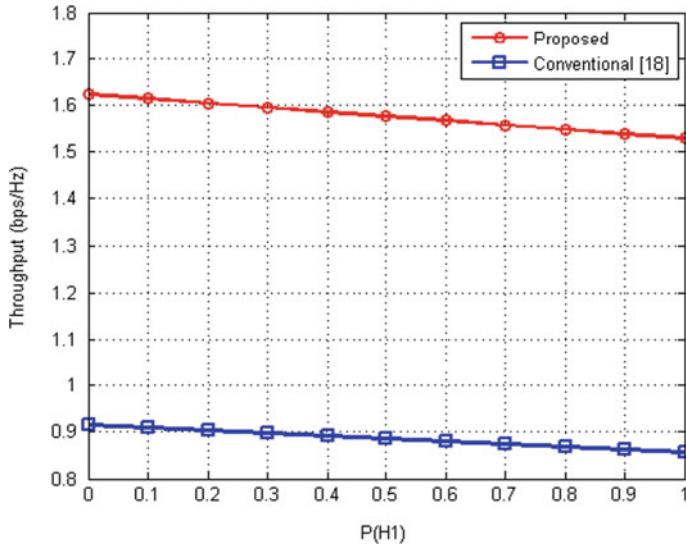


Fig. 4 Throughput versus probability of PU presence in a channel

radian θ_s . It can be inferred that UAV throughput increases, in the beginning, attains maximum value at $\theta_s = 0.18$ radian and then start decreasing. This is due to the reason that the sensing time increases with an increase in θ_s , as a result, transmission time and therefore throughput decrease. Further, it is also evident from the graph that the proposed spectrum sharing scheme outperforms the conventional scheme and achieves significantly high throughput for all values of sensing radian under similar system parameters.

The UAV-based SU throughput with respect to the likelihood of PU presence in a channel is shown in Fig. 4. It is evident from the graph that the throughput of the system decreases as the likelihood of PU presence in a channel increases.

4 Conclusion and Future Scope

The UAV-based spectrum sharing communication system is analyzed in this paper while considering the sensing throughput trade-offs. To maximize the UAV-based CR throughput, the sensing radian is optimized. For the calculated optimum values of the sensing radian, the secondary throughput is maximized further by using transmission mode diversity. Simulated results validate the effectiveness of the proposed scheme. It has been perceived that the proposed scheme outperforms the conventional scheme by switching between two different transmission modes.

The work can be extended in the future by incorporating virtual spectrum sharing to enhance spectrum sensing performance. Moreover, energy harvesting solutions can also be used for efficient power management.

References

1. Santana GMD, Cristo RS, Dezan C, Diguet JP, Osorio DPM, Branco KRLJC (2018) Cognitive radio for UAV communications: opportunities and future challenges. In: 2018 International conference on unmanned aircraft systems, ICUAS 2018, pp 760–768. <https://doi.org/10.1109/ICUAS.2018.8453329>
2. Saleem Y, Rehmani MH, Zeadally S (2015) Integration of cognitive radio technology with unmanned aerial vehicles: issues, opportunities, and future research challenges. *J Netw Comput Appl* 50:15–31. <https://doi.org/10.1016/j.jnca.2014.12.002>
3. Li B, Fei Z, Zhang Y (2019) UAV communications for 5G and beyond: recent advances and future trends 6(2):2241–2263
4. Bala I, Ahuja K, Nayyar A (2021) Hybrid spectrum access strategy for throughput enhancement of cognitive radio network. In: Sharma DK, Son LH, Sharma R, Cengiz K (eds) Microelectronics and telecommunication engineering. Lecture notes in networks and systems, vol 179. Springer, Singapore, pp 105–122. <https://doi.org/10.1007/978-981-33-4687-1>
5. Pan Y, Da X, Hu H, Zhu Z, Xu R, Ni L (2019) Energy-efficiency optimization of UAV-based cognitive radio system. *IEEE Access* 7:155381–155391. <https://doi.org/10.1109/ACCESS.2019.2939616>
6. Zhang H, Da X, Hu H, Ni L, Pan Y, Seo J (2020) Spectrum efficiency optimization for UAV-based cognitive radio network. *Math Probl Eng* 2020. <https://doi.org/10.1155/2020/2497542>
7. Bala I, Bhamrah MS, Singh G (2019) Investigation on outage capacity of spectrum sharing system using CSI and SSI under received power constraints 25(3):1047–1056. <https://doi.org/10.1007/s11276-018-1666-7>
8. Bala I, Bhamrah MS, Singh G (2017) Rate and power optimization under received-power constraints for opportunistic spectrum-sharing communication. *Wirel Pers Commun* 96(4):5667–5685. <https://doi.org/10.1007/s11277-017-4440-8>
9. Bala I, Bhamrah MS, Singh G (2017) Capacity in fading environment based on soft sensing information under spectrum sharing constraints. *Wirel Networks* 23(2). <https://doi.org/10.1007/s11276-015-1172-0>
10. Rana V (2014) Resource allocation models for cognitive radio networks : a study. *Int J Comput Appl* 91(12):51–55
11. Sethi R (2013) Performance evaluation of energy detector for cognitive radio network. *IOSR J Electron Commun Eng* 8(5):46–51. <https://doi.org/10.9790/2834-0854651>
12. Rubeen R, Bala I (2015) Throughput enhancement of cognitive radio networks through improved frame structure. *Int J Comput Appl* 109(14):40–43. <https://doi.org/10.5120/19259-1016>
13. Liu X, Li F, Na Z (2017) Optimal resource allocation in simultaneous cooperative spectrum sensing and energy harvesting for multichannel cognitive radio. *IEEE Access* 5(8):3801–3812. <https://doi.org/10.1109/ACCESS.2017.2677976>
14. Fan L, Zhao R, Gong FK, Yang N, Karagiannidis GK (2017) Secure multiple amplify-and-forward relaying over correlated fading channels. *IEEE Trans Commun* 65(7):2811–2820. <https://doi.org/10.1109/TCOMM.2017.2691712>
15. Liu X, Chen K, Yan J, Na Z (2016) Optimal energy harvesting-based weighted cooperative spectrum sensing in cognitive radio network. *Mob Networks Appl* 21(6):908–919
16. Thilina KM, Choi KW, Saquib N, Hossain E (2013) Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE J Sel Areas Commun* 31(11):2209–2221. <https://doi.org/10.1109/JSAC.2013.131120>

17. Bala I, Ahuja K, Energy efficient framework for cognitive radio networks. *Int J Commun Syst* (forthcoming). <https://doi.org/10.1002/dac.4918>
18. Liu X, Guan M, Zhang X, Ding H (2018) Spectrum sensing optimization in an UAV-based cognitive radio. *IEEE Access* 6(8):44002–44009. <https://doi.org/10.1109/ACCESS.2018.2862424>

Blockchain Technology with Supply Chain Management: Components, Opportunities and Possible Challenges



Ayasha Malik, Abhijit Kumar, Jaya Srivastava, and Bharat Bhushan

Abstract Supply chains have been challenged by giving transparency and confidence between contributors as well as investors for guaranteeing efficiency. Conversely, a few challenges are tough to solve as supply chain records may be exchanged by their contributors. The advent of blockchain is an unbreakable distributed ledger technology and has been acknowledged as an encouraging resolution to these encounters. The paper introduced the blockchain and examined the various components of the newly developed blockchain to uncover various supply chain challenges. The paper emphasises the acquisition of supply chain challenges and their management. In addition, this paper describes the classification of blockchain technology and elements of consensus algorithms. Furthermore, the paper outlines the characteristics of the blockchain structure for supply chain management, sums up the possible solutions for the blockchain for its effective implementation in future supply chain mechanisms and discusses a few challenges along with opportunities left in future research.

Keywords Source · Structure · Transparency · Blockchain · Security · Distributed ledger · Confidentiality · Supply chain management · Consensus algorithms

1 Introduction

The supply chain is a group of successive phases that transfer to design manufactured goods. Every phase could be managed by single or multiple corporations, contractors or investors who take part in the production, distribution, storage or circulation of

The original version of this chapter was revised: The author “J. Srivastava” affiliation has been updated. The correction to this chapter is available at

https://doi.org/10.1007/978-981-16-8721-1_74

A. Malik (✉) · A. Kumar · J. Srivastava
Noida Institute of Engineering Technology (NIET), Greater Noida, India

B. Bhushan
School of Engineering and Technology (SET), Sharda University, Greater Noida, India

the manufactured goods to user [1]. A chain of supply is having a vital role in the universal budget; the impact of supply chain businesses account is more than 82% as of global employment. Big companies are exporting their association lines to cost-effective areas to reduce construction charges. Processes close to the supply chain are further subdivided and hosted by a growing number of partners. Supply chains have converted into universal, multifaceted and co-depend on each stage. Supply chain includes a variety of investors and customers with a wide range of processes across multiple stages. It is not easy to monitor all the procedures, resources and proprietorship at various stages. In addition, supply chain categories are often available in diverse locations and sometimes in different nations. Supply chain congestion poses the management challenges in the management of an active administration of supply chain.

The corporation's purpose to uncover the growing supply chain difficulty by accepting diverse novel types of machineries like barcode, radio-frequency identification (RFID) and a high-frequency direction finder to gather data straight as of supply chain procedures plus phases. Data statistics are additional machinery used continuously in stock administration and demand forecasting [2]. Nowadays, supply chains prerequisite is even much trustable than ever. Disturbance of the supply chain may result in important losses for corporations with shorter and longer terms and raise the cost of end-users. Such a corporation's prerequisite to originate faster resolutions to encounter the vigorously varying needs. Productions and consumers propose new knowledge requirements on manufactured goods such as legitimacy, source, superiority and sustainability. These challenges go hand in hand with contributors with manufactured goods purchases in supply chain. On the other hand, the stored data in the supply chain could be exchanged by the participants of the supply chain, and at that time, the data may not be accessible to consumers. To solve the challenges defined in real estate, data records should be kept unchanged and available. Blockchain is encouraging machinery that satisfies several procurement tasks. Blockchain is a scalable as well as flexible ledger that delivers a reliable data record which never is managed and hacked by a central authority [3]. The motivation of this study is enumerated as follows:

- This work highlights the blockchain technology as well as supply chain management for the better life for people.
- This work emphasised the collaboration of nature of blockchain to the supply chain management to introduce even better facility that can tackle people's problems.
- This work described several components, opportunities and challenges with their solution in context with the two combined termed that is supply chain management and blockchain.

The remainder of the paper is organised as follows. Section 2 introduces supply chain management to accomplish extraordinary competence and effectiveness. Section 3 presents possible challenges that occur in supply chain and its proper management (administration) to reduce difficulties. Section 4 highlights the blockchain technology with its classifications and components of consensus algorithms. Section 5 distinguished the structures that have been projected for proper

administration of supply chains. Section 6 discusses the possible resolutions to the supply chain challenges. Finally, Sect. 7 concludes with conclusions.

2 Supply Chain Management

The chain of supply could stay very composite because this might have a lot of phases, and all the associated people are essential to monitoring the manufactured goods production line by line in each phase. The count of phases can increase according to difficulty of the manufactured goods. In addition, a chain of supply can be a group of multiple integrated chains of supply since few manufactured goods can work as components or synthetic components of another [4].

For example, dealers deliver raw ingredients to dispensation units that produce components of a composite manufactured goods. Those components are accumulated and wrapped by the producer to obtain the finalised manufactured goods. These finalised manufactured goods are delivered by a retailer or supplier. Affected parts can be dispersed to several positions wherever logistics management can be controlled by traders and exporters in national lines. Additional dissemination of manufactured goods is controlled by dealers which brings them ultimately to the consumers. The efficiency of the resources, components and manufactured goods depends on the way they are delivered in the chain also on the price.

Specific purposes of supply chain management are asset administration, warehouse administration, contract administration, transference, secretarial and other purposes. Typically, manufactured goods management permits over the transaction list or contract made between participants. Dealings must be recorded exactly and dependably. Dependability between supply chain groups confirms smooth communications and no outbursts. In addition, for newly familiarised business partners, secondary technology that delivers circumstantial information of the partners intricate can speed up the procedure of construction of such relations [5]. Figure 1 shows the internal flow of supply chain between materials to customers. At first, the raw

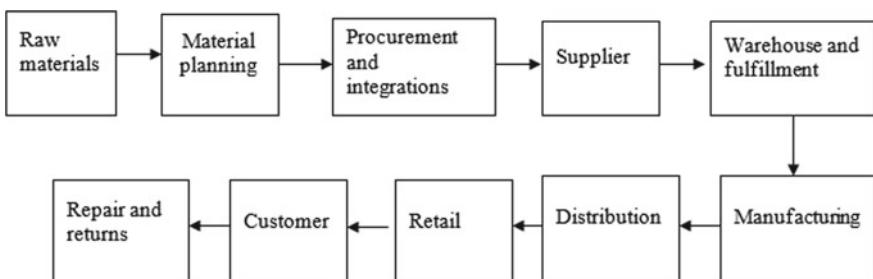


Fig. 1 Flow of supply chain

materials get collected, then it goes for proper planning then procurement and integration performed over it; furthermore, processed materials are handed over to the supplier and stored in the warehouse. After that, it will get handed over to the manufacturing team, then to the distributors, then to the retailers and lastly to the customer, but if any discrepancy found in the materials, then it returns for repairs.

3 Challenges in Supply Chain with Its Management

While the technology is digital and automatic with numerous supply chain management roles, other challenges continue to make transactions well-organised, dependable and protected. This section described the challenges of providing supply chains with the aim of ensuring their competence and faith in their shareholders.

- Source: Data files of ownership over time. The main functions delivered by source are tracking and tracing. It is frequently complicated to trace the resources of construction and the main source of the manufactured goods.
- Performance enhancement: The performance of supply chain could be well-defined by a variety of main pointers, like how much time a manufactured good devotes to every phase of supply chain, manufactured goods' production costs and construction yield [6].
- Quality assurance and quality control: It is acquiescence with a manufactured goods, construction or supply chain of a diversity of quality features imposed by shareholders, consumers or controlling organisations.
- Sustainability: Supportable supply chain should reflect its influence over the surroundings and use ecologically welcoming resources and procedures, in order to decrease greenhouse gas emanations in its phases. The supply chain is accountable for its construction of heat-retaining pollutants and additional impurities in a reliable and impartial manner. Such responses may be helpful in assessing the ecological impact of participants and controllers [7].
- Transparency: It denotes the acquisition and discovery of information about the provision of suppliers, stockholders, buyers and regulatory entities. Openness keeps participants informed of the nature of processes and resources. Transparency in the sales list has revealed that it has an encouraging impression on the status of the corporate.
- Data Secrecy: Sensitive and transactional data like economic accounts need to be accessed only by other shareholders. Examples of this data contain material costs, remunerations and residues. Any other data relating to communications between individuals must be kept trustworthy and authenticated [8].

4 Blockchain

A blockchain is a flexible dispersed incontrovertible ledger, which is used to record and noted the dealings made among dissimilar consumers without going to consolidated and trusted party. Consistency is the energetic force of the blockchain, which strengthens faith between consumers by giving a long-lasting and validated transaction record. The blockchain consists of a peer-to-peer system created by contributor organisations, a dispersed ledger with static data blocks, recorded transactions, smart business contracts and consensus algorithms that determines the applicant for the succeeding block. Blockchain's participation may be a customer, a simple customer request (light nodule) or a mineworker (complete nodule). A blockchain consumer interconnects via a customer nodule, and the customer is termed as contributor, who creates transactions [9].

The block is connected in sequence to a formerly documented block via a hash cursor. The hash identifier covers the hashed details of the preceding block's content and ensures the block order and data integrity. The outcome is a spreadable, immutable distributed ledger. Each data block comprises a certified business group and a metadata header including evidence of block authentication and a hash identifier that directing towards the preceding block. Figure 2 shows the blockchain components.

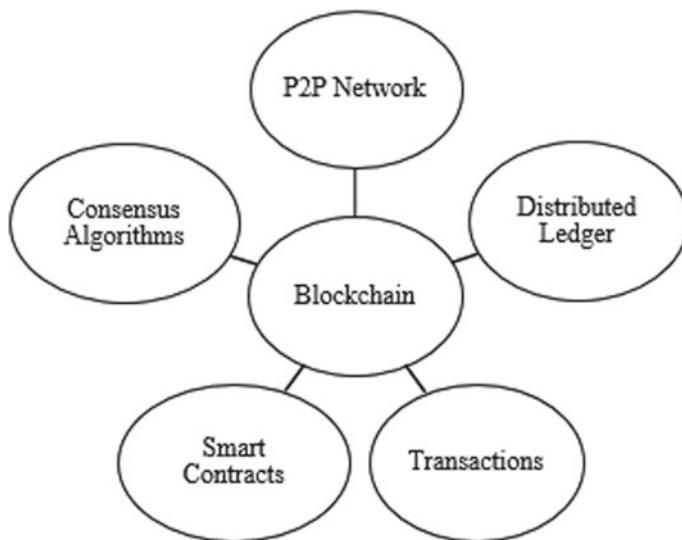


Fig. 2 Components of blockchain

Table 1 Comparison of public blockchain and private blockchain

Public [10]	Private [10]
Openly available to the network as well as consensus algorithms	Limited availability to the network as well as consensus algorithms
Difficult consensus algorithms	Easy consensus algorithms
Less throughput	High throughput
Accessible consensus algorithms	Consensus algorithms having limited accessibility
Enticement procedures required	Not any need for enticement procedures

4.1 Classification of Blockchain

Blockchain can access dissimilar levels of local power sharing and can be divided into public and private properties. Those are discussed below.

- Public property: In this property, the blockchain having the network access and contribution in the consensus algorithm is exposed to person who desires to contribute. Contributor ownership and their replacement at any time, lacking any need to disclose their actual personality. These blockchains use a compatibility procedure that outlines strict guidelines for adopting projected blocks and an incorporated promotion approach that recompenses reliable contribution. Conversely, decentralisation is accomplished through a low business fee, that is, the total number of transactions per second.
- Private property: In this property the blockchain having the network admission along with contribution in the consensus algorithm for sync is limited; contributors of this blockchain are obligatory to authorise before they take part in blockchain. As participants' individualities in this blockchain are recognised by registered members, mischievous actions could be noticed by the Byzantine Fault Tolerant (BFT) algorithm [10]. Table 1 illustrates the differences between public and private blockchains.

4.2 Components of Consensus Algorithms in Blockchain Technology

Consensus algorithms of blockchain describe the fixed group of guidelines for mineworkers or authoriser to approve a mutual reality. Figure 3 presents the consensus algorithms of blockchain that can comprise the succeeding five mechanisms discussed below. However, all five components are not implemented by all blockchain consensus algorithms.

- Block proposal is a procedure where mineworkers or authoriser chooses the subsequent supporter of a block. For safety motives, a block suggestion methods

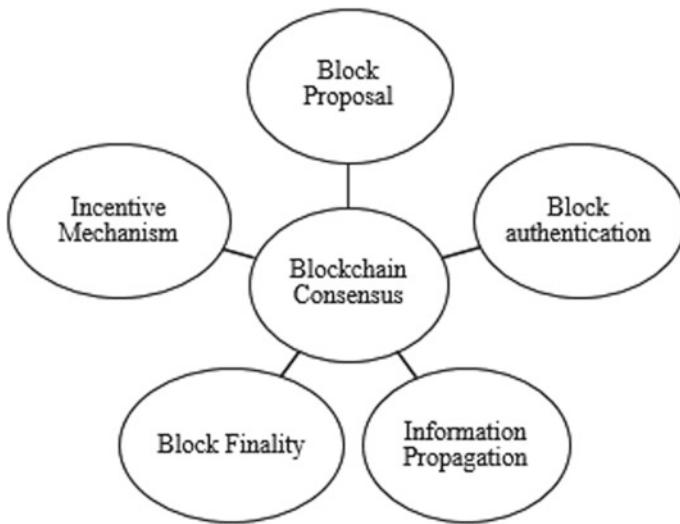


Fig. 3 Mechanisms of consensus algorithm

in permissionless blockchains need the lowest inter-block suggestion period. Furthermore, mineworkers or authoriser are obligated to deliver a resistance of their calculation exertion to receive the correct data to suggest the succeeding block of blockchain.

- Block authentication is the procedure to confirm that acknowledged the blocks that are philologically accurate. Block in blockchain becomes effective after it contains the resolution to a cryptanalysis problem along with the certified digital autographs of the contributors in the communications that are documented in block [11].
- In information propagation, complete nodules transmission blocks to the open system. They are essential to monitor the one-to-many transmission or distribution procedures of a discussed and selected agenda.
- Block finality is the procedure that is used for recording of irreversible block after it has been dedicated to a distributed ledger. As soon as it is acknowledged and confirmed by the participating end, the block is inserted into a local replica of the distributed ledger which is preserved for full nodes.
- Incentive mechanisms are used to avoid irregular or spiteful activity by compensating compliant miners. In a permissionless blockchain, pseudo-anonymous miners are used for data security and extraction. Consequently, mineworkers can come together and lift illegal blocks and make that illegal as well as harmful block inoperable. Another consensus algorithm could use incentive approach to prevent mischievous behaviour by miners. In addition, consensus algorithms may also impose fines on offensive behaviour of mineworkers [12]. Table 2 encapsulates the framework for solution to overcome the challenges that come under products associated with blockchain for supply chain.

Table 2 Blockchain frameworks

Products	Challenges [11, 12]	Framework for solution [11, 12]
Food	Transparency, sustainability, reassurance of food quality with proper administration, enhancement of overall performance, source	Ethereum, Hyperledger, Double blockchain, BigchainDB, Carbon footprint chain
Wood	Source	Ethereum
Entertainment and media	Transparency	TransICE
Pharmacological	Data confidentiality, quality assurance and quality control, Performance improvement, source	Hyperledger, Gcoin, QuarkChain
Postage	Source	Exonum
Auto-motorised	Source	Ethereum
Digital goods	Source	Ethereum
Wine	Source	MultiChain

5 Characteristics of Blockchain Structure for Management of Supply Chain

Blockchain delivers a comprehensive communiqué atmosphere along with safety assurances that designed for the employment of vigorous and affordable use of supply chain management. The main characteristics of blockchain structure for its use in supply chain management.

- Privacy with data verification: Certain blockchain structures permit complex data for transactions before they are integrated into a blockchain ledger, using cryptographic techniques such as hashing, encryption, decryption and digital signature. The blockchain structure could also limit the availability and reflectivity of blockchain content to only certified consumers.
- Lightweight contraction algorithms: Blockchain structures must sustenance lightweight compliance techniques to accomplish extraordinary communications. Supply chains can be essential to maintain huge size of dealings installed by different IoT devices [13].
- Smart decision-making contracts: Employee-led smart contracts are required to be established and preserved the status of one ledger. Smart agreements use a group of procedures that describe the occupational concept of a supply chain management system
- Quick retrieval: Tracing manufactured goods structure as it goes through various phases of its supply chain needs sustenance for data acquisition in a blockchain

approved and acquired by the supply chain. This information are used in corresponding choices or manufactured goods observing; while it is a safe and consistent record-keeping data construction, the blockchain is not active in resolution of asked queries due to the lack of reference in its data organisation.

- Flexible authentication algorithms: Blockchain agendas should be flexible to permit the description of authentication methods to assimilate various use of supply chain. Peer's certification strategies use to ensure acceptance or rejection of acquired blocks and transmission to renew the ledger of blockchain [14].

6 Proposed Solutions to the Challenges that Occur in Supply Chain

The preceding challenges encountered in supply chain in collaboration with blockchain technology are recounted and gave resolutions for the supply chain challenge. They are divided into groupings on the basis of the exact problem faced by supply chain challenges.

- Source: Source has proven to solve three challenges, tracking goods source, tracing goods and classifying counterfeit manufactured goods.
- Performance enhancement: The main performance attribute of the supply chain is transference throughput. Blockchain transference details are required by service providers and supervisory agencies. For instance, a double chain is a projected resolution for an integrated and accessible organisation [15].
- Quality assurance and quality control: The high amount of health apprehensions in food consumption triggers the blockchain adoption to note the excellence of processing and thus the excellence of food manufactured goods.
- Sustainability: Source has also played the main part in analysing the resilience of commodities. Chlorofluorocarbon is used to store the carbon imprint of a manufactured goods supply chain by emphasising transference. There is a growing need for these data logs on supervisory bodies to assess the sustainability of supply chain. Conversely, there is also a requirement to consolidate the quantity of carbon that can be supplied by a supplier in such a way that the production of manufactured goods or provisions is sustainable.
- Transparency: This blockchain agenda permits for clear record-keeping of all dealings that take place in the logistics of this supply chain. With the help of transparency, an optimal solution for all the challenges that occur in supply chain is achieved. The apprehension for transparency derives from consumers' requests to appear on food produces. Blockchain may guarantee fair requests in the supply chain.
- Data confidentiality: Data confidentiality is a concern that raised in many supply chain studies, but many proposed activities do not address it [16]. This discourses the anxieties on safety by encoding the data files that are stowed on the block with the public keys of the contributors who are certified to admittance the block.

7 Opportunities and Challenges Occur in Blockchain Technology in Combination with Supply Chain Management

There are lots of opportunities and challenges that occur during the implementation of a blockchain structure for supply chain management. Some of these opportunities as well as challenges are discussed below.

- **Consistency:** Consistency supports transparency, tracking, asset verification and initiation that promotes blockchain acceptance in supply chain processing. For an instance, consistency can be able to help and maintain data records of supply chain verification, tracking and emergence.
- **Tracking accuracy:** To accomplish maximum tracing accuracy, the blockchain requests to note a huge figure of dealings (transmission). The challenge stayed in the volume of stowage prerequisite to handle a distributed ledger of blockchain. As the count of transmission increases, the dimensions of the blockchain's ledger are also increasing.
- **Source:** Source needs the recording of manufactured goods ownership through a supply chain. The blockchain should allow you to retrieve the goods in a proper sequence of records as of its source to the destination. A second data organisation, like a record table of hash function, could remain required to work through blockchain documented data [17].
- **Innovative prototypes of supply chain management:** Reliability is a significant matter underneath a catastrophic/ruinous occurrence that could disturb single or multiple supply chain phases. Original blockchain prototypes and supply chain prototypes are essential to build to cover this problem.
- **Throughput:** It may prerequisite to recording a huge figure of transmission in supply chains. The value of the goods might be higher than other manufactured goods. In fact, in 90-day period, a single car constructor usually delivers about 10 million calls to its contractors. The trade-off between safety and presentation should be well-adjusted on the blockchain.
- **Budget and inconvenience:** Along with the incorporation of big data technology, blockchain can assemble big data related to sales, customers and supply chain. While blockchain technology helps the supply chain management after solving some of the surviving problems, it similarly familiarises supplementary start-up and operating costs. Blockchain technology has the needs of computer setup, data transmission, data assortment and incorporation in present management of supply chain.
- **Security:** While blockchain can solve many supply chain challenges, it can also bring some safety issues. The accomplishment of the blockchain lies in reducing local power. However, high computational power collection, risking and elective on a few nodes can intimidate power allocation. Mining diversity and authentication are important to maintain that characteristic. The characteristic of the blockchain technology depends on the safety of its compatible techniques [18].

8 Summarization, Conclusion and Future Research Direction

This paper presented the introduction of supply chain with its procedure, characteristics and surviving challenges on the way to create it extra active and well-organised. Additionally, the paper described blockchain technology that could enhance the administration of supply chains. Moreover, the paper highlighted the surviving blockchain structures to uncover the challenges of supply chain. Furthermore, the paper enumerated the leftover challenges of supply chains wherever blockchain technology could still catch its use. The paper described possible opportunities and challenges of blockchain technology in combination with supply chain management. In addition, we want to study more in this domain and will provide enhanced version of security with proper collaboration and adoption of new technologies along with ease. As future research, we will focus to protect the blockchain technology integrated with supply chain management while designing a safe supply chain architecture that will be able to protect the data as well as goods from all the harmful vulnerabilities.

References

1. Eltantawy R (2011) Supply management governance role in supply chain risk management and sustainability. Supply Chain Manage- New Perspect. <https://doi.org/10.5772/1999>
2. Amadeo K (2019, June) How supply chain affects the U.S. economy. <https://www.thebalance.com/what-is-the-supply-chain-3305677>
3. Supply chain services (2020) International trade administration. <https://www.trade.gov/supply-chain-services>
4. Pop C, Cioara T, Antal M, Anghel I, Salomie I, Bertoncini M (2018) Blockchain based decentralized management of demand response programs in smart energy grids. Sensors 18(2):162. <https://doi.org/10.3390/s18010162>
5. Malik A, Gautam S, Abidin S, Bhushan B (2019) Blockchain technology-future of IoT: including structure, limitations and various possible attacks. In: 2nd International conference on intelligent computing, instrumentation and control technologies (ICICICT), Kannur, India, pp 1100–1104. <https://doi.org/10.1109/ICICICT46008.2019.8993144>
6. Goyal S, Sharma N, Bhushan B, Shankar A, Sagayam M (2020) Iot enabled technology in secured healthcare: applications, challenges and future directions. In: Cognitive internet of medical things for smart healthcare, pp 25–48. https://doi.org/10.1007/978-3-030-55833-8_2
7. Gautam S, Malik A, Singh N, Kumar S (2019) Recent advances and countermeasures against various attacks in IoT environment. In: 2019 2nd International conference on signal processing and communication (ICSPC). <https://doi.org/10.1109/icspc46172.2019.8976527>
8. Goyal S, Sharma N, Kaushik I, Bhushan B (2021) Blockchain as a solution for security attacks in named data networking of things. In: Security and privacy issues in IoT devices and sensor networks, pp 211–243. <https://doi.org/10.1016/b978-0-12-821255-4.00010-9>
9. Malik A (2020) Steganography: step towards security and privacy of confidential data in insecure medium by Using LSB and cover media (December 12, 2020). SSRN Electron J. <https://doi.org/10.2139/ssrn.3747579>
10. De Giovanni P (2020) Digital supply chain Through IoT, design quality, and circular economy. In: Dynamic quality models and games in digital supply chains, pp 57–89. <https://doi.org/10.1007/978-3-030-66537-1>

11. Gaitán Cremaschi D (n.d.) Sustainability metrics for agri-food supply chains. <https://doi.org/10.18174/380247>
12. Saxena S, Bhushan B, Ahad MA (2021) Blockchain based solutions to secure IoT: background, integration trends and a way forward. *J Network Comput Appl* 103050. <https://doi.org/10.1016/j.jnca.2021.103050>
13. Shakhbulatov D, Arora A, Dong Z, Rojas-Cessa R (2019) Blockchain implementation for analysis of carbon footprint across food supply chain. In: 2019 IEEE international conference on blockchain (Blockchain). <https://doi.org/10.1109/blockchain.2019.00079>
14. Bozic D, Bateman AH (2018) Evaluating social transparency in global fashion supply chains. In: Eco-friendly and Fair, pp 165–173. <https://doi.org/10.4324/9781351058353-16>
15. Faria C, Correia M (2019) BlockSim: blockchain simulator. In: 2019 IEEE international conference on blockchain (Blockchain). <https://doi.org/10.1109/blockchain.2019.00067>
16. Bhushan B, Khamparia A, Sagayam KM, Sharma SK, Ahad MA, Debnath NC (2020) Blockchain for smart cities: a review of architectures, integration trends and future research directions. *Sustain Cities Soc* 61:102360. <https://doi.org/10.1016/j.scs.2020.102360>
17. Morana J (2013) The economic aspect of sustainable supply chain management. In: Sustainable supply chain management, pp 1–51. <https://doi.org/10.1002/9781118604069.ch1>
18. Nakhai I, Jafari S (2010) Developing smart and active packaging of inventory model in drug supply chain for special diseases. In: IEEE International conference on management of innovation & technology, Singapore, pp 550–555. <https://doi.org/10.1109/ICMIT.2010.5492762>

Preventing and Detecting Intrusion of Cyberattacks in Smart Grid by Integrating Blockchain



Avinash Kumar, Bharat Bhushan, and Parma Nand

Abstract The limitation of fossil fuels has increased the amount of carbon in the atmosphere. The emergence of electricity-based utilities such as electric vehicles, electric-based cooking utensils, and increased electrified train systems has given birth to smart grid (SG). Moreover, the cost of the traditional power hyphen-based system comes at very high price for the consumers. There are also various security flaws in traditional SG system. The SG is based on a centralized system, which again makes them prone to many attacks. Also, it fails when renewable energy source comes into the picture. This paper deeply analyzes the infrastructure, weakness in managing power distribution, and security flaws in the SG. The paper also covers the vital solutions taking both preventive and curative measures for the security issues in SG. The variable elements of SG, such as consuming price and variable demands of power, are deeply analyzed to make the system more efficient and economical for both producer and consumer.

Keywords Blockchain · Smart grid · Transaction energy · Cybersecurity · Renewable energy · Channel attack · Distributed generator · Advanced meter infrastructure

1 Introduction

The earlier few decades have seen a sharp decline in the availability of fossil fuels. This depreciated the efficiency of a power system in the sense of electric power generation, transmission over cable, and distribution to consumers. Moreover, the increase in human population has also increased the demand for energy consumption. The use of traditional power generation system, which highly depends upon fossil fuels, has spiked the amount of carbon in the atmosphere that is very alarming to

A. Kumar (✉) · B. Bhushan · P. Nand

School of Engineering and Technology (SET), Sharda University, Greater Noida, India

P. Nand

e-mail: parma.nand@sharda.ac.in

not only the human being but it is also harmful to the rest of the living creatures on this planet. The need for energy is a never-ending demand, and it is still growing at a faster rate [1]. In order to reduce the dependency on fossil fuels, the power system is now generating energy through renewable sources such as wind, solar, and water. Though these seem very efficient, the distribution of energy to the consumer is not an easy process in comparison to the traditional power system [2]. In order to tackle these issues, smart grid (SG) got evolved from the traditional power system [3]. The commencement of SG has made the possibility of a bi-directional exchange of energy and information. The SG consists of a smart meter, control system, and interconnected device for improved utilization of the renewable power generation system. Though SG seems very useful, it lacks some of the features such as the centralized concept involved in SG increases delay in the response, and it makes the SG very costly as well [4]. The cost to customer is also not very effectively calculated because there could be off-pick and on-pick in demand for the energy by the consumer.

Unfortunately, SG lacks the potential to calculate a proper revenue taxing mechanism. Therefore, the integration of blockchain with SG could bring a proper solution for bill management. It has also known that centralized system-based concept like SG is more prone to cyberattacks. The intrusion in the system is easier because of a longer transmission channel. This could be tackled by blockchain technology. Blockchain will provide a transparent and trusted platform for information transactions in the entire chain containing multiple nodes or blocks. The security triads that are confidentiality, integrity, and availability (CIA) could be achieved using blockchain [5]. The payment mechanism using blockchain cryptocurrency is very useful in managing the rise and fall of demand of any transaction-based system.

Many works on the use of blockchain have been done before, but this paper tries to reduce the security flaws and anomalies in distribution of energy in SG. The paper tries to cover the most important energy-saving, payment management, and security management in SG. In summary, the major contribution of this work can be enumerated as below.

- This paper presents the necessity for modification of SG to meet state-of-the-art implementation.
- This paper suggests the vulnerabilities evolving in SG leading to a data breach.
- This paper also compares and contrasts the technological features of centralized SG with the suggested decentralized blockchain mechanism.
- This paper also presents the economic revenue taxation issue in the renewable energy-based SG.

The remainder of this paper is organized as follows. Section 2 introduces blockchain technology and its important terminologies. Section 3 deals with issues involved in utilizing blockchain over SG. Section 4 deeply explains the implementation of blockchain applications in SG. Section 5 explains the role of blockchain in SG for cybersecurity, and finally, Sect. 6 summarizes the conclusion of the work.

2 Blockchain Technology

Blockchain is a set of blocks containing information. The system is a chronologically formed chain where the main aim is to transfer information in a ciphered form, which performs a truth check in case of any illegal intrusion is done by an adversary within the chain. There are various important concepts and terms involved in it that are further discussed in the below subsections.

2.1 *Concept and Terminology*

The definition as suggested above signifies blockchain as a collection of blocks. The first block of chain is called Genesis that acts as the foundation for the chain. The emerging new blocks are added to each other, which are directly or indirectly connected to Genesis. These blocks also contain hash apart from the information. The hash acts originally checks for the information as it represents signature or fingerprint for the block. Hence, the modification due to intrusion or attacks in the block will modify the original hash [6]. Thus, the hash in the chain reflects the security anomaly. Thus, the feature of blockchain vitalizes it as a secure option for the industries [7]. The tampering of information within a block modifies the hash of this block only; it does not hamper hashes of other blocks connected to the victimized block. Though only one block is victimized, the rest of the blocks associated with a victimized block are made invalid, and hence, one tampered block could result in invalidating all connected blocks within the chain [7]. The working of blockchain is initiated by node, it broadcast the transaction in an entire network of the blockchain, then transaction is verified using hashes by nodes, and after successful verification, the transaction is accommodated with existing block and is made permanent and immutable. The hashes, which acts as signature contributes security to the blockchain in a most suitable way.

2.2 *Type and Distribution Feature of Blockchain*

The hashes, despite providing security in the form of signature, could be tampered with using super computer without being noticed. Therefore, various mechanisms have been proposed, which are known as consensus in blockchain [8]. These proposed consensuses that verify the transaction work prior to the addition of new block or node to the existing chain. Hence, the blockchain could expand securely without losing integrity of the information stored in the older and newly added blocks or nodes. The working of consensus is done on the basis of discretely predefined time fashion. The time interval signifies the time incurred from the initiation of a transaction to the time of its addition to the chain. The time of confirmation depends upon the size of the

Table 1 Blockchain versus shared database

Feature	Blockchain technology	Shared database
Operations	Adding new block	Creation/insertion/updation/deletion
Disintermediation	Blockchain follows this	Not adopted
Robustness	Completely robust	Partially robust
Confidentiality	Completely confidential	Partially robust

node or block, volume of the transaction as well the mechanism used for consensus. There are four major consensus mechanisms for blockchain. The first one is called proof of work (PoW) which is most widely used in blockchain where miners are the decision-making element for a new node or block addition; Ethereum and Bitcoin fall under this consensus. The second one is known as proof of stake (PoS), which does not include miners for validation, rather it uses validators for the stake within the block. Despite being energy efficient, PoS is not recommended for industries as that of PoW. The third consensus is called proof of authority (PoA) which supports chain working only using approved accounts. The last one is the Practical Byzantine Fault Tolerance (PBFT), which uses two primary and secondary replicas. Secondary works only when the primary is compromised.

The blockchain has distributed architecture that makes it more useful in comparison to other technologies. In addition, the decentralized feature of blockchain makes it more resilient toward any cyberattacks. Table 1 tries to compare and contrast blockchain technology with the traditional shared database system.

2.3 Real-Life Blockchain Utilization

Blockchain is now beneficial for numerous systems such as the Internet of things (IoT), smart healthcare, currency transactions, and many others where data are transferred from one part to another. The most vibrant application of blockchain is cryptocurrency. Bitcoin, Ripple, and Ethereum are the few most famous cryptocurrencies based on blockchain [9]. Some of the major applications of blockchain are summarized in Table 2.

The blockchain is also used for technology transfer. Dubai government started using blockchain for making a roadmap for business and established open platform that is used for transfer of technologies to rest of the world [18]. The blockchain was also adopted by the United Nations (UN) for distributing food among the citizens of Pakistan in the year 2017, which was adopted because of the transparent and secure monitoring feature of blockchain [19]. The above all example reflects the vitality of blockchain implementation for a corporate and social cause.

Table 2 Blockchain uses

Field of use	Implementation
Market	Quality monitoring [10] Data transfer and bill monitoring [11]
Government law	Identity as well as registry management [12] Ownership transfer proof [13]
Smart healthcare	Patients' record repository [14] Digital wallet management [15]
Science	Analysis of crowd [16] Peer to peer management of resources [17]

3 Blockchain Utilization Issues

The above sections represent useful implementation and features of blockchain. Despite its usefulness, there are some issues with the blockchain. The below subsections covers the important issues involved in utilizing blockchain.

3.1 *Flexibility in System and Negative Pricing Issues*

The inclusion of renewable energy (RE) in power systems based on the SG has increased drastically. The rise and fall of price in power systems depend upon the various conditions, and few are operation of machine, generation, and change in demand. When the demand is deficient, the price falls below the variable cost, and when the demand is very high, the price rises very high above the variable cost. The low demand results in a negative value of the price. Negative pricing is proportional to the market demand that is directly related to the power generation as well as reduction of resources. The negative pricing challenges the flexibility of pricing, which affects the risk management. The negative pricing heavily affects the power system based on RE because the generation is almost constant despite the low demand. The tax incurred by thermal production units faced these negative pricing issues due to pricing concept [20]. The mitigation for negative pricing was also attempted by Department of Energy (DE) in the year 2017 [21]. The negative pricing problem becomes more prevalent when the subsidies are applied for a certain amount in the revenue collection process. The use of cryptocurrency mining could be a probable solution for the demand response (DR) processes, which cause negative pricing [22, 23]. Therefore, integration of blockchain's cryptocurrency with the power system in SG would result in a more economically viable power distribution system.

3.2 Management of Energy and Arbitrage

The energy consumption in many nations is calculated using computer-assisted technology, which is termed as arbitrage. The energy from the storage system is efficiently distributed using the arbitrage. The large energy storage system is used for balancing load by reducing the difference between the on-pick as well as off-pick time. This will help in improving the economical as well as technical mechanism for the modern power system. The adoption of cryptocurrency mining could be used in some systems as alternative for the energy storage system. The concept could be used because it acts as a load during low load duration, while during consumption, it acts as a charging source. The charging helps the smart power system to balance load as it consumes the extra unused generated power. Also, scheduling of various blockchain S-based operation that uses cryptocurrency mining during off-peak hours might be used to fill the load demand and nullifies the use of mining process during on-pick time. This would also help to meet the power need for storage system even when the system does not have sufficient infrastructure for storage. The extra energy developed by power system could also be transferred to other locations where there is low energy generation; the surplus energy cost could be easily calculated using cryptocurrency. This will increase the efficiency of power system market. Therefore, the arbitrage could be easily tackled by cryptocurrency, which is based on blockchain.

3.3 Trading Issue for Smart Grids

The traditional power system is based on the conventional payment system, which includes currencies of respective nations. The use of cryptocurrency, which is based on blockchain, is a great alternative for modern as well as traditional power systems [24]. The use of cryptocurrency could be considered more suitable for modern technologies such as smart homes and smart healthcare, where distribution of energy is done using SG. Hence, it could be inferred that cryptocurrency is a type of electricity that is converted for easy transferring. The grid is controlled by a two-way management system for energy transmission between the grid administrator and the consumers. Therefore, economic and security control must be considered in an SG-based system that could be easily achieved by the blockchain-based technology. The use of peer-to-peer (P2P) decentralized systems as well as centralized transactions among the market needs efficiency, resiliency, security, and reliability. Therefore, the increased system such as electric vehicle (EV), smart homes, smart healthcare, and other related systems which entirely depends upon electricity can use blockchain for viability of their bill management. Blockchain technology could be very crucial in the payment mechanism of wind energy-based systems combined with SGs. The system should be tackled using appropriate scheduling distribution. If the energy produced by the windmill is more than the consumer's requirement, the crypto mining could be

used to preserve them or transfer them to other locations. Therefore, the trading issue lies in the SG system and could be potentially resolved using blockchain technology.

3.4 Management Issue in Smart Grid

Transaction energy (TE) platforms based on blockchain technology could be a new solution for power management. The transaction of P2P energy using digital currency is growing rapidly. Blockchain can perform intrusion detection within the network of the multi-micro grid. The use of consensus makes this threat detection a possible outcome as the modification of hash signals the possibility of intrusion in the system. The integration of blockchain technology can ensure improved functionalities such as negative pricing management, controlling of many grids authorizations with the help of DR analysis. The management of demand is tackled by increasing the time duration for the low as well as high load of power generation. The reduction in time duration of demand for the low power generation could be achieved most efficiently by cryptocurrency mining. Moreover, DR also supports the management of price spikes when there is scarcity of power supply. Therefore, the uncertainties in demand as well supply could be achieved by DR mechanism.

4 Blockchain Application in Smart Grid

Blockchain is very useful for secure transactions, and also, it requires high computation as it involves huge number of blocks or nodes. There are many reasons for its application in SG, which are explained in the below subsections.

4.1 Motivation Behind Blockchain Technology in Smart Grid

As per the discussions in the previous section, it is clear that the new infrastructure of grid has come into a format where terms such as negative pricing, transaction security, and market control are crucial parameters that are considered at priority level. Moreover, the SG is now using digital communication, modern infrastructure, which is transforming legacy traditional grid to more intelligent access, efficient transaction, accurate data, and secure delivery network-based system. The change and modernization in grid system have been brought about due to random climate change and the need for the sustainable energy. The motivation for these transformations and adopting state-of-the-art features comes from the need of sustainable renewable energy and use of distributed energy in order to reduce the dependency for the fossil fuel. Also, the prevention from unauthorized access, modification of information and detection of intrusion are other set of features that motivate the use of blockchain

in SG. The traditional grid system is conceptualized on the power distribution from longer distance transmission, but the SG reduces the distance between the producer and the consumer via implementation of distributed renewable energy system. In recent time, energy Internet (EI) power system concept has been introduced that is considered as next generation grid because it takes into account information security, economic viability, and energy efficiency of power system. The EI main objective is to provide wide opportunities to facilitate seamless collaboration of clean and green renewable resources of energy with the SG in order to provide interactions among various SGs to function in an autonomous manner for efficient energy generation and distribution in a secure way. EI want to treat the power grid load sharing capability in a way as it does with load balancing of data over the internet.

4.2 Smart Metering Using Blockchain Technology

The development of advanced meter infrastructure (AMI), increased consumers, companies, launch of electric-based vehicles, and renewable resource system-based producer in SG network has made the entire system more interactive with the automation and two-way communication support of smart meter. The smart meters are far more intelligent than the traditional meter, and this is because smart meter supports billing, monitoring, appliance access control, and troubleshooting. The transaction in AMI is usually done using a wide area network (WAN), and data are stored in a centralized database or over the cloud. The centralized database system might face traditional issues such as privacy breaches and single-point failure. Also, the addition of more connections can increase in delayed response, higher jitter, increased latency, and other network-related problems. AMI consists of billing for smart meters at home, offices, and electric vehicles that consist of huge amounts of sensitive data of consumers and producers as well. This may include details of payment cards of consumers. These payment data are shared at wider network in AMI, which increase the chance of data breach. These data could also reveal the other sensitive information such as address of the consumer, their identity, and other sensitive data associated with the payment system. Also, the trust problem exists in the centralized system, both consumers and producers face transparency issues in the system. Integration of blockchain is one of the most feasible solutions to the vulnerabilities existing in the AMI. Blockchain is highly secure because of its various options available in the form of consensus. Moreover, the system is decentralized, which means it can eradicate the vulnerabilities found in AMI due to a centralized database system. The blockchain also ensures secure transactions; hence, the breach of privacy and loss of sensitive information is achievable using it. The chances of payment card details that are usually revealed in traditional systems can be resolved using blockchain technology.

4.3 Decentralized Trading of Energy Using Blockchain Technology

The new concept of bi-directional flow of energy makes the consumer act as a producer when considering SG in modern times. Installation of solar panels on houses connected to grid is an example of this concept where unused energy of house is transmitted to the grid and the consumers even get paid. The SG aims to increase consumer, producer as well as prosumer (consists of both consumer and producer) in the energy trading market. The small generating units, which includes an energy storage unit, micro-grids, electric-based vehicles, are used for sharing energy with each other in order to reduce off-pick or on-pick load, reducing load on the main grid, encouraging a green environment, reducing energy loss in transmission, and many other that make SG an economically and environmentally more suitable power generation system. Though these targets seem very useful for SG, it is not achievable using the traditional system of transaction, monitoring, and access control of grid system. Also, the market seems closed to the consumer as they have limited options of payment through traditional mechanisms. Therefore, the use of blockchain will not only revolutionize the SG in making it easier for making transactions, rather it will also make the SG more secure. Blockchain will make the consumer more confident on SG as blockchain makes any system transparent and decentralized, which increases the security and response time.

5 Role of Blockchain Technology for Cybersecurity in Smart Grid

Blockchain could bring significant changes in the security parameters such as confidentiality and integrity. Therefore, the coming subsections present vitality of blockchain for achieving cybersecurity parameters in SG.

5.1 Field Management and Channel Security

Blockchain is widely used in SG for field management and communication. The smart system (SS), such as smart homes and smart healthcare, uses the concept of IoT where nodes are interconnected using the concept of blockchain to provide secure P2P interaction in SG to measure reliability, data quality, and other management related to filed devices. The blockchain helps to distribute data with less delay and helps to make the replica of the data to make it available locally to the nearest receiver or consumer. Also, the use of AMI integrating with SG and taking blockchain for a transaction would make the entire system more resilient to channel-based attacks. Blockchain technology can easily track the intruder attacks performed by the insider

of the organization in the SG. Moreover, the man-in-the-middle (MITM) attacks could be easily traced based upon the distortion of original information, which has a unique hash value or signature. The attacks prevention and detection will help the SG to measure the filed data accurately. It will also make the communication between interconnected devices more secure by making the network tamper-proof.

5.2 *Synchronization of Blockchain for Power Generation*

The modern SGs are based on renewable powered by distributed generator (DG) in place of a traditional centralized power generator. Moreover, the centralized system reveals more sensitive information rather than the decentralized one. Hence, the integration of blockchain technology will enhance the security of SG at highest level. The breach of private and sensitive data can be prevented. Though blockchain helps in improving the security of the SG, it very important to synchronize the implantation according to the policy and rule of the SG. A failure to synchronization could result in an anomaly for the generation and distribution of power by the prosumer (consists of both consumer and producer). The anomaly could expose private and sensitive information that could be gathered by the hacker to exploit the SG.

6 Conclusion and Future Research Directions

The SG is an essential implementation of power system. The paper significantly explains the need for the SG and various important terminologies involved in it. The paper also covers the security aspects of the blockchain and tries its best to map the eradication of vulnerability by using blockchain. Moreover, the paper also highlights some of the major real-life implementations of blockchain in order to show its vitality when integrated with the SG. The paper also suggests solutions for the negative pricing, which is a significant issue of SG. Finally, the paper concludes by presenting the vital aspect of preserving private and sensitive data over SG. The paper has certain limitations. The testbeds for blockchain-integrated SG could provide more analysis for SG and provide efficiency for whole grid process. Also, the testbeds will produce any shortcomings of this integrated system during testing phase only; this will reduce the overhead cost that might occur in system due to failures occurring in actual implementation. Hence, inclusion of testbed in future would be very useful pretesting of real SG to be implemented.

References

1. Bhushan B, Sahoo G (2019) E^2 SR²: an acknowledgement-based mobile sink routing protocol with rechargeable sensors for wireless sensor networks. *Wireless Netw* 25(5):2697–2721. <https://doi.org/10.1007/s11276-019-01988-7>
2. Yoldaş Y, Önen A, Muyeen SM, Vasilakos AV, Alan İ (2017) Enhancing smart grid with microgrids: challenges and opportunities. *Renew Sustain Energy Rev* 72:205–214. <https://doi.org/10.1016/j.rser.2017.01.064>
3. Dileep G (2020) A survey on smart grid technologies and applications. *Renew Energy* 146:2589–2625. <https://doi.org/10.1016/j.renene.2019.08.092>
4. Bhushan B, Sahoo G (2017) Recent advances in attacks, technical challenges, vulnerabilities and their countermeasures in wireless sensor networks. *Wireless Pers Commun* 98(2):2037–2077. <https://doi.org/10.1007/s11277-017-4962-0>
5. Sharma N, Kaushik I, Agarwal VK, Bhushan B, Khamparia A (2021) Attacks and security measures in wireless sensor network. In: Intelligent data analytics for terror threat prediction, pp 237–268. <https://doi.org/10.1002/9781119711629.ch12>
6. Beck R (2018) Beyond bitcoin: the rise of blockchain world. *Computer* 51(2):54–58. <https://doi.org/10.1109/mc.2018.1451660>
7. Bhushan B, Sinha P, Sagayam KM, Andrew J (2021) Untangling blockchain technology: a survey on state of the art, security threats, privacy services, applications and future research directions. *Comput Electr Eng* 90:106897. <https://doi.org/10.1016/j.compeleceng.2020.106897>
8. Moubarak J, Filiol E, Chamoun M (2018) On blockchain security and relevant attacks. In: 2018 IEEE Middle East and North Africa communications conference (MENACOMM). <https://doi.org/10.1109/menacomm.2018.8371010>
9. Bhushan B, Khamparia A, Sagayam KM, Sharma SK, Ahad MA, Debnath NC (2020) Blockchain for smart cities: a review of architectures, integration trends and future research directions. *Sustain Cities Soc* 61:102360. <https://doi.org/10.1016/j.scs.2020.102360>
10. Lee C, Mueller J (2019) Can blockchain unlock the investment Africa needs? *Innov: Technol Governance Global* 12(3–4):80–87. https://doi.org/10.1162/inv_a_00277
11. Information technology. Cloud computing. Service level agreement (SLA) framework (n.d.) <https://doi.org/10.3403/30316174u>
12. Kumar A, Abhishek K, Bhushan B, Chakraborty C (2021) Secure access control for manufacturing sector with application of ethereum blockchain. *Peer-to-Peer Network Appl.* <https://doi.org/10.1007/s12083-021-01108-3>
13. Thorwirth N (2020) The decentralized rights locker. *SMPTE Motion Imaging J* 129(3):56–62. <https://doi.org/10.5594/jmi.2019.2957687>
14. Haque AK, Bhushan B, Dhiman G (2021) Conceptualizing smart city applications: requirements, architecture, security issues, and emerging trends. *Expert Syst*. <https://doi.org/10.1111/exsy.12753>
15. Patel M (2017) Blockchain approach for smart health wallet. *IJARCCE* 6(10):131–135. <https://doi.org/10.17148/ijarcce.2017.61022>
16. Wang J, Li M, He Y, Li H, Xiao K, Wang C (2018) A blockchain based privacy-preserving incentive mechanism in crowdsensing applications. *IEEE Access* 6:17545–17556. <https://doi.org/10.1109/access.2018.2805837>
17. He Y, Li H, Cheng X, Liu Y, Yang C, Sun L (2018) A blockchain based truthful incentive mechanism for distributed P2P applications. *IEEE Access* 6:27324–27335. <https://doi.org/10.1109/access.2018.2821705>
18. AlBarghuthi NB, Ncube C, Said H (2019) State of art of the effectiveness in adopting blockchain technology—UAE survey study. In: 2019 Sixth HCT information technology trends (ITT). <https://doi.org/10.1109/itt4889.2019.9075108>
19. Blue J, Condell J, Lunney T (2019) Proving yourself: addressing the refugee identity crisis with Bayesi-chain probability & digital footprints. In: 2019 International symposium on networks, computers and communications (ISNCC). <https://doi.org/10.1109/isncc.2019.8909145>

20. Fanone E, Gamba A, Prokopcuk M (2013) The case of negative day-ahead electricity prices. *Energy Econ* 35:22–34. <https://doi.org/10.1016/j.eneco.2011.12.006>
21. Staff Report to the Secretary on Electricity Markets and Reliability. Energy.gov (n.d.) <https://www.energy.gov/staff-report-secretary-electricity-markets-and-reliability>
22. Bhushan B, Sahoo C, Sinha P, Khamparia A (2020) Unification of blockchain and internet of things (BIoT): requirements, working model, challenges and future directions. *Wireless Netw.* <https://doi.org/10.1007/s11276-020-02445-6>
23. Saxena S, Bhushan B, Ahad MA (2021) Blockchain based solutions to Secure IoT: background, integration trends and a way forward. *J Network Comput Appl* 103050. <https://doi.org/10.1016/j.jnca.2021.103050>
24. Goyal S, Sharma N, Kaushik I, Bhushan B (2021) Blockchain as a solution for security attacks in named data networking of things. In: Security and privacy issues in IoT devices and sensor networks, pp 211–243. <https://doi.org/10.1016/b978-0-12-821255-4.00010-9>

Design of Low Switching Pattern Generator for BIST Architecture



Sachin Kumar Pandey and C. Paramasivam

Abstract During testing, the power dissipation is more than normal working. Among many reasons, one reason is due to a lot of transitions occur between bits of test patterns generated. The work advised here has a low power pattern generator for a built-in-self-test (BIST) design. The proposed design is a low power, improved version of a conventional linear feedback shift register (LFSR). The proposed designed named half-start-half-stop pattern generator (2HS-PG) is comprised of conventional LFSR, redesigned with add-on circuitry which brings down the total number of switching between the bits, results in low switching power during testing. Implementation of both conventional LFSR and 2HS-PG design are done in Cadence Genus tool, using 90 nm standard cell technology library and simulation is done in Xilinx ISE tool. Simulation result and analysis shows that more than 30% dynamic power reduction, when patterns are generated using proposed pattern generator.

Keywords BIST · LFSR · Bit-swapping LFSR (BS-LFSR) · Pseudo random pattern generator (PRPG) · Low power pattern generator

1 Introduction

Increased automation and self-connectivity are just several trends in the modern electronic industry. All of them lead to a steady increase in the number and complexity of electronic controlled systems. Hence, the safety and reliability issues [1, 2] are becoming more severe, thus requiring advanced methods in more sophisticated and robust approaches for both hardware and software security. Testing of the chip also varies with an increase in the technology enhancement, and complexity in the circuits is multiplying; hence, more issues are arising like testing of the design. Testing of the chip also plays a vital role; once the chip is manufactured, it is very essential to test whether it is working as per the specifications. The circuit is tested for any error or fault, mainly stuck-at faults that may be present in the signal lines. During

S. K. Pandey (✉) · C. Paramasivam
Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Bengaluru, India

testing, power wastage is more compared to typical operation mode, and for BIST application, it is more critical because chips have a limited source of power.

In CMOS circuits, average power dissipation is split into three categories. They are:

1. Dynamic or switching power consumption—This is the power consumption when the circuit is operated. When logic is switching from one value to another value, during these transitions, power is consumed. Out of total power, around 90% of power is consumed in the form of dynamic power.

$$P(\text{dyn}) = C * V dd^2 * F \quad (1)$$

2. Static power consumption—It is also called leakage power, it occurs due to the leakage current when the circuit is holding a constant value, i.e., it is not switching.

$$P(\text{leak}) = V dd * I(\text{leak}) \quad (2)$$

3. Short-circuit power consumption—Power consumption during transitions when current flows from power supply to ground.

$$P(\text{short}) = V dd * I(\text{short}) \quad (3)$$

In the previous years, the pronounced word is carried out in the field coupled to low-power design like high-performance scan flip-flops [3], clock gating [4], weighted random pattern generation [5], bit-swapping pattern generator which uses multiplexers to reduce the total switching [6]. Design of LFSR which is based on decreasing the switching activity between the consecutive generated patterns [7]. Fundamentally, testing of the circuits can be done using automatic test pattern generation (ATPG) and BIST. ATPG is an automatic design that is used to test the circuit by generating test patterns with the help of different algorithms. BIST is an extra mechanism added which enables the machine to test itself, and various approaches are used in modeling of BIST along with test compression techniques [8, 9], BIST with capture-per-clock [10], embedded deterministic test to reduce the memory to store test patterns [11], and low-power fault diagnosis method [12]. The leading purpose of BIST is to reduce the time taken for testing in ATE, reducing the overall complexity. The main modules of the BIST are pattern generator which is a conventional linear feedback shift register (LFSR), circuit-under-test (CUT), response analyzer (RA), and BIST controller as designed previously [13]. BIST controller controls the operation, i.e., normal mode or testing mode. During normal mode, the circuit performs usual working, but during the testing mode, pattern generator gets active and feed circuit with random patterns; the response of the circuit is feed to the response analyzer which generate signature value. This signature value is compared with golden response saved in the memory; if value matches, then the circuit is fault-free, else it is faulty. Figure 1 shows a basic BIST architecture block diagram.

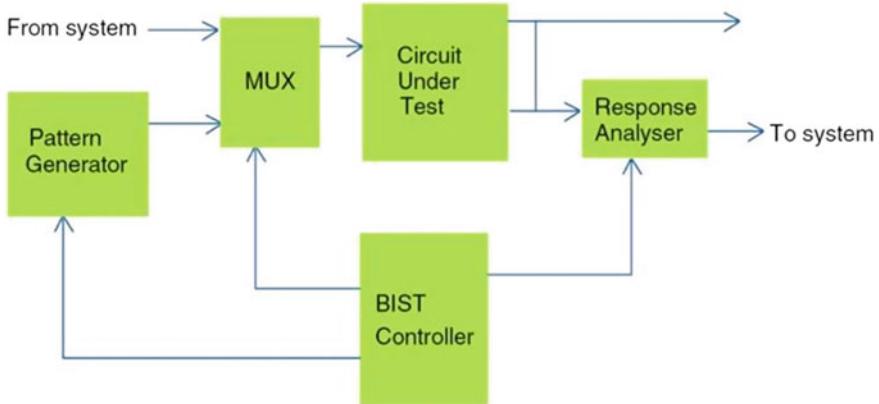


Fig. 1 Typical BIST block diagram [13]

In conventional BIST, LFSR is used as a pattern generator. Standard LFSR is composed of flip-flop and XORed feedback. An n -bit LFSR generates only $2^n - 1$ patterns. Another problem is that there is a lot of stitching between the pattern; hence, more switching power. To rectify this problem, a modified version bit-swapped LFSR (BS-LFSR) [14] is proposed which generates test patterns and also with reduced switching power. As BS-LFSR reduces the switching power, it increases the overall area. To overcome this issue, modified design of pattern generator is proposed named as half-start-half-stop pattern generator (2HS-PG). Proposed design 2HSPG is a combination of two half circuits, in which one will work for a certain period and the other will remain in a stable state and, hence, reducing power consumption.

This paper presents the advantage of using 2HS-PG as a pattern generator for BIST. It also focuses on exhibiting the advantage of using 2HS-PG over LFSR and BS-LFSR [14] by comparing the power and area reports of both the designs. In this paper, Sect. 2 imparts a complete explanation about LFSR and BS-LFSR design. Section 3 demonstrates the design implemented and methodology flow. Section 4 gives an absolute analysis of the results obtained. The result is shown in the form of diagrams, graphs and tables, and Section 5 concludes the paper.

2 Methodology

2.1 Linear Feedback Shift Register

In the BIST application, LFSR is used to provide test patterns to the circuit-under-test. N-bit LFSR consists of n flip-flop and feedback which is composed of XOR gate. There are two types of LFSR design-

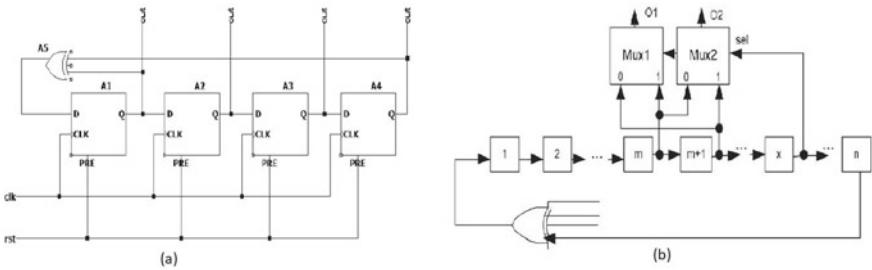


Fig. 2 **a** 4-bit conventional LFSR design, **b** BS-LFSR design [14]

- External LFSR (standard form)
- Internal LFSR (modular form).

LFSR is represented by characteristic polynomial

$$x^n + x^{(n-1)} + x^{(n-2)} + \cdots + x + 1 \quad (4)$$

Based on the polynomial, taps are selected which determine the number of distinct output pattern generated. With an increase in transitions between bits generated, the switching power associated is high, and it will increase further with an increase in the size of pattern generator. Figure 2a shows the LFSR structure of 4-bit; for maximum patterns, given polynomial is $x^4 + x^3 + 1$. The 4-bit polynomial will generate a total of $2^4 - 1$ patterns, i.e., 15 patterns.

2.2 Bit-Swapping LFSR

BS-LFSR [14] is a kind of modified design of LFSR which is capable of generating low switching patterns. The circuit design is combination of conventional LFSR and 2×1 MUX. The BS-LFSR is designed upon basic observations on the design and output pattern generated by LFSR.

Lemma 1 *For an n-bit LFSR (either external or internal) bit-length of LFSR $n > 2$. Swapping between cells is done if they are adjacent cells, the third cell is used to swap the output value of the swapping cell, when the third cell value is 0 swapping is done between two adjacent cells connected with MUX and when the value is 1 swapping is not done.*

As shown in Fig. 2b design of BS-LFSR, output of two adjacent cells is swapped with the help of a multiplexer based on the output of the selection line.

3 Design Implementation

LFSR is the most used topology to implement PRPG. Although LFSR's are very simple to implement, they consume lot of power during testing. To reduce power consumption during testing, a set of strategies are often used. The strategy includes disabling the part of the circuit which is not performing any functional operations during a particular time frame; thus, by disabling the circuit part, power consumption can be reduced. Figure 3a shows the proposed design of the 4-bit half-start-half-stop pattern generator (2HS-PG). To implement a 4-bit design, two 2-bit modified designs of pattern generator are used, and they are combined to form a 4-bit design. Consider Fig. 3a of 2HS-PG design, the total number of patterns generated is 2^4 , i.e., 16 patterns. The first 2-bits are generated by half of the design and the other 2-bits are generated by the next half design. To generate all combination of 1 and 0, modification is done in the typical LFSR design, and additional $(n-1)$ input NOR gate is added in the feedback path. Both $(n/2)$ -bit designs are combined to form an n-bit design of 2HS-PG. The working principle of design is that clock for half design is disabled by clock controller for particular period, and for other half, design clock is continuously working; by this, switching between bits can be decreased. This design also saves clock power by disabling it for certain period of time. Figure 3b shows the architecture of the controller for gated clock to generate clock at particular instance. N -bit down counter is initialized with all bits to zero; n -bit register is binary coded as $2^n - 1$; for example, if clock is needed at fourth period then 4-bit register will store binary value of 3, i.e., 0011.

As shown in Table 1, total transitions for LFSR are 36; for BS-LFSR, 26 transitions and for proposed design, 2HS-PG, only 18 transitions between test pattern bits.

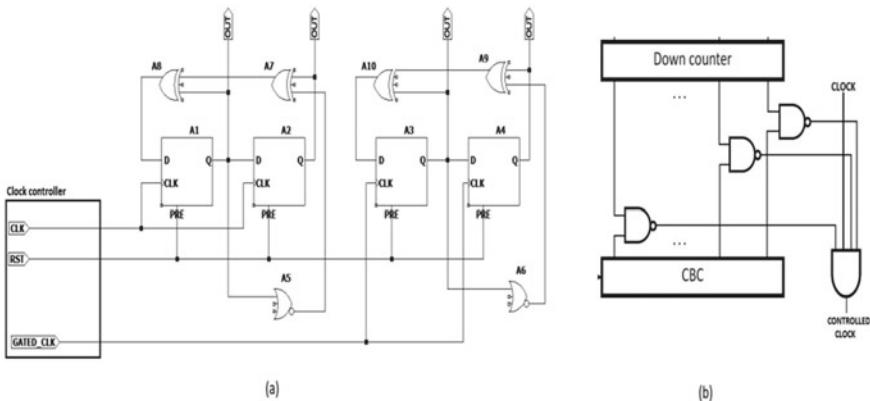


Fig. 3 a 4-bit 2HS-PG design, b Clock controller design [11]

Table 1 4-bit LFSR, BS-LFSR and 2HS-PG test patterns

Test patterns			
Clock	LFSR	BS-LFSR	2HS-PG
1	1111	1111	1111
2	0111	0111	0111
3	1011	1011	0011
4	0101	0110	1011
5	1010	1010	1101
6	1101	1101	0101
7	0110	0101	0001
8	0011	0011	1001
9	1001	1001	1100
10	0100	0100	0111
11	0010	0001	0000
12	0001	0010	1000
13	1000	1000	1110
14	1100	1100	0110
15	1110	1110	0010
16	1111	1111	1010
No. of switching	36	26	18

4 Simulation Results

Implementation of design is done in the Cadence Genus tool, and area reports are obtained. Simulations of conventional LFSR design and modified design 2HS-PG design have been done using Xilinx ISE 14.7 using Verilog language. The target device family used for the implementation is the Xilinx Spartan 3E. Power reports are obtained by simulating designs for different values of n.

Figure 4 shows the output waveform of pattern generators. It can be verified from the simulation result that switching in 2HS-PG is very less when compared to LFSR and BS-LFSR. Switching power is obtained by using an Xpower analyzer; all the designs are simulated for an identical time period and resultant power is calculated. The total power of 4-bit 2HS-PG is 0.044W, for BS-LFSR, it is 0.049W and for LFSR, it is 0.051W; about 13.7% power is saved by using a modified design (Fig. 5).

It can be verified that as n -value is increasing, there is more percentage reduction in total power. Modification in design results in area overhead, Cadence Genus tool is used to implement the design in a 90 nm library. Technology schematic of 4-bit 2HS-PG design generated by Xilinx ISE is shown in Fig. 6.

From Fig. 7, it can be evident that for large design dynamic, power saving will be more than 30%. The proposed pattern generator, 2HS-PG, is used to model low-power BIST architecture for the circuit C17 of ISCAS-85 benchmark designs (Table 2).

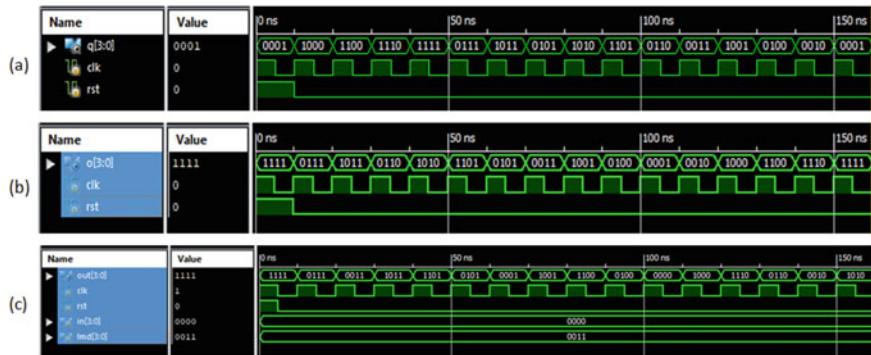


Fig. 4 Output waveform of pattern generated by 4-bit design. **a** LFSR, **b** BS-LFSR and **c** 2HS-PG

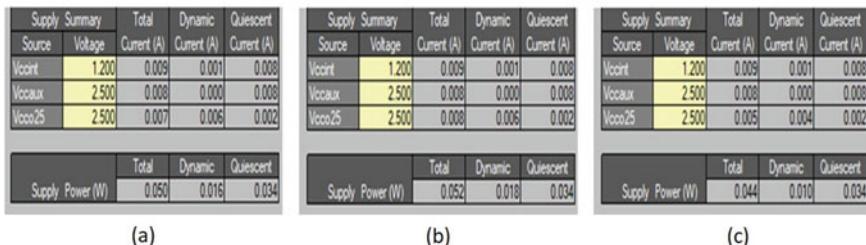


Fig. 5 Power report generated by Xpower analyzer of 4-bit design **a** LFSR, **b** BS-LFSR, **c** 2HS-PG

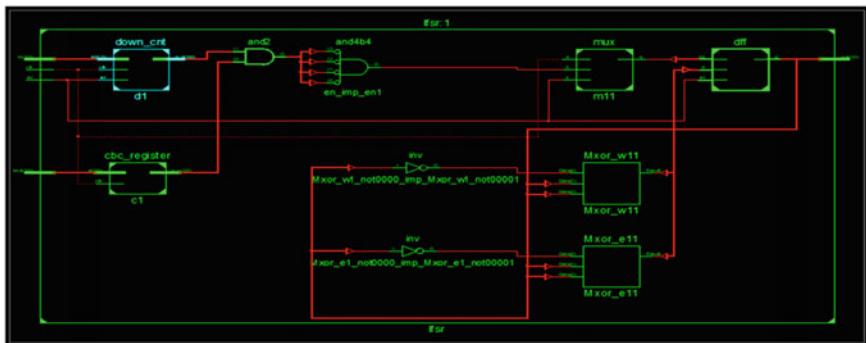


Fig. 6 Technology schematic of 4-bit 2HS-PG architecture

5 Conclusion

Testing of the design before and after manufacturing provides information whether circuit is working as per functionality. Sometimes, it is very critical to test the circuit

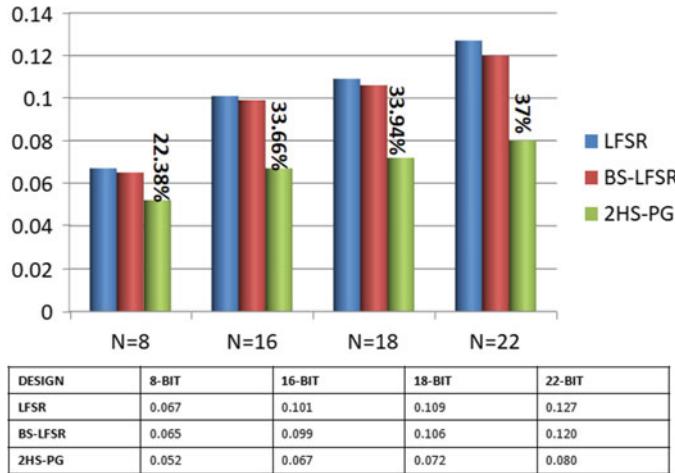


Fig. 7 Total power (in Watts) consumed by pattern generator for different length

Table 2 Total area generated in Cadence Genus using 90 nm lib

Total area in nm			
Design	8-Bit	16-Bit	18-Bit
LFSR	185.197	286.287	376.179
2HS-PG	192.258	429.162	489.483

every time it is switched-on, like SOC's, auto-motives, signal processors, FPGA's, etc. [15]. The work proposed in this paper is mainly focused on designing a modified pattern generator design to optimized BIST architecture. To generate all possible unions of 0s and 1s, the design of an n -bit 2HS-PG is successfully implemented. Simulation results verified that 2HS-PG generates all pattern combination of 0s and 1s. By minimizing the activity factor of flip-flop's and adding extra gated feedback will result in more than one-third reduction in switching power. As the number of bits (N) is increased, the more is the power dissipated, and it will increase vigorously as the length increased. Hence, using of 2HS-PG as a pattern generator is beneficial for power reduction in BIST design.

5.1 Future-Scope

Proposed design minimizes total dynamic power and for large design, power saving is more, but modification leads to area overhead. In future, noble work can be done in order to minimize the area to generate clock controller. Some techniques like FSM, gated clock models can also be used to control the clock of the design.

References

1. Kogan T, Abotbol Y, Boschi G, Harutyunyan G, Martirosyan N, Zorian Y (2018) Advanced uniformed test approach for automotive SOCs. In: 2018 IEEE international test conference (ITC). IEEE, New York, pp 1–10
2. Hatti K, Paramasivam C (2020) Design and implementation of enhanced PUF architecture on FPGA. *Int J Electron Lett*, pp 1–14
3. Eedupuganti K, Murty NS (2017) High performance and power aware scan flip-flop design. In: 2017 IEEE international conference on computational intelligence and computing research (ICCIC), pp 1–4
4. Aloisi W, Mita R (2008) Gated-clock design of linear-feedback shift registers. *IEEE Trans Circuits Syst II Express Briefs* 55(6):546–550
5. Zhang X, Shan W, Roy K (2000) Low-power weighted random pattern testing. *IEEE Trans Comput Aided Des Integr Circuits Syst* 19(11):1389–1398
6. Abu-Issa A, Quigley S (2008) Bit-swapping lfsr for low-power bist. *Electron Lett* 44(6):401–403
7. Tehranipoor M, Nourani M, Ahmed N (2005) Low transition LFSR for BIST-based applications. In: 14th Asian Test Symposium (ATS'05). IEEE, New York, pp 138–143
8. Roy A, Anita J (2017) Pattern generation and test compression using presto generator. In: International symposium on security in computing and communication. Springer, Berlin, pp 276–285
9. Anita J, Sudheesh P (2016) Test power reduction and test pattern generation for multiple faults using zero suppressed decision diagrams. *Int J High Perform Syst Archit* 6(1):51–60
10. Moghaddam E, Mukherjee N, Rajski J, Solecki J, Tyszer J, Zawada J (2018) Logic bist with capture-per-clock hybrid test points. *IEEE Trans Comput Aided Des Integr Circuits Syst* 38(6):1028–1041
11. Liu Y, Mukherjee N, Rajski J, Reddy SM, Tyszer J (2020) Deterministic Stellar BIST for automotive ICs. *IEEE Trans Comput Aided Des Integr Circuits Syst* 39(8):1699–1710. <https://doi.org/10.1109/TCAD.2019.2925353>
12. Kumar CN, Madhumitha A, Preetam NS, Gupta PV, Anita J (2019) Fault diagnosis using automatic test pattern generation and test power reduction technique for VLSI circuits. In: 2019 3rd International conference on trends in electronics and Informatics (ICOEI), pp 412–417
13. Devika K, Bhakthavatchalu R (2017) Design of efficient programmable test-per scan logic BIST modules. In: 2017 International conference on microelectronic devices, circuits and systems (ICMDCS). IEEE, New York, pp 1–6
14. Abu-Issa AS, Quigley SF (2009) Bit-swapping lfsr and scan-chain ordering: a novel technique for peak- and average-power reduction in scan-based bist. *IEEE Trans Comput Aided Des Integr Circuits Syst* 28(5):755–759
15. Harshitha Y, Paramasivam C (2020) Design of low complexity high performance LUT based feed-forward FFT architecture. In: 2020 IEEE international conference for innovation in technology (INOCON). IEEE, New York, pp 1–5

Connecting Blockchain with IoT— A Review



**R. Anusha, Mohamed Yousuff, Bharat Bhushan, J. Deepa, J. Vijayashree,
and J. Jayashree**

Abstract Current trends have diagnosed the emergence of two technologies by integrating them. The two technologies are “Blockchain” and “Internet of Things(IoT).” The IoT stands for providing “smart” features to all the products where it steps into it. For example, IoT Smart City, Smart E-Healthcare system, Smart Home, Smart domestic appliances, Smart Agriculture exist. The next technology is called blockchain, which establishes the P2P (peer-to-peer) network, decentralized and non-tampering technique exists. Thus the marriage between blockchain and the IoT poses many advantages. This paper reviews the intersection of these two recent research areas for the past three years. We hope this paper supports the new researchers and engineers interested in blockchain to build future blockchain–IoT systems.

Keywords Blockchain · IoT · Blockchain_IoT Architecture

1 Introduction

The IoT was coined by Kevin Ashton called “Computer Everywhere.” This IoT is a new, rapidly growing technique that enables people to communicate with each other and their surroundings. It has four types of wings on it such as wireless sensor network (WSN), machine-to-machine (M2M), radio-frequency identification (RFID) and supervisory control and data acquisition (SCADA). IoT has many interdisciplinary communications of fields such as healthcare, smart city, logistics and Industry 4.0. This IoT has a significant impact on the evolution of Industry 4.0 [4].

R. Anusha (✉) · M. Yousuff · J. Vijayashree · J. Jayashree
Vellore Institute of Technology, Vellore, Tamilnadu 632014, India

B. Bhushan
Sharda University, Greater Noida, Uttar Pradesh 201310, India

J. Deepa
Veltech Rangarajan & Dr Sakunthala R&D Institute of Science & Technology, Avadi, Chennai 600062, India

The next most enormous, regressive technique on every researcher is called a blockchain. The following are decentralization, immutability, consensus, mechanisms, distributed ledger, transparency, security, privacy, traceability, reliability, distribution and peer-to-peer network [6]. IoT has many security challenges that can be recovered by using blockchain. The marriage between blockchain and other emerging technology tends to non-tamper-proof property on it. Thus, the blockchain and IoT have provided many advantages for the future build system [2].

This paper describes blockchain and IoT separately. It also depicts their concepts separately. The gap in research of both fields is how to integrate these two technologies and in what way IoT get benefited from the usage of blockchain. The layers of both blockchain and IoT are tabulated in this paper. Many researchers stated that the real, analyzed, decentralized approach thus enables for further intricated details about blockchain–IoT system requirements and architecture [10]. There are only 35 papers that have the real implementation of blockchain–IoT platforms [3]. In this paper, we filled those research gaps.

Section 2 of the paper discusses the concept of blockchain and the disadvantages of IoT. Section 3 elaborates on the architecture of blockchain and IoT, also about the layers of blockchain and IoT and the problems of IoT. Section 4 instructs about the benefits of IoT from the blockchain. Section 5 deliberates the conclusion. Finally, in Sect. 6 we included the future scope of the technology.

2 Blockchain Concepts

This part deals with the concepts of blockchain.

2.1 Architecture

There is no standard architecture for IoT and blockchain. The modeled architecture of blockchain and IoT is shown in Fig. 2. Blockchain is called a distributed database [11]. The participants can store every transaction copy in a distributed ledger; no central authority monitors the transaction, which is sent to be *distributed*, a *decentralization* property. The structure of blockchain is displayed in Fig. 1. Blockchain provides high security naturally. How do these characteristics work? Because it stores the previous hash value [5]. Once the changes happened, the hash value may differ. Thus, it cannot be altered, erased or tampered with. This enables *immutability* and *security*. The special concept called *transparency* enables a peer-to-peer network, i.e., all the nodes can able to see the data available in a block [8].

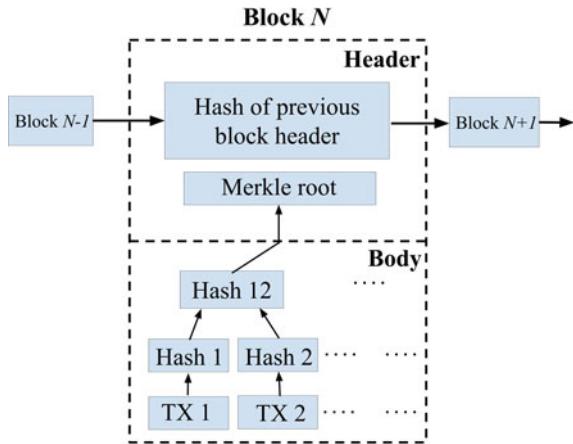


Fig. 1 Blockchain structure (*Note Tx*-Transaction)

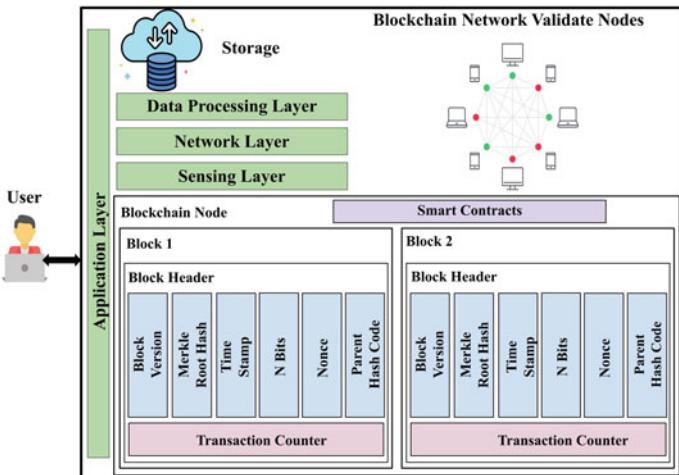


Fig. 2 Modeled blockchain–IoT architecture

2.2 Types of Blockchain

The blockchain is categorized based on their characteristics like decentralization, i.e., the usage will be based on permission provided to the user. **Private Blockchain** is completely based on a particular concern to maintain the blockchain nodes. In **Public Blockchain**, no permission is required for nodes; anybody can join and access the network blockchain publicly. **Hybrid/Consortium Blockchain** is based on both public and private.

2.3 Consensus Algorithm of Blockchain

When the block has to join in the network of successive blocks, so that has to be validated by some process that is said to be consensus algorithm (CA). It is a protocol that maintains tamper-free and also avoids invalid transactions. This means the process of 2V's (verification and validation) has been done by giving the puzzle to solve it to the user wish to join in the network, maintained by miners, a small number of fees deducted by the system based on the platform that we use, i.e., for fees. Many mechanisms are available.

Proof of Work (PoW) The main concept of this work is to maintain the forgery nodes as well as true nodes. The ability of the system involves symmetry to the system's computational and also the resource power in this work. Since blockchain is a decentralized network, it has to be selected on a node to record all the transactions. For this, random selection is chosen in case that is prone to security threats. So that particular node has to show that it is not vulnerable. In this phase, the nonce and cryptographic hash are used; in general, it will be the SHA-256 hash function. The user has to solve the difficult problem in order to get incentives and to be the added successive blocks. Once the changes in the nonce occur, it might lead to a change in Miner's value.

The miners keep on checking the target value; once the node touches the particular value, that miner block is forwarded to all over the network, then all the further nodes have to verify, i.e., all other nodes miners will check for a hash value to verify. Once it is verified, it will be added to the blockchain. Simultaneously, multiple blocks can be successively added to the network when the target value is found in all other miners. Thus, this makes branches in the network; PoW has a longer chain with authentication. But it requires high computational power and also resources. Primecoin, the selection of prime number, exists in mathematical research.

Proof of Stake (PoS) This work does not waste any resources, but it was the best alternative to PoW. In PoS, the miners with heavy blocks are the minimum prone to a security attack, and the miner with more coins can produce the next one. This makes the domination by the wealthier nodes with other nodes in the network. In order to solve this, some examples of PoS were Peercoin and Blackcoin. Peercoin is based on the coin-age (that the miner possesses) i.e., how long or older the miner in the network gets the priority. Blackcoin is based on randomness. PoW is slowly moving to PoS [12].

Practical Byzantine Fault Tolerance (PBFT) The Byzantine flaws can be tolerated by these mechanisms. Generation 3.0, i.e., Hyperledger Fabric, uses this PBFT mechanism. This PBFT has some phases like prepared, committed. The main node will be chosen first based on specific rules. This maintains the transaction order. Every block is added to the network at every round. So each phase, the nodes get the votes of one-third. This work makes every node to be highlighted in the network. The best example of blockchain, NEO (Antshares) coined in 2017, uses the special delegated PBFT(dPBFT), is used to vote for the recorded transaction.

Delegated PoS (DPoS) is delegated to publish the blocks and tries to validate them. If fewer nodes are there in the network, it is easy to verify and maintain the transactions as soon as possible. Block size and block intervals are altered, which can be further supported by DPos. BitShares also uses DPoS.

Ripple work has many sub-networks within the trusted existing networks. In this work, nodes have two main parts namely, consensus process participating server (CPPS) and client to transfer funds (CTF) only. The existing network has the unique node list (UNL) to prevent forgery. This may check UNL for every transaction based on the agreement it will operate.

Tendermint is same as PBFT. This is based on the proposer node, which is used to perform further transactions. It is divide by voting, CommittingPreviously (CPrev), Commit. This CPrev will be kept on broadcasting over the network before the process starts; thus, the nodes which get two or three CPrev will tend to commit the work. The commit phase then enables it and sends it to the block.

2.4 Disadvantages of IoT

Recently, many researchers are talking about the IoT and its advantages and integration with other enabling technologies like blockchain, artificial intelligence (AI), machine learning (ML), deep learning DL, etc., Even though it has many merits, it has some disadvantages.

Scalability is about quantity of work the device can do. We all know, now the world will not run without smart gadgets like phones and TV. Thus, the IoT makes everything to be actually smart; this makes them difficult to handle the huge amount of data sending, receiving, etc., over the network, increasing like cancer cells, day by day. This is called scaling of data in IoT. Thus, the handling of resource utilization for those huge data by the IoT framework.

Heterogeneity In an IoT system the different nodes/systems and intersecting other technologies, so each platform should demand authorization policy to enable this property is the most significant difficulty for IoT.

Privacy Since the IoT has a vast number of users, it should maintain the user's privacy to prevent unauthorized users from accessing the data from the user [9].

Security The system should handle a large amount of data so that many users may try to behave like trusted parties and try to access other data; thus, maintaining the security of all the users participating over the network was quite tricky in IoT [7]. The security threats of IoT differ from layer to layer. In some of the attacks like denial of service (DoS), man-in-the-middle attack, side-channel attack, cross-site scripting attack, etc., all the parts of IoT like hardware part also prone to security attacks.

Access Control and Transparency Trusting the people over the network by the nodes was very difficult, and transparency is another big impact in IoT because of privacy and security.

Interoperability Due to the large amount of data processing in systems, the data transfer and knowledge about the data between the internodes in the network was quite difficult in IoT.

Immutability The changes of data in the network lead to break security, and privacy issues of participating nodes may exist. This is the threat identified, and several attacks that are classified still exist in research. This immutability was very difficult in IoT.

3 Architecture of Blockchain-IoT

Figure 2 depicts the combined architecture of both domains. All the users interact with the system through application layers. The transaction request is recorded and maintained by the blockchain nodes, and as well as it will be verified and validated by the consensus protocol. Once that results in a true and legitimate identity, it will be attached to the blockchain network and data sent to all the participants. Blockchain intersects with IoT to provide security, privacy, decentralized platform and so on.

3.1 Layers of IoT and Blockchain

The IoT has many things to be connected in a single network, i.e., heterogeneous in nature. There is a uniform or standard architecture or uniform layered architecture for IoT. The layered architecture of IoT and blockchain has been discussed below in Tables 1 and 2. Both technologies have some unique features on the different layers.

Table 1 Layers of blockchain

S. No.	Layers	Explanation
1.	Data layer	The chain structure has the data processed by the tree root structure with a timestamp that says the exact time
2.	Network layer	It enables peer-to-peer network where the participants can join and leave the network as per their wish. It has distribution property on it
3.	Consensus layer	Verification and validation done by the block by itself using miner which provides a puzzle to solve it for authentication [POW, POS, DPOS, PBFT]
4.	Contract layer	It is a self-executing code; based on these conditions, it gets executed automatically, i.e., the two unknown people, that makes trust
5	Application layer	The final end product like cryptocurrency, healthcare, etc., whereas the user can access the blockchain through applications

Table 2 Layers of IoT

S. No.	Layers	Explanation
1.	Middleware layer	It is a base layer in which data is processed and stored
2.	Network layer	This layer performs the processing, transmission, data sending, retrieving, etc.
3.	Sensing/Perception layer	IoT has the data collector called sensors and actuators which helps to collect the data from the physical world to connect it into the visual world
4.	Application layer	These are end products, where the smart activities are done on many applications

4 Benefits from Blockchain to IoT

These two following figures depict in what way the IoT gets benefits from blockchain. The trilemma concept of blockchain is also introduced with three phases. The following are

A-B(decentralization) deals with availability and accessibility, consistency and transparency among all the participants [11].

A-C(Security) The main goals of security are confidentiality, integrity and availability. All the goals of the distributed ledger in the blockchain has been maintained with control mechanisms [4].

B-C(Scalability) This phase based on non-functional requirements like latency, throughput, etc., which depends upon the transaction processed in blockchain [1].

5 Conclusion

The drawback of IoT such as security of data, privacy, interoperability and scalability, are unknown to the people. They struggled to overcome the issues involved in IoT for many years. The blockchain is the solution for the demerits of IoT. The blockchain is new upcoming technology well known for its protection and many high qualities make everyone look aside on it. By integrating blockchain and IoT, there will be many advantages like trust, security, privacy, resiliency, transparency, data management, etc., and many other merits. In this paper, we have discussed the basics of blockchain and IoT separately and the disadvantages of IoT and how the IoT benefits from the Blockchain.

6 Future Scope

There is lot to be done in IoT, especially security, scalability and more. These all can be perfectly achieved using blockchain. So in future, many apps of BC-IoT will be in use.

References

1. Al Sadawi A, Hassan MS, Ndiaye M (2021) A review on the integration of blockchain and iot. In: 2020 International conference on communications, signal processing, and their applications (ICCSA), pp 1–6. <https://doi.org/10.1109/ICCSPA49915.2021.9385757>
2. Assiri A, Almagwash H (2018) IoT security and privacy issues. In: 2018 1st International conference on computer applications information security (ICCAIS), pp 1–5. <https://doi.org/10.1109/CAIS.2018.8442002>
3. Bhushan B, Sahoo C, Sinha P, Khamparia A (2021) Unification of blockchain and internet of things (biot): requirements, working model, challenges and future directions. *Wireless Networks* 27(1):55–90. <https://doi.org/10.1007/s11276-020-02445-6>
4. Bhushan B, Sharma N (2021) Transaction privacy preservations for blockchain technology. In: Gupta D, Khanna A, Bhattacharyya S, Hassanien AE, Anand S, Jaiswal A (eds) International conference on innovative computing and communications. Springer, Singapore, pp 377–393
5. Bhushan B, Sinha P, Sagayam KM, Andrew J (2021) Untangling blockchain technology: a survey on state of the art, security threats, privacy services, applications and future research directions. *Comput Electr Eng* 90:106897. <https://doi.org/10.1016/j.compeleceng.2020.106897>
6. Ferrag MA, Derdour M, Mukherjee M, Derhab A, Maglaras L, Janicke H (2019) Blockchain Technologies for the internet of things: research issues and challenges. *IEEE Internet Things J* 6(2):2188–2204. <https://doi.org/10.1109/JIOT.2018.2882794>
7. Hameed A, Alomary A (2019) Security issues in IoT: a survey. In: 2019 International conference on innovation and intelligence for informatics, computing, and technologies (3ICT), pp 1–5. <https://doi.org/10.1109/3ICT.2019.8910320>
8. Haque AKMB, Bhushan B, Dhiman G (2021) Conceptualizing smart city applications: requirements, architecture, security issues, and emerging trends. *Expert Syst*. <https://doi.org/10.1111/exsy.12753>
9. Latif S, Zafar NA (2017) A survey of security and privacy issues in IoT for smart cities. In: 2017 Fifth international conference on aerospace science engineering (ICASE), pp 1–5. <https://doi.org/10.1109/ICASE.2017.8374288>
10. Lockl J, Schlatt V, Schweizer A, Urbach N, Harth N (2020) Toward trust in internet of things ecosystems: design principles for blockchain-based IoT applications. *IEEE Trans Eng Manage* 67(4):1256–1270. <https://doi.org/10.1109/TEM.2020.2978014>
11. Saxena S, Bhushan B, Ahad MA (2021) Blockchain based solutions to secure IoT: background, integration trends and a way forward. *J Network Comput Appl* 181:103050. <https://doi.org/10.1016/j.jnca.2021.103050>
12. Tosh DK, Shetty S, Liang X, Kamhoua C, Njilla L (2017) Consensus protocols for blockchain-based data provenance: challenges and opportunities. In: 2017 IEEE 8th Annual ubiquitous computing, electronics and mobile communication conference (UEMCON), pp 469–474. <https://doi.org/10.1109/UEMCON.2017.8249088>

DAMONA: A Multi-robot System for Collection of Waste in Ocean and Sea



Riya Sil, Anwesha Das, Firdous Shamim, Aninda Chowdhury, Ritam Mukherjee, Sazid Ali, and Rohit Sharma

Abstract About 70% of the Earth's surface is covered with water. The abundance of water on the surface of the earth makes it vulnerable and an easily accessible open ground for the disposal of waste, thus making it tremendously exposed to pollution by anthropogenic activities. These wastes include unwanted oil spills from the commercial and industrial belts, litters from construction sites, oil spills from leakage of oil tankers, etc. The unwanted wastes in the water body threaten the aquatic ecosystem as it depletes the amount of dissolved oxygen in the water. Marine plastic pollution has impacted at least 267 species worldwide, including 86% of all sea turtle species, 44% of all seabird species, and 43% of all marine mammal species. In this paper, the authors have proposed an effective solution to the worldwide problem of surface water pollution. The main objective of this paper is to supply a multirobot cooperation system for the collection of surface water pollutants with the help of a mother-ship and small unmanned surface vehicles working in conjunction to achieve the goal.

Keywords Waste management · Water-waste · Robotics · Surface water waste

1 Introduction

The surface water, one of the most perceptible forms of water on earth, includes rivers, sea, ocean, etc. [1]. The water bodies are most susceptible to anthropogenic activities. These activities lead to the deposition of large quantities of waste (garbage) daily from industrial and municipal discharges [2]. Plastic is one of the most useful inventions, but these do not biodegrade; therefore, it floats and break into microplastics. Around 8–12 million tons of plastics are dumped in oceans every year [3]. The marine fauna

R. Sil (✉) · A. Das · F. Shamim · A. Chowdhury · R. Mukherjee · S. Ali
Department of Computer Science & Engineering, Adamas University, Kolkata, India

R. Sharma
Department of Electronics & Communication Engineering, SRM Institute of Science & Technology, Ghaziabad, India

eat this microplastic or even take them in during respiration [4]. But cleaning such huge bodies of water is a challenging task [5].

In this paper, the authors have proposed a system of distributed bots that are capable of coordinating with each other to clean a designated area of water effectively and efficiently without any human intervention other than collecting the gathered garbage from a common dumping vessel.

Section 2 discusses about the background study. Section 3 focuses on the related works done in this field. Section 4 thoroughly explains our proposed model and the performance analysis. Section 5 discusses about the result and analysis of the proposed model. Section 6 concludes our paper with a direction to future scope of this research work.

2 Background Study

According to a study conducted by Clarke [6], it is found that the majority of plastics found in the ocean are not plastic bottles and bags, rather the broken-down piece of these plastic objects due to the action of small critters, seawater corrosion, and ultraviolet lights from the sun. These broken-down pieces of plastic are known as microplastic and amount to around 92% of the total plastic found in the ocean. From a recent study [7], it was found that plastics can be found throughout the vertical length of the ocean. The devices that are to work in open waters (ocean, sea, etc.) have to deal with waves of an average height of 3 m [8].

One such phenomenon known as biofouling is where bacteria and diatoms produce a biofilm on rigid floating bodies, providing an ideal habitat for algae and protozoans which in turn attracts organisms such as sponges, mollusks, and crustaceans leading to an increase in weight of hundreds of kilograms per square meter of the surface that is submerged [9].

3 Related Research

In the related research section, the authors have investigated some of the existing solutions that provide similar services and listed their limitations in Table 1.

4 Proposed Model

Here, the authors have proposed a distributed system of floating waste collecting bots that communicate with each other to divide a marine litter-filled area into smaller quadrants that can be individually cleaned by bots. The authors have also outlined

Table 1 Comparison of research work

S. No.	Research paper	Description	Limitation
1	A Water Surface Cleaning Robot: A Floating Bot that Cleans Garbage [10]	The bot has an arm with a net attached to it with a belt that is submerged in water to collect floating wastes	Have no facility of autonomous mode and have not been tested in ocean water
2	Underwater Robotics: Surface Cleaning Techniques, Adhesion and Locomotion Systems [11]	It aims to collect waste from the water surface using arms creeping in various directions using motors and vacuum cups	The paper only provides the theoretical abstract model of the bot but no implementation of it
3	Unmanned Floating Waste Collecting Robot [12]	The paper is based on a Bluetooth-controlled bot system consisting of two propellers, two dc motors, battery, conveyor belt, collector box, sensors, etc.	It is manually operated as it has been accessed by remote. The communication system is not authentic
4	Ocean Cleanup System 001: An Autonomous Floating Plastic Capture System that Accumulates Floating Plastic [13]	A floating U-shaped system to catch and concentrate garbage	Dependent on ocean currents to collect the garbage
5	Waste Shark [14]	A semi-autonomous aqua drone to clean floating plastic and debris from the water surface	The mode of operation is confined to manual (currently) and can only be used on semi-confined water surfaces
6	Ocean Cleanup Interceptor [15]	A passive cleanup solution that is designed to be stationary and catch drifting floating garbage in a river	The passive solution, not suitable for open water use and can cause navigation problems for water vessels

rules that needed to be followed in case of under or over availability of bots for the job and in case malfunction.

4.1 Motivation and Objective

Given the location of the area of the patch of garbage is known, a portion of such a patch needs to be cleaned with a cost-effective, efficient, and semi-automated or automated solution. Such problems are being tackled by the ocean cleanup [13, 15] using giant floating nets that catch plastics and debris passively. But this approach is heavily dependent on the wave motion to guide the garbage into the net. Other solutions like WasteShark [14] use a single bot to move about with active locomotion facilities and catch garbage. They also have limited power delivery, resulting in less range and work time.

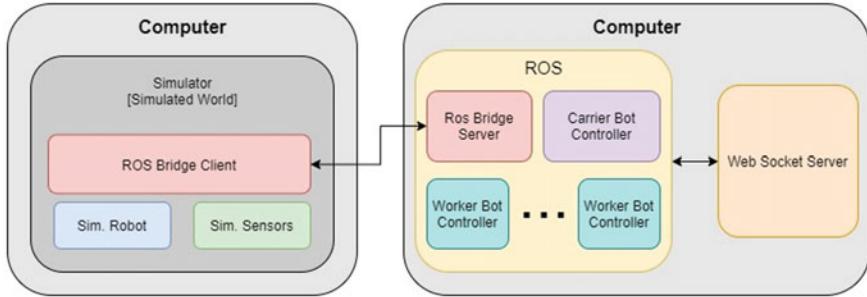


Fig. 1 System overview

4.2 System Architecture

The system described in Fig. 1 is divided into two sections, the simulator, and the controller. The simulator is the section that handles the simulation of the bots, garbage, sensors, and environment (gravity, ocean, winds, etc.). It gets control information for the bots and sends the sensor data through the ROS bridge client. This is a client library for C# used to interface with a ROS [16] environment on the computer or remotely using Web sockets. The corresponding module that serves as the server for the ROS Bridge client runs on the controller contains the carrier bot controller, worker bot controller(s), and a Web socket server [17].

The carrier bot controller is responsible to communicate and control the carrier bot simulation in the simulator [18]. Similarly, the worker bot controller(s) controls the worker bot(s) in the simulator. All the bots connect to the web socket server including the carrier bots. The Web socket server acts as a bookkeeper, which acts as the arbitrator between bots and is responsible for sector creation, sector assignment, and reassessments [19]. On the ROS side of the system architecture, we can see in Fig. 4 the use of topics and services. Each worker bot is using the publisher-subscriber model using topics to convey details as the current condition, position, and velocity [20].

4.3 Proposed Strategy

The proposed strategy is categorized into two sections: (i) cleaning strategy (ii) maneuver planner. The former distributes the assigned area equally into the number of agents in the system while the latter helps in planning the points for a smoother traversal of bots. Figure 2 shows the overview of internal architecture bot controller.

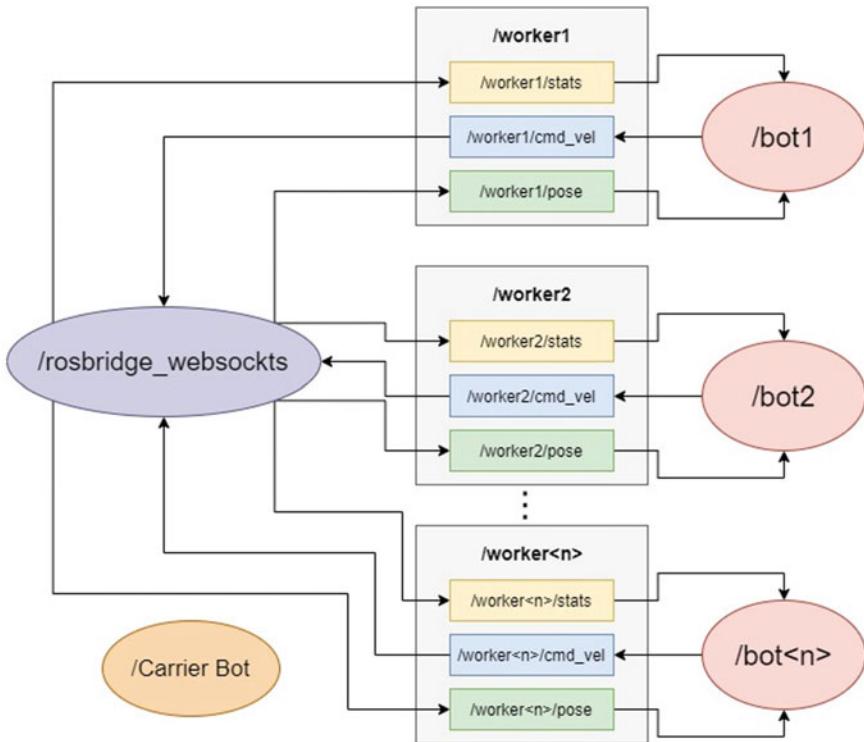


Fig. 2 Internal architectural overview of Bot controller

5 Simulator Environment and Robot Design

Here, authors have discussed about the utilization of simulator for the simulated environment and the creation, modeling, and texturing of the two bot types and the 3D software used for the various types of garbage that are usually found floating as marine litter.

5.1 Carrier Bot

Carrier bot acts as the mother of all the small bots as shown in Fig. 3. It is the controlling unit of all the worker bots. The bot is autonomous and eco-friendly [21].

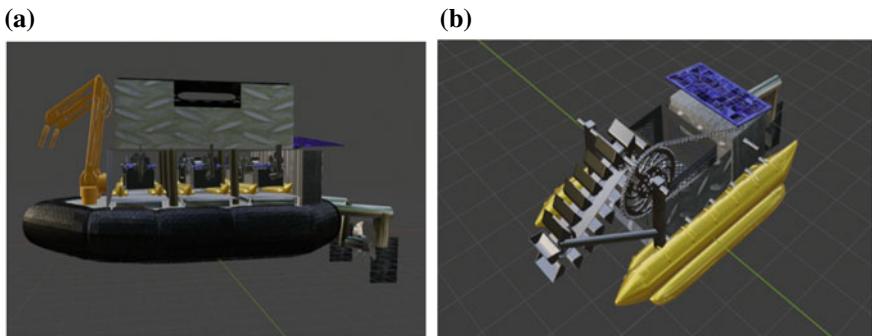


Fig. 3 **a** Carrier Bot design in blender **b** Worker Bot design in blender

5.2 Worker Bot

The worker bot as shown in Fig. 4 is the most important bot among the two bots because the whole waste collection is performed by it [22]. The bot is designed in such a way that it collects the waste more simply.

6 Result and Analysis

Figure 4 portrays the simulation conducted for validation of the proposed strategy. Here, we spawn the carrier bot, worker bot, and the garbage based on the request

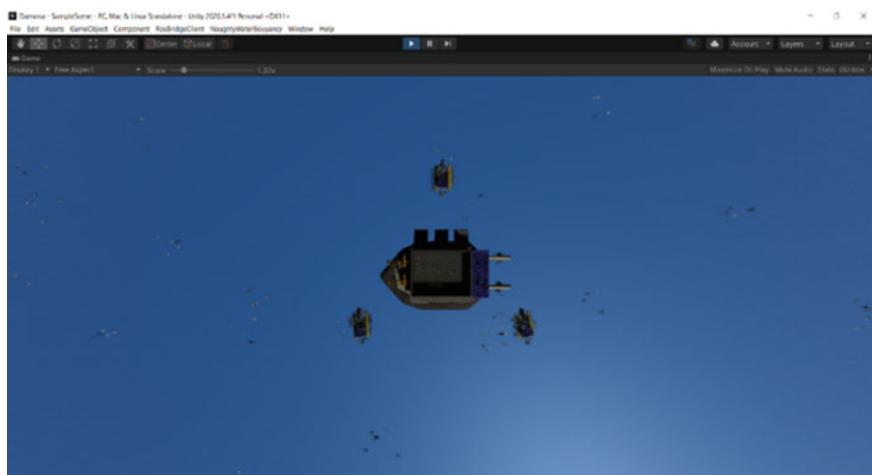


Fig. 4 Top view of the simulation environment with all of its component spawned

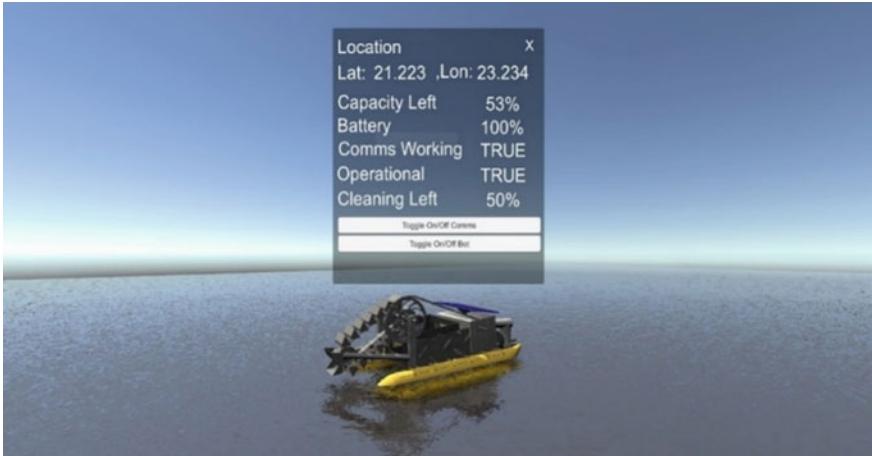


Fig. 5 Bot's UI representing its current stats and condition

sent from ROS [23]. The request specifies the required details like the location of each bot along with how many bots to be spawned, the radius of action, and such details which are needed to initialize the simulation environment. In Fig. 5, the code can be observed running and managing each bot's velocity and other parameters in real-time. This is reflected in the simulation environment [24]. The simulator also gives feedback about how each action is being performed via topics. The ROS environment and the scripts take the feedback and perform the necessary actions or decision-making [25].

6.1 Experimental Results

7 Conclusion and Future Scope

The proposed strategy can guide the bots to explore the area which is to be cleaned effectively. It also gives a formulated strategy to handle the division of work in multi-robot systems. The algorithm has been designed by abstracting a few key concepts like map formation, garbage detection in real-life scenarios via sensors, and such. The abstraction was done to solve a big problem by completing blocks of the puzzle one by one. But it is ensured that the abstracted concepts would not affect the working of the proposed algorithm and they can work after few adaptive measures. The algorithm has been written in Python, the simulated world was designed in Unity, and the bot is modeled in Blender. Further, the authors will improve the cooperation, design, garbage detection, and planning methods. Research and real-world tests need to be performed to be free from any software-induced boundary of simulators.

References

1. Water Science School. Surface Water, USGS.gov (n.d.) <https://www.usgs.gov/special-topic/water-science-school/science/surface-water>. Date of Access 11 June 2021
2. Moulder D (1983) Global marine pollution bibliography. Ocean dumping of municipal and industrial wastes. Mar Pollut Bull 14(9):362. [https://doi.org/10.1016/0025-326x\(83\)90404-6](https://doi.org/10.1016/0025-326x(83)90404-6)
3. Jambeck JR, Geyer R, Wilcox C, Siegler TR, Perryman M, Andrady A, Narayan R, Law KL (2015) Plastic waste inputs from land into the ocean. Science 347(6223):768–771. <https://doi.org/10.1126/science.1260352>
4. ScienceDaily (2016, April 13) Microplastics harm freshwater fauna. ScienceDaily. <https://www.sciencedaily.com/releases/2016/04/160413111224.htm>. Date of Access: 11 June 2021
5. Renteria A, Alvarez-de-los-Mozos E (2019) Human-Robot Collaboration as a new paradigm in circular economy for WEEE management. Proc Manuf 38:375–382. <https://doi.org/10.1016/j.promfg.2020.01.048>
6. Clarke C (2021, January 19) 6 Reasons that floating ocean plastic cleanup Gizmo is a Horrible Idea. KCET. <https://www.kcet.org/redefine/6-reasons-that-floating-ocean-plastic-clean-up-gizmo-is-a-horrible-idea>. Date of Access: 11 June 2021
7. Benton D (2017, April 7) Cleaning up the oceans is not a solution to the plastic problem. Inside track. <https://greenallianceblog.org.uk/2017/03/14/cleaning-up-the-oceans-is-not-a-solution-to-the-plastic-problem/>. Date of Access: 11 June 2021
8. Liu Z, Zhang Y, Yu X, Yuan C (2016, May 4) Unmanned surface vehicles: an overview of developments and challenges. Ann Rev Control. <https://www.sciencedirect.com/science/article/pii/S1367578816300219>
9. What Is Biofouling and How Can We Stop It? Marine Weather Data You Can Trust (n.d.) <https://www.sofarocean.com/posts/what-is-biofouling-and-how-can-we-stop-it>. Date of Access: 12 June 2021
10. Rahmawati E, Sucayah I, Asnawi A, Faris M, Taqwim MA, Mahendra D (2019) A water surface cleaning robot. J Phys: Conf Ser 1417:012006. <https://doi.org/10.1088/1742-6596/1417/1/012006>
11. Albitar H, Dandan K, Ananiev A, Kalaykov I (2016) Underwater robotics: surface cleaning technics, adhesion and locomotion systems. Int J Adv Rob Syst 13(1):7. <https://doi.org/10.5772/62060>
12. Akib A, Tasnim F, Biswas D, Hashem MB, Rahman K, Bhattacharjee A, Fattah SA (2019) Unmanned floating waste collecting robot. In: TENCON 2019 - 2019 IEEE Region 10 conference (TENCON). <https://doi.org/10.1109/tencon.2019.8929537>
13. Oceans. The ocean cleanup (2021, February 12) <https://theoceancleanup.com/oceans/>
14. Project: WasteShark—An autonomous catamaran to remove floating plastic debris in ports and harbours. DFKI GmbH (n.d.) <https://robotik.dfki-bremen.de/en/research/projects/wasteshark-1.html>
15. Rivers. The ocean cleanup (2021, June 4) <https://theoceancleanup.com/rivers/>
16. Dwhit (n.d.) dwhit/ros-sharp. GitHub. <https://github.com/dwhit/ros-sharp>. Date of Access: 12 June 2021
17. RaghavSaxena96 (n.d.) RaghavSaxena96/ROSBridge-Client. GitHub. <https://github.com/RaghavSaxena96/ROSBridge-Client>. Date of Access: 12 June 2021
18. Searchbot Simulator URL/User Agent Entry (n.d.) <https://botsimulator.com/>. Date of Access: 12 June 2021
19. Writing WebSocket servers—Web APIs: MDN. Web APIs, MDN (n.d.) <https://developer.mozilla.org/en-US/docs/Web/API/WebSocketsAPI/WritingWebSocketsServers>. Date of Access: 12 June 2021
20. EMQ Technologies Co., L (n.d.) Introduction to MQTT publish-subscribe model. emqx.io. <https://www.emqx.io/blog/mqtt-5-introduction-to-publish-subscribe-model>. Date of Access: 11 June 2021
21. Benefits of Robots. RobotWorx (n.d.) <https://www.robots.com/articles/benefits-of-robots>. Date of Access: 12 June 2021

22. Bots. Planetbase Wikia (n.d.) <https://planetbase.fandom.com/wiki/Bots>. Date of Access: 12 June 2021
23. Powering the world's robots. ROSorg (n.d.) <http://www.ros.org/>. Date of Access: 12 June 2021
24. Lateef F (2010, October) Simulation-based learning: Just like the real thing. *J Emergencies, Trauma, Shock.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2966567/>
25. Vinsloev (2021, January 30) Robotic software engineer — introduction to robotic operation system (ROS). Medium. <https://vinsloev.medium.com/robotic-software-engineer-introduction-to-robotic-operation-system-ros-da13b5f2a9ea>. Date of Access: 12 June 2021

Hardware Implementation and Comparison of OE Routing Algorithm with Extended XY Routing Algorithm for 2D Mesh on Network on Chip



Radha Velangi and S. S. Kerur

Abstract As per Moore's law, the number of transistors doubles every eighteen months this advance in technology grown to the level where SOC (System On Chip) became evident of this growth. This growth does not stop here further it turned into Network On Chip. But the bus technology which supported System On Chip cannot be used with Network On Chip Technology. Bus architecture when combined with NOC it was suffering from several issues like poor scalability, limited bandwidth and low resource utilization. To address these issues, routing algorithms and switching techniques are used. In this paper two such algorithms are compared. One is adaptive Odd even (OE) routing algorithms and another is Extended XY algorithm. It's been noticed that extended XY gives better performance in turns of Hop count and all the minimum paths are utilized in Extended XY routing algorithm unlike in OE few turns are prohibited.

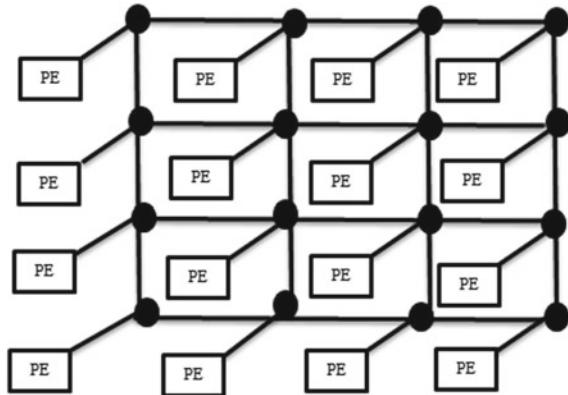
Keywords NOC (Network on Chip) · SOC (System on Chip) · OE (Odd Even) · XY · Extended XY

1 Introduction

System On Chip which had grabbed the market now getting turned onto Network on Chip due to the numerous advances in VLSI industry and need of more than million functionalities to be placed on single chip and carrying out the connectivity between them. When the core numbers were less it was bus architecture which extremely performed well but as the number of different cores began to increase the scalability

R. Velangi (✉)
Govt Polytechnic, Hubballi, India

S. S. Kerur
SDMCET, Dharwad, India

Fig. 1 4×4 Mesh NOC

became the bottle neck of bus architecture and hence switches and routing algorithms are the favourites of NOC.

A 4×4 mesh architecture is as shown in Fig. 1. It consists of PE (processing elements) and NI (Network Interface). Where PE can be any SOC or processor and NI actually holds the routing information and switching technique. Different topologies can be used to arrange the cores such as mesh, torus, star, etc., out of which mesh topology is the simplest and also more number of cores can be connected easily. Hence, mesh is preferred in NOC architecture [1].

Different switching techniques and routing algorithms can be used to administer the communication in NOC. Switching techniques are the one which decides data is moved in which format for example in packet switched network the data is moved from source to destination in packets where as in virtual cut through switching technique packets are further divided into flits and in circuit switched network before the data movement path between source and destination is decided [2].

Routing algorithm is used to route the data from source to destination smoothly by finding out the suitable output port for incoming input port. Routing algorithms can be divided into deterministic and adaptive, minimal routing and non-minimal routing, static and dynamic routing algorithms [3].

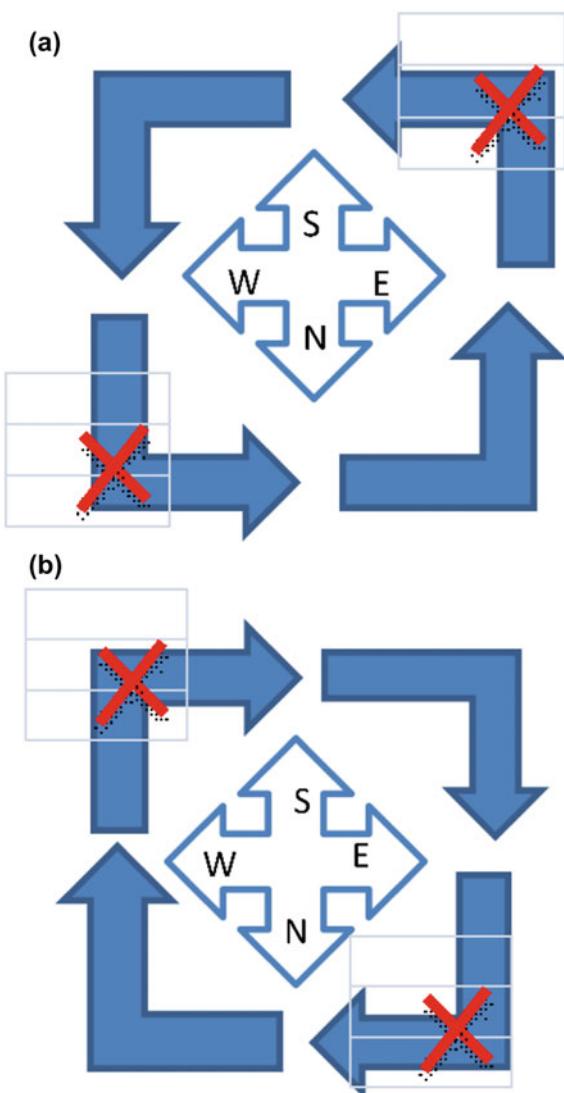
2 Literature Survey

“XY and OE routing algorithms”

In XY algorithm first, the X coordinator of the source and destination is matched, and then it matches with Y coordinator and there by finds the final destination. The direction of routing for XY direction is as shown in Fig. 2a, b.

Odd-Even routing algorithm is one of the adaptive routing algorithm. Unlike XY algorithm in OE algorithm few turns are prohibited. For even column East-North

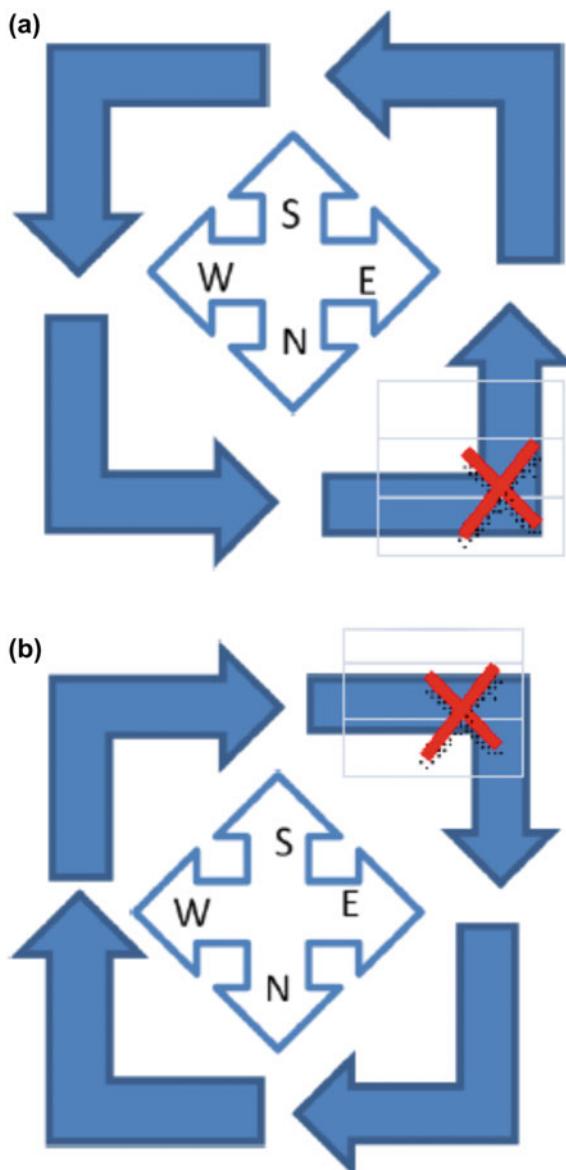
Fig. 2 Allowed turns in XY routing
a anti clockwise direction.
b Clockwise direction



and East-South turns are prohibited and for odd column North-West and South-West turns are prohibited this is shown in Figs. 3 and 4, respectively.

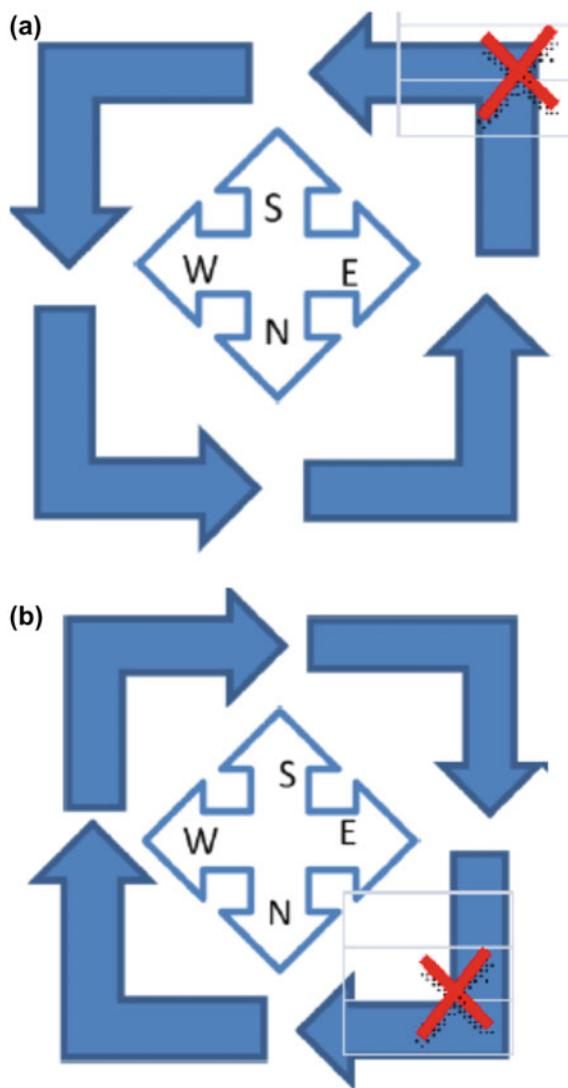
In Ref. [4] its shown that a 2-Dimensional 3×3 mesh topology NoC, XY routing algorithm and OE routing algorithm both are simulated and Percentage Load variation is compared and concluded that overall average latency of OE routing algorithm is less in comparison with XY routing algorithm, and in case of average throughput OE exhibits dominance over XY routing algorithm [4]. In Ref. [5], two-dimensional 2D-Mesh of 4×4 topology is chosen to compare packet delay and throughput of

Fig. 3 Prohibited turns in OE routing even column
a East South **b** East North



the two algorithm XY and OE. The results show that overall average delay of OE is less compared to XY and throughput is similar to that of XY. With NO dead lock and live locks. The results have shown that the ratio of OE routing algorithm is better compared with the ratio of XY routing algorithm. Hence, it has been concluded that the performance of OE routing algorithm is better in the two-dimensional Mesh topology [5].

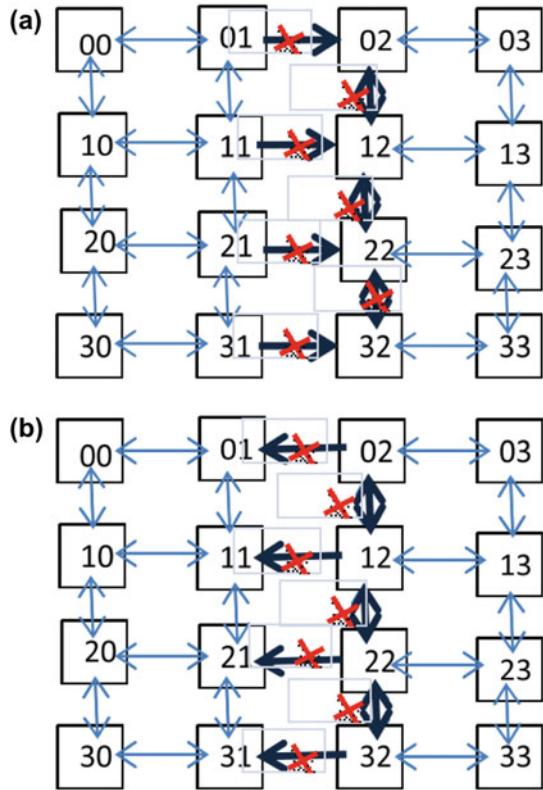
Fig. 4 Prohibited turns in OE routing odd column **a** South West **b** North West



On a 4×4 mesh using OE algorithm, for even column ES and EN turns are prohibited and for odd columns NW, and SW turns are prohibited which is as shown in Fig. 5.

As in Ref. [6] Extended 2D diametrical mesh can be considered as collection of four sub networks as shown in Fig. 6 where diagonally opposite corners of the each sub networks are connected directly. Like this added extra links contributes to reduce the diameter in 2D mesh when 2D mesh is expanded by a large number of IP cores. 4×4 diametrical 2D mesh with its sub networks and added extra links are as shown in Fig. 6a–e, respectively [6].

Fig. 5 Prohibited turns in OE routing algorithm **a** Even column **b** odd column



As per the extended XY algorithm depending on the source and destination nodes shortest path is calculated. If direct diagonal path available then diagonal path is chosen else normal XY routing algorithm followed its shown in Ref. [7] that the extra added e diametrical links in extended XY diametrical routing algorithm reduces the number of hops by providing the shortest possible path [7].

3 Methodology

OE Routing algorithm is designed using HDL where its designed for 4×4 mesh and core addresses are taken from 4'b0000 to 4'b1111 .rst signal is used to clear previous output once the source and destination address are available routing algorithm calculates the output path and if the prohibited turns encounters in path then next available path is chosen and the packets are forwarded. Module for OE algorithm is as shown below in Fig. 7.

Extended XY Routing algorithm is designed using HDL where its designed for 4X4 mesh and core addresses are addressed in same manner from 4'b0000 to 4'b1111

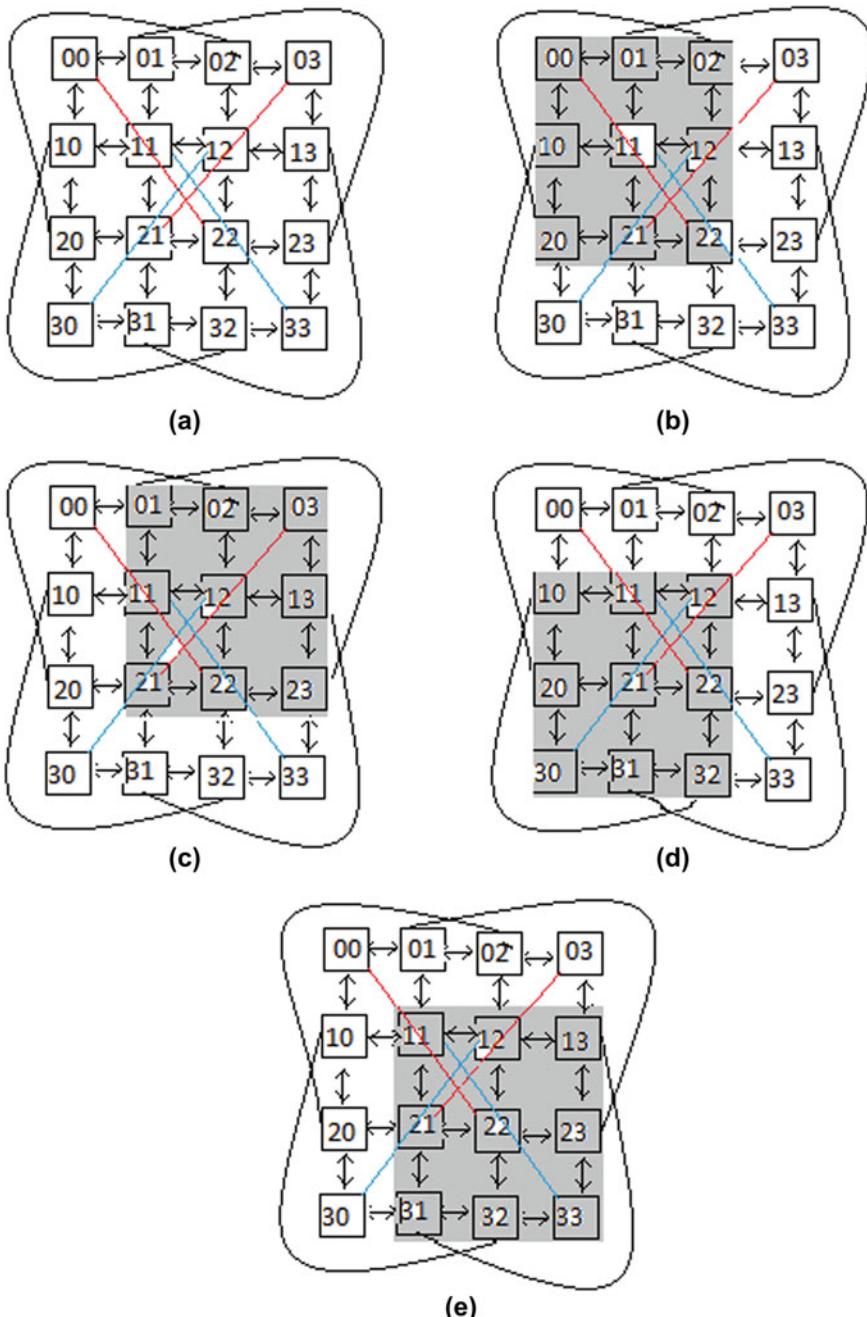
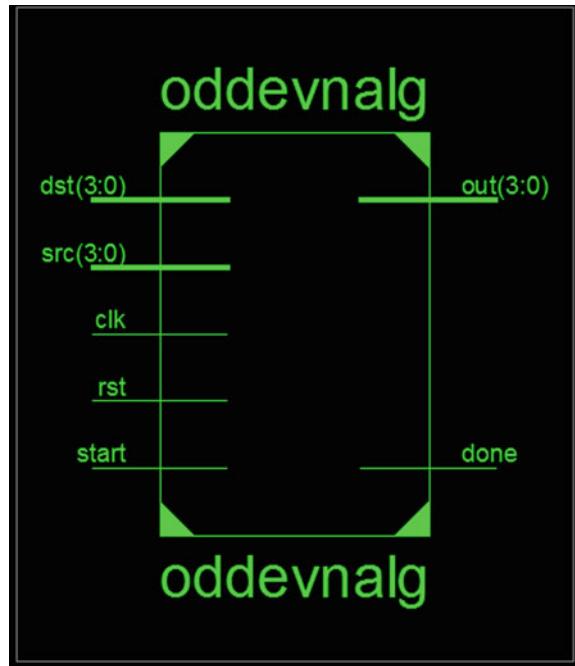


Fig. 6 **a** A 4×4 diametrical 2D mesh with 8 extra diametrical links. **b** First sub network with diagonal links from (00-22 & 20-02). **c** Second sub network with diagonal links from (01-23 & 03-21). **d** Third sub network with diagonal links from (10-32 & 30-12). **e** Fourth sub network with diagonal links from (11-33 & 31-13)

Fig. 7 OE routing algorithm module



.start signal is used to give green signal from router once it receives full packet once the source and destination address are available routing algorithm calculates the output path and prefers diagonal path if available else packet is routed through normal XY algorithm considering minimal path routing. Module for Extended XY algorithm is as shown below in Fig. 8.

4 Result Analysis

As shown in Table 1, where the source and destination with difference in path and hop counts chosen for OE and Extended XY routing algorithm it shows that Extended XY algorithms takes less hop count compared with OE algorithm. Total count of OE is 318 and that of Extended XY is 167. So the ratio is 0.52 which indicates in 50% cases Extended XY is choosing direct diametrical path in comparison to OE. Since unlike XY where there is possibility of load generation at centre which is avoided in OE by prohibiting few turns, by the below table its evident that even Extended XY also chooses alternate paths in 50% cases, and hence, it also avoids the load generation at the centre (Figs. 9 and 10).

Fig. 8 Extended XY routing algorithm module

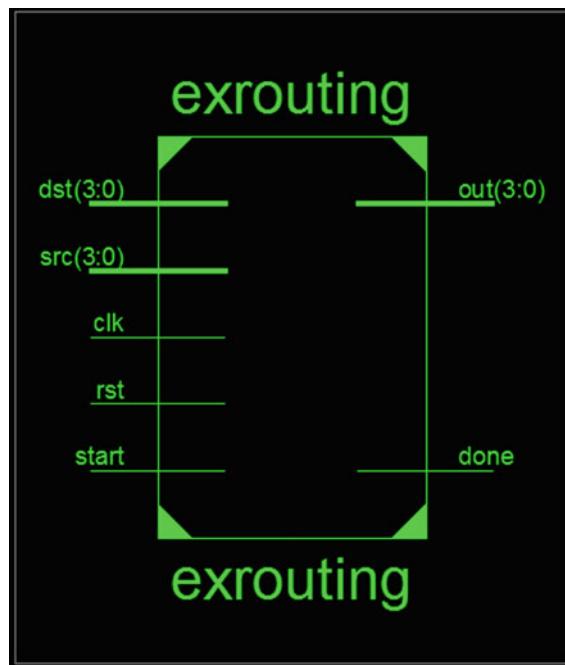


Table 1 Number of Hops taken by different source and destinations in OE and Extended XY algorithm

S. No.	Source–destination address	OE algorithm	No. of hops for XY algorithm	Extended XY	No. of hops for extended XY algorithm
	00-12	00-10-11-12	03	00-22-12	02
1	00-13	00-10-11-12-13	04	00-22-12-13	03
2	00-21	00-10-20-21	03	00-22-21	02
3	00-22	00-10-20-21-22	04	00-22	01
4	00-23	00-10-20-21-22	04	00-22-23	02
5	00-31	00-10-20-30-31	04	00-22-32-31	03
6	00-32	00-10-20-30-31-32	05	00-22-32	02
7	00-33	00-10-20-30-31-32-33	06	00-22-32-33	03
8	10-22	10-20-21-22	03	10-32-22	02
9	10-23	10-20-21-22-23	04	10-32-22-23	03
10	10-31	10-20-30-31	03	10-32-31	02
11	10-32	10-20-30-31-32	04	10-32	01
12	10-33	10-20-30-31-32-33	05	10-32-33	02

(continued)

Table 1 (continued)

S. No.	Source–destination address	OE algorithm	No. of hops for XY algorithm	Extended XY	No. of hops for extended XY algorithm
13	20-01	20-10-00-01	03	20-02-01	02
14	20-02	20-10-00-01-02	04	20-02	01
15	20-03	20-10-00-01-02-03	05	20-02-03	02
16	20-12	20-10-11-12	03	20-02-12	02
17	20-13	20-10-11-12-13	04	20-02-12-13	03
18	30-01	30-20-10-00-01	04	30-12-02-01	03
19	30-02	30-20-10-00-01-02	05	30-12-02	02
20	30-03	30-20-10-00-01-02-03	06	30-12-02-03	03
21	30-11	30-20-10-11	03	30-12-11	02
22	30-12	30-20-10-11-12	04	30-12	01
23	30-13	30-20-10-11-12-13	05	30-12-13	02
24	30-22	30-20-21-22	03	30-12-22	02
25	30-23	30-20-21-22-23	04	30-12-22-23	03
26	01-13	01-11-12-13	03	01-23-13	02
27	01-22	01-11-21-22	03	01-23-22	02
28	01-23	01-11-21-22-23	04	01-23	01
29	01-32	01-11-21-31-32	05	01-23-33-32	03
30	01-33	01-11-21-31-32-33	05	01-23-33	02
31	02-10	02-01-00-10	03	02-20-10	02
32	02-20	02-01-00-10-20	04	02-20	01
33	02-21	02-01-11-21	03	02-20-21	02
34	02-30	02-01-11-21-31-30	05	02-20-30	02
35	02-31	02-01-11-21-31	04	02-20-30-31	03
36	03-10	03-13-12-11-10	04	03-21-11-10	03
37	03-11	03-13-12-11	03	03-21-11	02
38	03-21	03-13-23-22-21	04	03-21	01
39	03-22	03-13-23-22	03	03-21-22	02
40	03-31	03-13-23-33-32-31	05	03-21-31	02
41	03-32	03-13-23-33-32	04	03-21-31-32	03
42	11-23	11-21-22-23	03	11-33-23	02
43	11-32	11-21-31-32	03	11-33-32	02
44	11-33	11-21-31-32-33	04	11-33	01
45	12-20	12-11-10-20	03	12-30-20	02
46	12-30	12-11-10-20-30	04	12-30	01
47	12-31	12-11-21-31	04	12-30-31	02

(continued)

Table 1 (continued)

S. No.	Source–destination address	OE algorithm	No. of hops for XY algorithm	Extended XY	No. of hops for extended XY algorithm
48	13-20	13-23-22-21-20	04	13-31-21-20	03
49	13-21	13-23-22-21	03	13-31-21	02
50	13-30	13-23-33-32-31-30	05	13-31-30	02
51	13-31	13-23-33-32-31	04	13-31	01
52	13-32	13-23-33-32	03	13-31-32	02
53	21-02	21-11-01-02	03	21-03-02	02
54	21-03	21-11-01-02-03	04	21-03	01
55	21-13	21-11-12-13	03	21-03-13	02
56	22-00	22-21-20-10-00	04	22-00	01
57	22-01	22-21-11-01	03	22-00-01	02
58	22-10	22-21-11-10	03	22-00-10	02
59	23-00	23-13-03-02-01-00	05	23-01-00	02
60	23-01	23-13-03-02-01	04	23-01	01
61	23-02	23-13-03-02	03	23-01-02	02
62	23-10	23-13-12-11-10	04	23-01-11-10	03
63	23-11	23-13-12-11	03	23-01-11	02
64	31-02	31-21-11-01-02	04	31-13-03-02	03
65	31-03	31-21-11-01-02-03	05	31-13-03	02
66	31-12	31-21-11-12	03	31-13-12	02
67	31-13	31-21-11-12-13	04	31-13	01
68	31-23	31-21-22-23	03	31-13-23	02
69	32-00	32-31-21-11-01-00	05	32-10-00	02
70	32-01	32-31-21-11-01	04	32-10-00-01	03
71	32-10	32-31-30-20-10	04	32-10	01
72	32-20	32-31-30-20	03	32-10-20	02
73	32-11	32-31-21-11	03	32-10-11	02
74	33-00	33-23-13-03-02-01-00	05	33-11-01-00	03
75	33-01	33-23-13-03-02-01	04	33-11-01	02
76	33-02	33-23-13-03-02	04	33-11-01-02	03
77	33-10	33-23-13-12-11-10	05	33-11-10	02
78	33-12	33-23-13-12	04	33-11-12	02
79	33-20	33-23-22-21-20	04	33-11-21-20	03
80	33-11	33-23-13-12-11	04	33-11	01
81	33-21	33-23-22-21	04	33-11-21	02



Fig. 9 OE routing algorithm waveforms

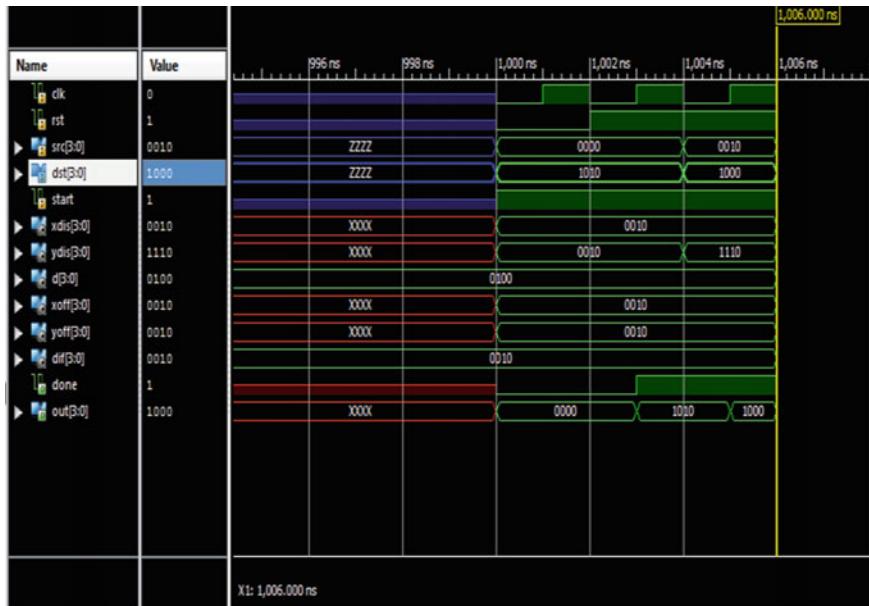


Fig. 10 Extended XY routing algorithm waveforms

5 Conclusion

From the carried, comparative analysis shows that Extended XY is more efficient than OE routing algorithm and takes less switching activity. In 50% cases of 4×4 mesh Extended XY routing algorithm is taking less hops to arrive at the destination and also avoiding load generation at the centre of the chip.

References

1. Pasricha S, Dutt N (2008) On chip communication architectures networks on chip, system on chip interconnect. Elsevier Publications
2. Mutha R (2011) Network on chip-design aspects. In: Innovative conference on embedded systems, mobile communication and computing (ICEMC2) 2011 Proceedings published by international journal of computer applications (IJCA)
3. Atienza D, Angiolini F, Murali S, Pullini A, Benini L, De Micheli G (2007) Network-on-chip design and synthesis outlook. *VLSI J* 41:340–359
4. Singh JK, Swain AK, Reddy TN, Mahapatra KK (2013) Performance evaluation of different routing algorithms in network on chip. In: 2013 IEEE Asia Pacific conference on postgraduate research in microelectronics and electronics
5. Hao P, QII H, Jiaqin D, Pan P (2011) Comparison of 2D MESH routing algorithm in NOC. 978-1-61284-193-9/11/\$26.00 ©2011 IEEE
6. Ghosal P, Karmakar S (2012) Diametrical mesh of tree (d2d-Mot) architectures novel routing solution for NOC. *Int J Adv Eng Technol* E-ISSN 0976-3945. IJAET/Vol.III/ Issue I/January-March, pp 243–247
7. Ghosal P, Das TS (2012) Improved extended XY on-chip routing in diametrical 2D mesh NOC. *Int J VLSI Des Commun Syst (VLSICS)* 3

Smart Medical Robotic Kit



Devesh Sonker, Akanksha, and Ranjeeta Yadav

Abstract The system is designed to keep a track of the patients' vitals as well as with the medication he is provided by monitoring the dosage, their timings and providing with the medication. The autonomous robot has a medication box for this purpose; it also receives user's commands as well as his vitals, the user has a heart rate monitoring sensor and a body temperature measurement sensor on a wrist band which is connected to the main controller and is connected with the help of an independent server, the vitals and the medication can be tracked with the help of an application for smartphones and various measures which can be taken to improve patient's health and the system uses the data to predict future hazards using the data accumulated in the database from the wrist band.

Keywords Computer vision · Image processing · Machine learning · Cloud database · Raspberry Pi · ROS

1 Introduction

The world is revolutionizing with artificial intelligence, and with new concepts such as deep learning and neural network having a breakthrough, it is possible for a machine to learn and act like human; the behavioral characteristics of autonomous robots are just a part of the beginning. They will soon be working in all such places where humans risk their lives; humans shall only supervise their jobs for quality control and management. With the ease in today's technology, the larger the database for a problem, better the inferences made by the model. Computer vision is

D. Sonker (✉) · Akanksha · R. Yadav
ABES Engineering College, Ghaziabad, Uttar Pradesh, India
e-mail: devesh.17bec1043@abes.ac.in

Akanksha
e-mail: akanksha.17bec1069@abes.ac.in

R. Yadav
e-mail: ranjeeta.yadav@abes.ac.in

an interdisciplinary field where visual inputs are provided to the database for testing and training of the model on various parameters, such as accuracy, size of the model, and normalization. Image processing uses algorithms to process an image with the help of a computer. These images are further stored in the database and used for learning purpose of the model.

With the technological node growing exponentially and the devices being developed in nanometers, it is possible that the future devices might be invisible to the naked eye due to their size. The system is a self-learning model which has preprocessed validation set and builds the database for accurate outcomes; the database keeps on updating itself on daily basis and keeps updating the cloud database for anomalies.

ROS is a platform which helps us build a network of devices, actuators, and sensors and multiplex their data. All the devices form clusters or nodes which can publish information and any node can simply subscribe to the information for actuation, planning, controlling, graphing, etc.

The main idea behind the project is to aid the patients which require more medical attention and are prone to ailments. The proposed system is a step-by-step implementation of a self-driving robot for specific tasks. Our system was created with the goal of assisting those who require medical assistance while also reducing human effort and contact. The concept proposes to complete some duties for the patient that necessitate human connection.

2 Literature Review

The need for autonomous robots in the healthcare sector has increased as per the current scenarios; automation and accuracy can be achieved by the use robots and their assistance. Limited contact can be maintained by the introduction of robots in this field. Various implementations have been seen over the years using different technologies to track the patients and provide timely medication to them. Some of them are mentioned below along with their references: The need for autonomous robots in the healthcare sector has increased as per the current scenarios; automation and accuracy can be achieved by the use of robots and their assistance.

Limited contact can be maintained by the introduction of robots in this field. Various implementations have been seen over the years using different technologies to track the patients and provide timely medication to them. Some of them are mentioned below along with their references:

In reference [5], the device has been designed for elderly people to provide medication on their prescribed timings. The device recognizes each medicine (pill and capsule) and keeps a record of the medication being taken by the patient. The patient is reminded by the device to take his medicine based on the preset timings.

In reference [4], a robot has been developed to determine the location of beds in a hospital ward where the robot uses computer vision to mark the beds, and along with this, the robot also keeps a track of the medication details and the dosage timings.

The robot now moves toward each bed to remind the patient to take the medicine at their respective times. The robot uses various algorithms to find the shortest path between the beds and thus taking measures.

In reference [1], the authors have discussed about a simple object detection and path correction algorithm with the help of a biped robot. The robot uses an ultrasonic and electronic compass to detect the information around it and use to avoid obstacles and correct its path moving along the course.

In reference [2], the robot uses a 2D LIDAR to map the surroundings in the form of spatial information. The robot uses convex hull algorithm to achieve obstacle detection and avoidance.

In reference [3], the authors have described various small physiological wearable sensors which can be used to monitor the vitals of a patient. The author states the purpose of these sensors to detect the anomalies at an early stage to reduce the risk of life endangerment. The authors have also discussed about the various modes of communication and transmission of data in the modules. Their purpose is to inform the researchers for future reference.

In reference [4], the authors have created a GSM-based monitoring system for the patients in ICU, their vitals through various sensors which measure parameters like blood pressure, body temperature, and heart rate. In the case the patient is not accompanied by medical personnel, the device triggers an alert message to a doctor and the hospital so that the patient can be taken care of by responding to the emergency. The monitored data is also stored in a local memory on the device so that the doctor may analyze the data.

In reference [5], a device has been developed for the patients which have Alzheimer's, as the patient tends to forget to take medication. This portable medication unit reminds the patient to take his medicine on prescribed timings.

In reference [6], a Bluetooth-based health tracking device has been developed which remotely monitors the patients vital such as heart rate and O₂ level through a mobile application in a visualized format.

In reference [7], a RFID-based system has been developed to keep a track of the patient and creates a database which is updated on each visit of the patient to a hospital. The device simultaneously records all the vital information from the sensors on the patient and updates them in the database.

3 Proposed System

The proposed system is a progressive implementation of autonomous robot for specified purposes. Our system has been designed to serve a purpose for people who need medical attention and to reduce the human effort and contact. The design proposes to accomplish certain tasks scheduled for the patient which require human interaction. It has various measures of input stream coming from different nodes of different format such as analog data from sensors, data over the local network created, and video feed from the camera.

The system is a model created as an individual identity which keeps on analyzing its surroundings and keeps on integrating them to itself. The model can analyze the previous data to predict what the instance might be. While encountering any new identity in the environment, it is supposed to acknowledge the parameters based on the health of the patient.

The system deals with the following points:

1. Measuring the temperature of the body.
2. Measuring the heartbeat or pulse.
3. Acquiring and logging all the information onto a local database.
4. Keeping a record of the medicines and alerting the patient for his dosage.
5. Avoiding obstacles in the path (robot).
6. Sensing the human and recording his face with the help of computer vision.
7. Using a passive infrared sensor to sense any human body present nearby.
8. A virtual data log presented to all those who want to view the patient's vitals (Fig. 1).

4 System Architecture

The system is divided into parts which are as follows:

- i. An autonomous robot
- ii. A wearable device.

The robot has a Raspberry pi 4 as the main master which controls the slave node as receives data over the Internet. The main controller Raspberry Pi has a 64-bit quad-core Cortex-A72 processor, 2 GB LPDDR4 ram, 2 micro HDMI ports, and a gigabit Ethernet port which enables it to connect to any network with the help of a RJ-45 cable. The device is accompanied with a display to represent all the necessary information. The Raspberry pi has a Wi-Fi module which makes it easy to connect to the slave node.

The robot has ultrasonic sensors attached to its head which gives a field of view of 180°. Along with this, a camera module is attached to it which enables to give us a live feed as well as implement algorithms of computer vision. A motor driver L298N drives the DC motors attached to the robot for its locomotion. A battery voltage regulator and distribution board to regulate and provide with the required voltage supply to the system. The camera is mounted on a stand which has two degrees of freedom which is provided with two micro servos, and can be controlled with the help of an external input changing the angle of the camera.

The slave node comprises of a SoC which is an Atmel ATmega-328P microprocessor along with a display and buttons. It is also provided with an ESP8266 12-E module. The microcontroller and oscillator are surface mount technologies to reduce the size of the node. The buttons have various features that are alerting the local authorities or activating the robot and traversing through the menu on the display. The surface mount technology reduces the surface area of the chip and thus reducing

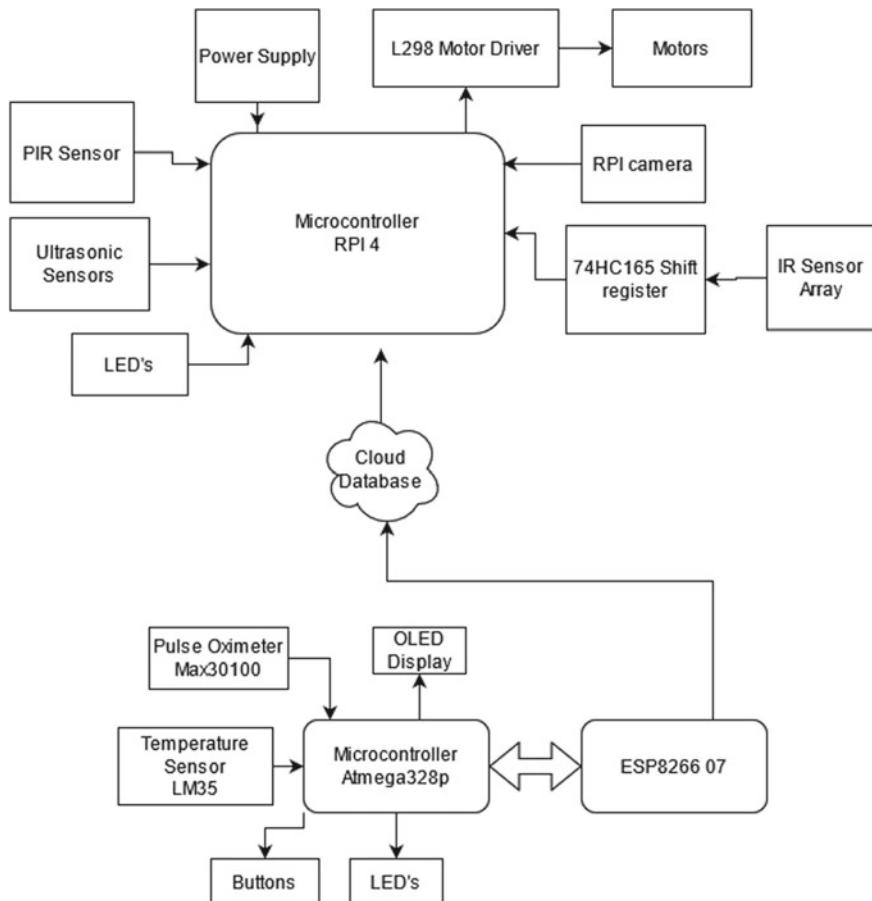


Fig. 1 Block diagram of the robot and wearable device

the size of the module. The module is connected to the main controller through a dedicated connection, i.e., through wireless fidelity, and the data is thereafter stored in a database which can be retrieved on various platforms with various extensions.

5 System Implementation

The robot has a Raspberry Pi mounted along with a motor driver connected to motors with wheels attached to their sides. The Raspberry pi is connected to two ultrasonic sensors, a passive infrared sensor, the medication box unit, LEDs as indicator, DC-DC buck boost convertor, and a LiPo battery for power supply.

The robot is called upon by a command received by the watch the patient wears; the command is simply an affirmation for the robot to move. The robot uses the ultrasonic sensors to detect the obstacles and subsequently follow an algorithm for path correction. The path correction algorithm is based on the obstacles on the left of the robot as well as in front of the robot. The robot determines the distance between the obstacle and itself to correct its path to reach the target.

The PIR sensor detects any human presence nearby it which makes it stop at that point. Thus, confirming the presence of a patient and task being completed.

Lastly, the RPI camera confirms the human presence by detecting a human in the camera frame or a human face. The technology used in achieving this is computer vision. This requires a library, i.e., OpenCV which is programmed using Python.

The data received by the robot is stored onto the local memory of the RPI and the updated into a database.

The database can be visualized onto a website or mobile application with the help of a simple fetch from the database. The motor driver is also enabled with the help of GPIO pins of the RPI. The motor driver uses an H bridge which controls the direction of the motors.

The above figure shows how the IR sensors would be connected to the 74HC595 Shift register. The usage of the 74HC595 reduces the usage of number of ports and simultaneously converts the data of all the individual IR transmitter-receiver pair from a parallel input scheme to a serial bit stream. A sensor consists of an IR led, An IR photodiode, and resistors. The output voltage is taken from a voltage divider bias. The high voltage value indicates that the compartment is empty whereas a low voltage would apply that the compartment has medicine in it (Figs. 2 and 3).

The serial data is easier to read and thus easier to process to determine which compartment of the box is empty or needs a refill. To each transmitter and receiver pair, a single sensor is attached to a compartment of the medicine box provided for the safe keeping of the medication. The medication box is filled with medicines with time. If in case any compartment is empty, the medication box notifies the patient that

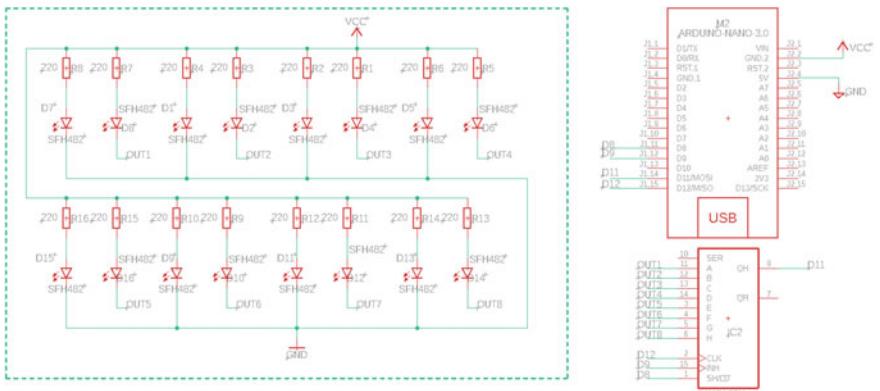


Fig. 2 Schematic diagram for IR sensor network using 74HC595

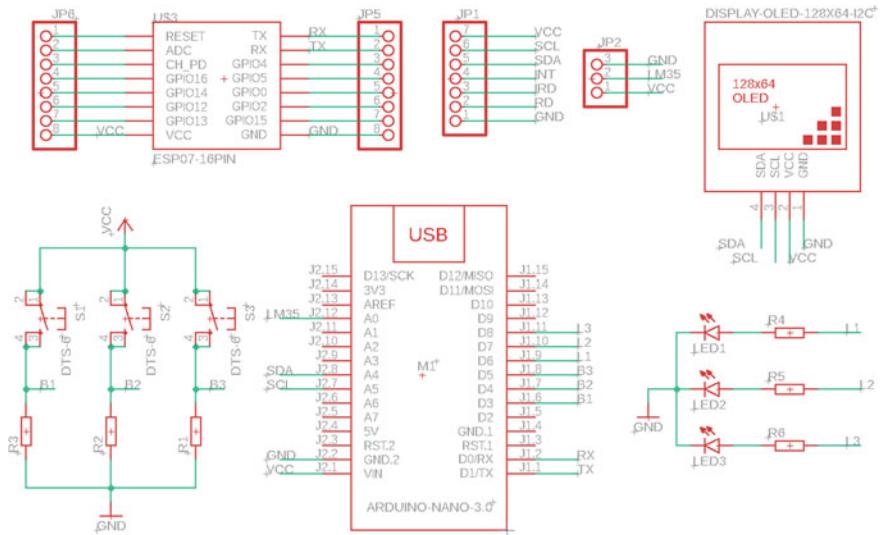


Fig. 3 Schematic diagram of the wearable device

a particular medicine compartment or multiple compartments are empty and need to be refilled.

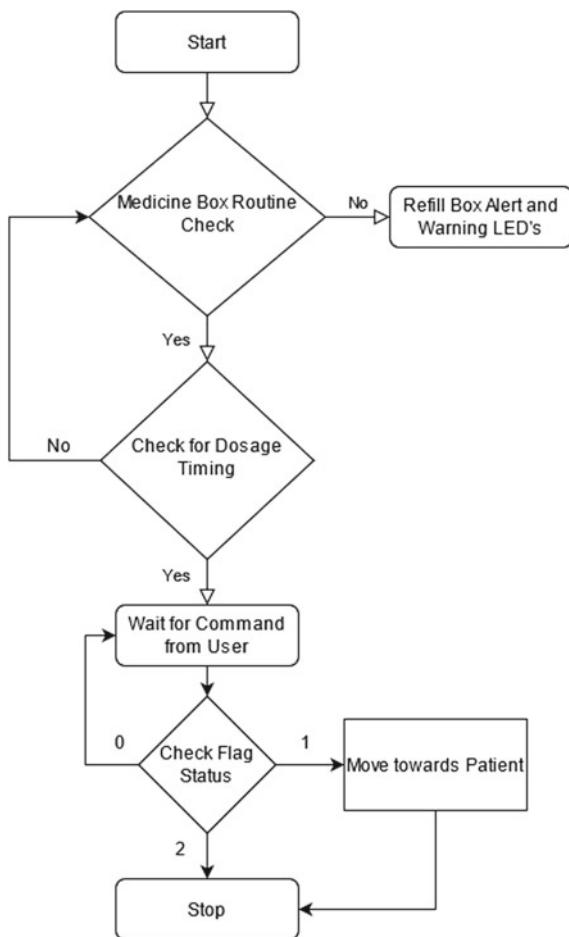
The wearable device has an onboard MCU which is an Atmega328p. The MCU is interfaced with an ESP8266 07 module to provide wireless functionality and as mode of communication. The sensors interfaced are an oximeter which measures the heart rate, oxygen saturation, and a temperature sensor which measures the body temperature. For the visualization of data, an OLED display is also provided along with LED indicators, and for the interaction of the patient with the device, buttons are provided. The data is read by the device and sent over to the autonomous robot which then processes the data and stores it. The buttons are also used to call the robot to the patient.

The wearable device sends off flag statuses to the main controller which determine the current state of the system. The statuses refer to the conditions in which it is idle, i.e., displaying the data of the vitals and simultaneously recording it. The other state being in motion when the autonomous robot is trying to reach the user. It also has a state of warnings in which it stops all operations to wait for the medication to be refilled (Figs. 4 and 5).

6 Conclusions

Our goal is to create a low-cost model that can create a localized map of an inbound area which can easily control any mechanized or electronic device to move from

Fig. 4 Flow chart for the autonomous robot



one point, i.e., source to another point, i.e., destination and in between avoiding all objects and obstacles and meanwhile analyze the surroundings with the help of a camera and other sensors and record the data. To prevent any harm by taking precautionary measures.

The Medicine Alert Unit focuses mainly on gathering the data from the sensors placed in position for each compartment which simply is each dose for the patient. Each sensor is a pair consisting of two IR LED's, an IR Led, and IR photodetector. The sensor data is converted from parallel inputs to a single streamline serial data which is easier to handle and process and subsequently reduces the number of I/O ports required by the microcontroller. This operation can also be termed as I/O expansion as the RPI GPIO are less, and to provide with more inputs, a PISO shift register is being used.

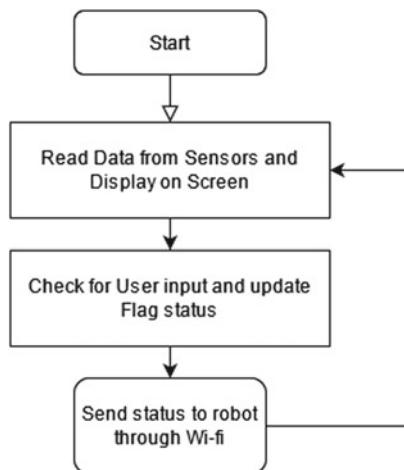


Fig. 5 Flow chart for the wearable device

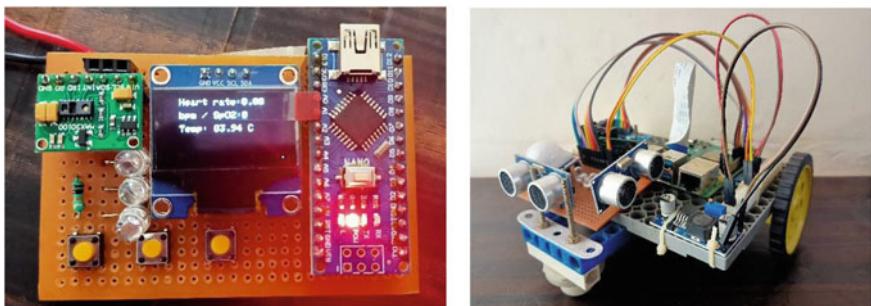


Fig. 6 The above figures are the images taken of the prototype implemented of the system

The autonomous robot can avoid and detect the obstacles, and it is able to correct its path. The device can identify any human being in its nearby region with the help of the PIR and camera module. The PIR sensor detects any human presence by measuring the difference in the temperature of the object in front of it. To make sure the medicine is being delivered to the correct user, face recognition algorithms can also be implemented in the system. The robot indicates any missing medication unless the device is preset (Fig. 6).

7 Future Potential

- The model can also be applied on drones and aerial vehicles.

- The model can easily reduce the hazard from areas where humans are prone to any risk.
- The model can be used in ICUs and other quarantine or isolation wards so that human contact can be minimized.
- The system is autonomous and can be implemented as a part of decision-making in neural networking and machine learning algorithms.
- The system has a great scope for its real time application in defense and military navigation algorithms and aerial path tracking.
- It can be implemented in medical institutions for its purpose of delivering medicine and reduce work load of the team by automating the tasks.

References

1. Suzuki T, Nakauchi Y (2010) Interactive medicine case system for elderly recipient. In: IEEE Conference paper
2. Fahimi F, Nataraj C, Ashrafiou H (2009) Real-time obstacle avoidance for multiple mobile robots. *Robotica* 27:189–198
3. Agarawala A, Greenberg S, Ho G (2004) The context-aware pill bottle and medication monitor. In: ACM Proceedings of 6th International conference on ubiquitous computing (UBICOMP 2004)
4. Jain NP, Jain PN, Agarkar TP (2013) An embedded, GSM based, multiparameter, real-time patient monitoring system and control—an implementation for ICU patients. In: IEEE Conference paper
5. Suzuki T, Jose Y, Nakauchi Y (2011) A medication support system for an elderly person based on intelligent environment technologies. In: IEEE Conference paper
6. Kamel M, Fawzy S, El-Bialy A, Kandil A (2011) Secure remote patient monitoring system. In: IEEE Conference paper
7. Joshi S, Dwivedi A (2020) RFID SYSTEM-used for monitoring and tracking patient (MTP). In: IEEE Conference paper
8. Vishal V, Gangopadhyay S (2018) CareBot: the automated caretaker system. In: IEEE Conference paper
9. Beer R, Keijers R, Shahid S, Mahmud A, Mubin O (2010) PMD: designing a portable medicine dispenser for persons suffering from Alzheimer's disease. In: Proceedings of the 12th International conference on computers helping people with special needs (ICCHP 2010), pp 332–335
10. Agarawala A, Greenberg S, Ho G (2004) The context-aware pill bottle and medication monitor. In: ACM Proceedings of the 6th International conference on ubiquitous computing (UBICOMP 2004)
11. Tsai C-H, Huang P-H, Cho W-C, Shih C-H, Chen C-Y, Shih B-Y, Chang C-J, Chen C-W, Chen T-H, Lee W-I (2010) Sonar-based obstacle avoidance using a path correction method for autonomous control of a biped robot for the learning stratification and performance. In: IEEE Conference paper
12. Jeong HK, Hyun KH, Kwak YK (2009) ELA+: goal-oriented navigation with obstacle avoidance for rescue robots. In: IEEE Conference paper
13. Ghorpade D, Thakare AD, Doiphode S (2017) Obstacle detection and avoidance algorithm for autonomous mobile robot using 2D LiDAR. In: IEEE Conference paper

Design of a Secure Blockchain-Based Toll-Tax Collection System



Debashis Das , Sourav Banerjee , and Utpal Biswas

Abstract The most advanced application of the Electronic Toll Collection (ETC) system is to collect the Toll-Tax Amount (TA) without slowing down a vehicle's speed at the toll plazas of the national highways. A few existing ETC systems suffer from various unexpected activities such as data security, transparency of the stored data, the privacy of users, and data immutability as these systems perform in a centralized platform. Blockchain is a secure technology for its nascent features such as decentralization, transparency, and data security. In this paper, a Blockchain-based Automated Toll-Tax Collection System (BATCS) has been proposed. The proposed system can collect an appropriate TA without stopping the vehicle while it passes the toll plaza. While vehicles cross the toll plaza, the predefined amount of tax will be deducted automatically from the bank account. The smart contract can authenticate the vehicle data and collect TA automatically at toll plazas. This research work provides Security, Trust, Transparency, and Privacy (STTP) in the field of the ETC system. The significant benefits of the BATCS concerning the RFID-based system are less fuel consumption and more time-saving for a vehicle. It also provides the zero-waiting time in the queuing line of the toll plazas.

Keywords Automated toll-tax collection system · Blockchain application · Decentralized electronic toll collection system · Intelligent internet of vehicular things · Smart contracts

D. Das · U. Biswas
University of Kalyani, Kalyani, India
e-mail: debashisdascse21@klyuniv.ac.in

U. Biswas
e-mail: utpalbiswas@klyuniv.ac.in

S. Banerjee
Kalyani Government Engineering College, Kalyani, India
e-mail: mr.sourav.banerjee@ieee.org

1 Introduction

The vehicle uses many types of intelligent devices to communicate and collect information from outside environments. The collected information can be useful for the Intelligent Internet of Vehicular Things (IIoVT) in traffic management. A fundamental problem is susceptibility to data. Therefore, there is a need to provide a novel solution that can provide Security, Trust, Transparency, and Privacy (STTP) to both communicating entities and secure vehicle data from malicious entities. Thus, Blockchain [1] can be incorporated with the Electronic Toll Collection (ETC) system [2] to collect an efficient Toll-Tax Amount (TA). This paper presents a Blockchain-based Automated Toll-tax Collection System (BATCS) to provide solutions to the challenges of the existing ETC system [3].

Blockchain is a decentralized, secure, transparent, and low-cost technology [4, 5]. It has a table of records named blocks, which are linked using cryptography. So, transactions can be secured and transparent to all associated organizations within the Blockchain network [6]. It is a new cutting-edge technology, where it provides different trusty solutions for various applications such as smart contracts, supply chains, home appliances, and healthcare management [1]. It allows users to form a collective agreement without involving middlemen or intermediaries such as a government and a company. Data privacy of Blockchain can provide confidentiality and transparency for all the sensitive data stored on it.

The number of vehicles has grown day by day. A vehicle needs to wait for a long time at toll plazas to pay TA. Thus, traffic congestion of the toll plazas has been risen using the hand-operated TA collection system. One of the existing ETC systems in India is Fastag [7] that can perform by using the Radio Frequency Identification (RFID) technology [8, 9]. However, there exist limitations, Alotaibi [10] which are the possibility of data modification, data security, fuel consumption, data privacy, and transparency of the transacted data over the internet. So, there is a need for an ETC system to avoid traffic congestion in the toll plaza. Thus, BATCS has been designed using Blockchain to address the existing limitations. It can also reduce waiting time and fuel consumption for the vehicle. Smart contracts have been used to authenticate the vehicle data and to collect TA automatically.

The major contributions of this proposed research work are summarized in the following:

- The BATCS framework has been introduced to collect the TA for the passing vehicle through the toll plaza without slowing down the speed.
- An algorithm has been provided that was implemented using the smart contract to verify the vehicle's data and to collect the TA automatically.
- An implementation flowchart of the proposed BATCS has been presented concisely in this paper.
- A comparative analysis of the proposed BATCS with the RFID-based system has been provided to compare the required time and fuel consumption.

The rest of this paper is summarized as follows. Section 2 describes the existing related works. In Sect. 3, the proposed framework has been described and implemented in detail. Section 4 provides experimental results and performance analysis of the BATCS framework with the existing RFID-based toll-tax collection system. Finally, Sect. 5 confers the conclusion and the future work.

2 Related Works

Automated Toll-Tax Collection System (ATCS) provides a significant contribution by minimizing the high traffic jams that have prompted the highways of crowded cities throughout the world. It is the easiest way to regulate congested traffic movements. The ETC system can find whether the vehicle has been registered or not and collect an appropriate TA, and then inform the administrator's facility about the activity of violation, transaction, and reconfiguring account.

Ito and Hiramoto [11] proposed a simulation-based ETC system, where they explained the traffic barrier. They also provided a solution to get reliable gate management and reconstructing a new design for toll plazas. The principal limitations of this system are the quality of the proposed model and irrelevant simulation time.

Hossain et al. [8] designed an RFID-based ATCS architecture. They provided a new advanced security feature that can counter the occurrence of possible crimes. In this system, the concerned authority can block a specific vehicle where the existing RFID-based system can detect RFID tags and collect TAs. But it suffers from data security, adoption of qualities, and inefficient features.

Nagothu proposed an ATCS applying the GPS and SIM of the GPRS [9] by using the longitude and latitude of the toll plaza's corner. A new method has been developed to find vehicles' accurate positions using the triangulation method. This system suffers from the inefficiency and weakness of the network's connectivity.

Ahmed et al. [12] implemented an RFID-based ATCS design. This system can solve traffic problems and maintain a little bit of transparency in the domain of the ETC system. It also reduces human error rates and manual labors. But it suffers from inadequate detection of tags, security and privacy issues, and equipment costs.

Jain et al. [13] proposed an ETC system using the OCR technology and RFID sensing to get the passing vehicle number and make a digital transaction using the unique identification. The main objective of this work was to reduce the over-collection of TAs and manage vehicle theft issues.

3 Blockchain-Based Automated Toll-Tax Collection System

This section describes the overall system framework of the proposed Blockchain-based Automated Toll-Tax Collection system (BATCS). The proposed framework is designed to collect the TA automatically. It can collect TA without stopping the

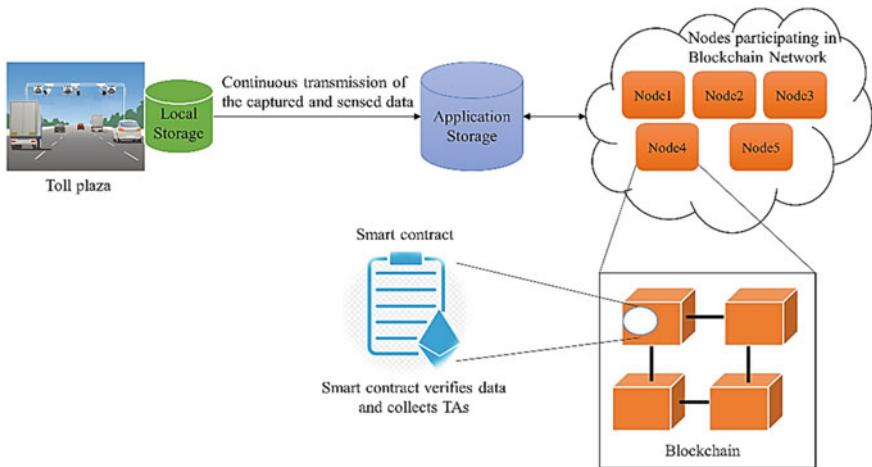


Fig. 1 System overview of the proposed framework

passing vehicle. In this framework, the Optical Character Recognition (OCR) method and image processing have been used to develop the BATCS. Smart contracts [14] have also been designed and implemented to verify the vehicle data [15]. The TA can be collected from the linked bank account of the vehicle owner based on the vehicle's type.

Figure 1 describes the system overview of the BATCS framework. The vehicle number plate is captured to retrieve the vehicle identification while it passes the toll plaza. Therefore, the captured data will be stored in the local storage of the toll plaza. The speed of the vehicle can be detected and stored on the same server. All the collected information will be transmitted to the application storage using the synchronous data transmission method. Every successful transaction will be stored on the Blockchain ledger. Finally, a notification will send to the vehicle's owner about the successful or failed transaction.

3.1 Methodology of the Vehicle Data Collection

Vehicle identification is required to collect the appropriate TA. The recognized vehicle data will be stored on the local storage of the toll plaza, and then, it will be sent to the Blockchain storage. The V_{Id} and V_{Speed} will be transmitted continuously to the Blockchain storage, where V_{Id} is the vehicle Id, and V_{Speed} is the vehicle speed. The vehicle speed can be detected using image processing methodology. Vehicle speed detection is required to know the V_{Speed} of the vehicle at the time of passing the toll plaza. In this proposed system, the randomly chosen threshold value of V_{Speed} shouldn't be more than 30 km/h. A vehicle number plate can be captured using the OCR technology while the vehicle will pass the toll plaza. The OCR method can

provide better identification of the V_{Id} . After processing the image, the V_{Id} is stored on the local server of the toll plaza and sent to the Blockchain application server frequently. Finally, the smart contract authenticates the V_{Id} and accesses all related data, and collects the TA.

3.2 Data Verification and TA Collection

An essential perspective of how smart contracts run in Ethereum is that they have their addresses in the Blockchain. If a node needs to access any methods defined by the smart contract, it can send a message to the address for the smart contract. Methods used in the smart contract can access data as input and return an output. Even methods used in the smart contract can store data. There is a need for a suitable mechanism to validate the received data after receiving it. Thus, the smart contract has been used in the BATCS framework. The TA can be collected based on the vehicle class, where the vehicle class is already associated with its vehicle number. Every vehicle has to be registered while buying it. Every vehicle owner needs to provide their identification and account details at the time of registration. The smart contract can identify the vehicle and the linked account number using the V_{Id} to collect the TA. The smart contract can retrieve the vehicle's class, the vehicle owner account, and account balance using the V_{Id} . All the information has already been taken at the time of registration and stored in Blockchain. Finally, the smart contract will collect the TA by accessing the V_{Class} and $AC_{Balance}$, where the V_{Class} is the vehicle class, and $AC_{Balance}$ is the account balance of the vehicle owner account.

Figure 2 illustrates the workflow diagram of the proposed BATCS framework. A fixed amount will be collected as a penalty when the V_{Speed} will be greater than 30. If the required TA is not available, the count value of Insufficient Balance (IB_{Count}) will be set to 1, and a reminder message will be sent to the vehicle's owner. Concerning this message, the vehicle owner should credit the account within a given time by the appropriate authority. If the IB_{Count} will be greater than 1, then that V_{Id} will be sent to the blacklist.

4 Experiment Results and Comparison Analysis

This section illustrates the experimental results of the incorporated smart contract of the proposed research work, which has the most vital role in the proposed BATCS framework. The smart contract is implemented and tested by accessing the Remix IDE. Smart contracts code can write and experiment with within solidity language using Remix that is a web-based platform.

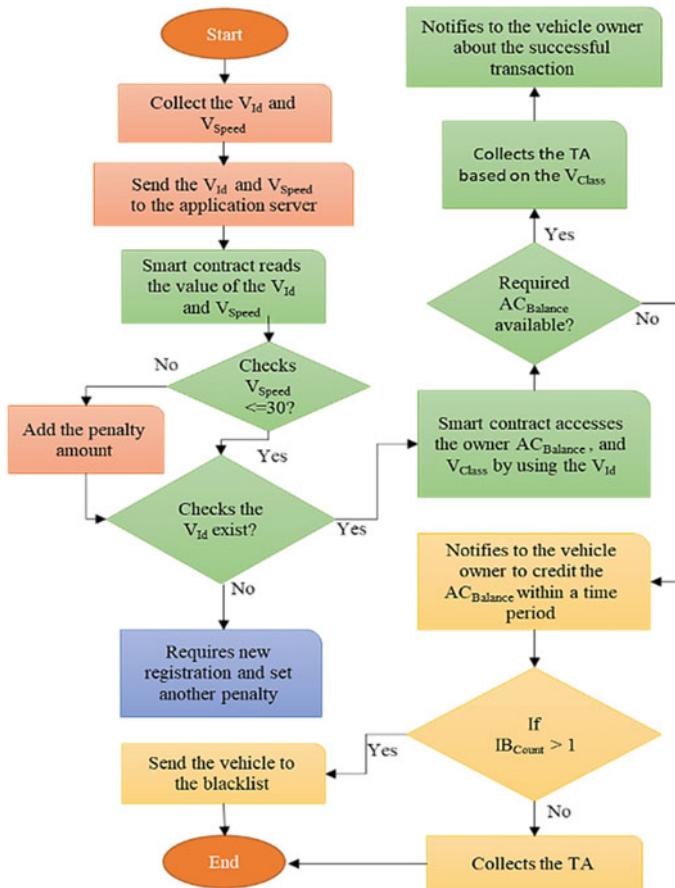


Fig. 2 Workflow diagram of the proposed framework

4.1 Required Time Analysis

Figure 3 shows that less time is needed using the proposed BATCS framework. In this system, the vehicle needs to pass the toll plaza at a threshold speed, where the maximum speed is 30 km/h. So, more time can save for the passing vehicle through the toll plaza. Thus, the proposed BATCS takes less time rather than RFID-based systems.

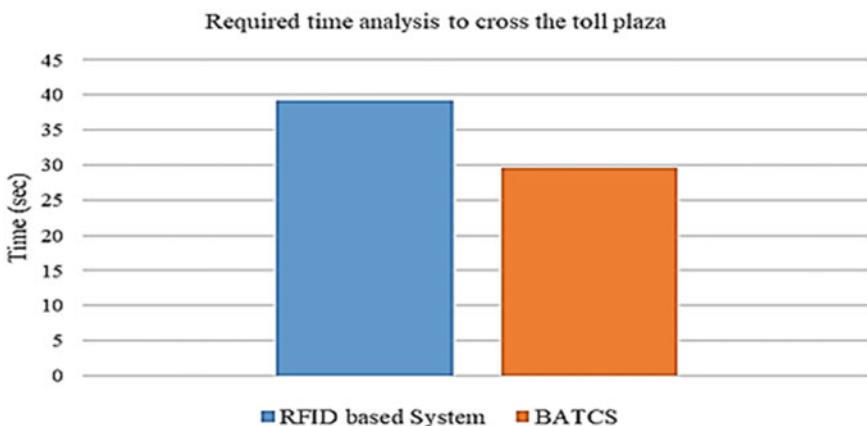


Fig. 3 Required time by a vehicle to cross the toll plaza

4.2 Fuel Consumption Analysis

Fuel consumption can be reduced using the BATCS concerning the RFID-based system that is shown in Fig. 4. The RFID-based system takes time to read the RFID tag and open the gate after scanning the tag. But, using the proposed BATCS, vehicles don't have to wait at the toll plaza. In this case, vehicles pass the toll plaza, and TA will be collected instantly.

Table 1 represents that the proposed framework is better than the RFID-based system after analyzing the required time and fuel consumption by a vehicle to cross the toll plaza.

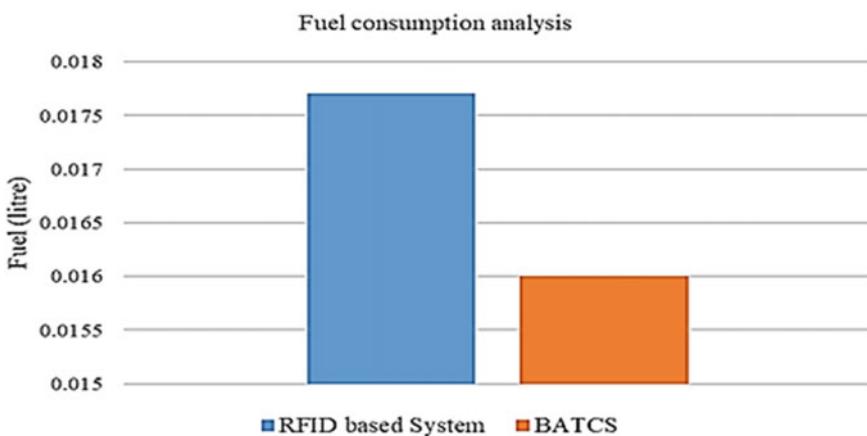


Fig. 4 Fuel consumption by a vehicle to cross the toll plaza

Table 1 Cost analysis of the BATCS with the RFID-based system

	RFID-based system	BATCS	Improvement by BATCS (%)
Required time (second)	39.24	29.58	32.65
Fuel consumption (liter)	0.0177	0.0160	10.62

5 Conclusion and Future Work

The ETC system is becoming the most common way for passengers to travel by bridge and highway. The ETC mechanism is a robust and fault-tolerant method to collect the TA at the toll plaza. The existing ETCs have data security, transparency, and privacy issues that can be eliminated using the proposed BATCS framework. It can give an adequate solution for the ETC system as it is a transparent, secure, and privacy-preserving framework. It can save a lot of time and fuel consumption. Every node participating in the Blockchain network can perceive the TA collection transactions. But, in the BATCS framework, it is not possible to modify or delete the data, as it is a decentralized platform. In the proposed system, an appropriate TA can be collected for the passing vehicle securely and transparently without having the toll barrier system at the toll plaza of the national highways. It can provide zero waiting time in the toll plazas. In the BATCS, stored data cannot be altered, as it is designed using Blockchain technology. Thus, data security can be achieved using the BATCS. Blockchain can provide a stable and profitable TA collection system in the domain of the ETC system. In future, we want to extend this work concerning vehicle security issues and to provide an efficient transport system for the smart city.

References

- Das D, Banerjee S, Biswas U (2021) A secure vehicle theft detection framework using Blockchain and smart contract. Peer-to-Peer Netw Appl 14:672–686. <https://doi.org/10.1007/s12083-020-01022-0>
- Inserra D, Hu W, Wen G (2018) Antenna array synthesis for RFID-based electronic toll collection. IEEE Trans Antennas Propag 66(9):4596–4605. <https://doi.org/10.1109/tap.2018.2851292>
- Arjroody AR, Ameri M, Hasheminejad SM (2009) Cost-benefit analysis of electronic toll collection (ETC) system in Iranian freeways (case study: Tehran-Qom Freeway)
- Banerjee S, Das D, Biswas M, Biswas U (2020) Study and survey on blockchain privacy and security issues. In: Williams I (ed) Cross industry use of blockchain technology and opportunities for the future. IGI Global, pp 80–102. <https://doi.org/10.4018/978-1-7998-3632-2.ch005>
- Das D, Banerjee S, Ghosh U, Biswas U, Bashir AK (2021) A decentralized vehicle anti-theft system using Blockchain and smart contract. Peer-to-Peer Netw Appl. <https://doi.org/10.1007/s12083-021-01097-3>

6. Jiang T, Fang H, Wang H (2019) Blockchain-based internet of vehicles: distributed network architecture and performance analysis. *IEEE Internet Things J* 6(3):4640–4649. <https://doi.org/10.1109/IJOT.2018.2874398>
7. FASTag: Electronic Toll Collection. <https://www.fastag.org/fasttag>. Last accessed 7 Feb 2020
8. Hossain R, Ahmed M, Alfasani MM, Zaman HU (2017) An advanced security system integrated with RFID based automated toll collection system. In: 2017 Third Asian conference on defence technology (ACDT), Phuket, pp 59–64. <https://doi.org/10.1109/ACDT.2017.7886158>
9. Nagothu SK (2016) Automated toll collection system using GPS and GPRS. In: 2016 International conference on communication and signal processing (ICCSP), Melmaruvathur, pp 0651–0653. <https://doi.org/10.1109/ICCSP.2016.7754222>
10. Alotaibi B (2019) Utilizing blockchain to overcome cyber security concerns in the internet of things: a review. *IEEE Sens J* 19(23):10953–10971. <https://doi.org/10.1109/JSEN.2019.2935035>
11. Ito T, Hiramoto T (2006) A general simulator approach to ETC toll traffic congestion. *J Intell Manuf* 17(5):597–607. <https://doi.org/10.1007/s10845-006-0023-3>
12. Ahmed S, Tan TM, Mondol AM, Alam Z, Nawal N, Uddin J (2019) Automated toll collection system based on RFID sensor. In: 2019 International Carnahan conference on security technology (ICCST), Chennai, India, pp 1–3. <https://doi.org/10.1109/CCST.2019.8888429>
13. Jain P, Dhillon P, Singh AV, Vats K, Tripathi S (2018) A unique identity based automated toll collection system using RFID and image processing. In: 2018 International conference on computing, power and communication technologies (GUCON), Greater Noida, Uttar Pradesh, India, pp 988–991. <https://doi.org/10.1109/GUCON.2018.8675073>
14. Hartel P, Homoliak I, Reijsergen D (2019) An Empirical study into the success of listed smart contracts in ethereum. *IEEE Access* 7:177539–177555. <https://doi.org/10.1109/ACCESS.2019.2957284>
15. Paik H, Xu X, Bandara HMND, Lee SU, Lo SK (2019) Analysis of data management in blockchain-based systems: from architecture to governance. *IEEE Access* 7:186091–186107. <https://doi.org/10.1109/access.2019.2961404>

An Overview of Security Services and Trust-Based Authentication Schemes in VANET



M. Gayathri and C. Gomathy

Abstract Communication has played a very significant role in mobile communication. MANET Mobile Ad Hoc Network which uses mobile nodes in a distributed way to communicate with all mobile nodes to reach the user. This mode of communication has caused the user to share information, get unknown information from the user, get a solution for a problem in real time. By using the principle of Mobile of Ad Hoc network a new technique called VANET, Vehicular Ad Hoc network was introduced which provides a way to Intelligent Transportation System. VANET uses a wireless medium for communication between vehicles. All messages which are transmitted are broadcasted messages so there is a huge opportunity for hackers or attackers to hack the communication. These attackers can drop packets that are kept for transmission, modify the messages and delay the communication. All the communication which are done in VANET is based on a trust mechanism because the communication is carried out by one node to an unknown vehicle which purely relies on the trust factor of that vehicle. Secure communication is one of the major issues to be concentrated on VANET. In this paper, different types of VANET Architecture, Security mechanism, Cryptographic Technique, Trust-Based Authentication schemes are discussed in detail.

Keywords VANET · Trust · DSRC · WAVE · ARIB · CALM · C2C

M. Gayathri · C. Gomathy

Electronics and Communication Engineering, College of Engineering and Technology,
SRM Institute of Science and Technology, Vadapalani Campus, No: 01, Jawaharlal Nehru Road,
Vadapalani, Chennai, Tamil Nadu, India
e-mail: gm0717@srmist.edu.in

C. Gomathy
e-mail: gomathyc@srmist.edu.in

1 Introduction

VANETs are used to provide communication between vehicles or unknown nodes. Through this communication, the user can contact or communicate with a nearby node, unknown node at anytime and anywhere. This mode of communication will reduce the traffic in urban areas by inducing traffic alerts to nodes, reduce accidents by providing a pre-alert warning to other nodes, it can also increase the safety of drivers and passengers in all hectic times, and hence, VANETs are known to be a future of intelligent transportation. VANETs can be used in safety applications and for entertainment purposes also [1]. The main question which arises when we talk about VANET is how communication is possible between vehicles during mobility. VANETs communication depends upon the pre-registration mechanism of vehicles with roadside units and onboard units. Roadside units are placed along the sides of the road to provide internet connectivity to the user. Onboard units are placed inside every vehicle which acts as a radio transmitter to communicate with the other nodes [2]. Since all the messages are transmitted through the wireless mode of communication to an unknown node. The data which are sent should be more secured and confidential. Hence, trusted authorities are used which is responsible for registering all trusted nodes with their OBU identity. It is also responsible for authenticating the user id of incoming and outgoing nodes to prevent malicious nodes to enter. Event data recorder (EDR) is one of the components of a vehicle that acts as a black box of a vehicle to track all the activities of the vehicle. It records the entire event which has happened before and after accidents. VANETs communicate using a dedicated short-range communication protocol (DSRC). Here, the communication is provided in between the vehicles which have high mobility in a single lane, two-lane and multi-lane in roads. VANET users can communicate with other nodes either by single hop or by multi hop if communication distance is increased. To have safe and secure communication trust plays a significant factor. This paper is organized into a different section, Sect. 2 explains about characteristics and limitation of VANET, Sect. 3 discusses VANET architecture, Sect. 4 explains about various security mechanisms in VANET, Sect. 5 concentrates on cryptographic technique, Sects. 6 and 7 concentrate on various trust models and trust-based authentication scheme in VANET, respectively.

2 Characteristics and Limitations of VANET

The major characteristics and features of VANET are depended on its behavior in real-time application. VANETs have high mobility in nature [3, 4] which causes some difficulties in routing scenarios in urban areas, effective diameter is used between vehicle communication, since every vehicle moves from one place to another it has dynamic network topology and high-computational ability [5]. The main limitation

of VANET is frequent disconnection of the network, fading of signals, limitation of bandwidth and time constraints.

3 VANET Architecture

The architecture definition of VANET was provided by many standard organizations such as ISO and IEEE. These standards include WAVE, ARIB, C2C, CALM [6].

WAVE (Wireless Access in Vehicular Environment) IEEE has started its operation using Wireless Access in Vehicular Environment (WAVE) which uses Dedicated Short range communication protocol (DSRC) [7–9]. This protocol is very useful for providing higher connectivity, used for safety applications in VANET. These services include standards that are recognized by the U.S National Intelligent Transportation system [10]. WAVE system supports many applications that in turn provide vehicle systems and drivers with real-time events, threats, tracking of vehicle accidents and also used for user safety.

ARIB (The Association of Radio Industries and Businesses) This ARIB was introduced to expand and develop the utilization of radio waves for the development of telecommunication systems and increase in transmission of broadcasted messages. It mainly focuses on emergency messages broadcasting in VANET [11].

C2C (Car-2-Car Communication Consortium) The automobile industry in Europe introduced this C2C VANET Standard. It is a slightly modified version of IEEE 802.11p. The physical layer focuses on DSRC and WLAN Standards. It focuses mainly on safety requirements in VANET. This requirement has led to the development of a protocol stack in Geographical packet distribution, data aggregation, congestion control in cross-layer and security services.

CALM (Communications, Air-interface, and Long and Medium range) [12, 13] This standard focuses on multiple media. It is the combination of different technologies in the physical layer. It communicates both in unicast and multicast. They deal with uninterrupted nodes and communication in VANET.

4 Various Security Mechanism in VANET

Security plays a prominent role in wireless communication to protect the information being theft or steeled by hackers. Security mechanisms can be classified into five categories [4].

Authentication: The identity verification of a sender and receiver is very important in node-to-node communication. In VANET, authentication is meant to verify the genuine node, lead it for communication and prevent unauthorized nodes to access communication. Malicious nodes can mislead the communication by injecting false messages, inserting unwanted messages in the communication path, false GPS spoofing which may lead to traffic, information theft and divert the path of a user

and so on. Hence, authentication of nodes is a significant process in communication that needs to be done to avoid unauthorized access which may collapse the entire infrastructure of safe communication.

Availability: The node for communication should be available all the time even if the node is infected. Only if the node is available for the communication, the sender node, roadside unit or trusted authorities will be able to reach the node to transfer information about which is trusted node and malicious node, exchange emergency information and so on.

Integrity: The originality of the message being transmitted must be maintained till it is received by the user [2, 3]. The messages which are sent by the sender can be altered by hackers. The hackers can reframe the messages and enclose false information in the packet and send it to the receiver. This may cause high information piracy.

Confidentiality: The messages which are sent by a node can be classified into many types as Emergency alert messages, Safety messages, Entertainment application messages, Banking or transaction messages. Entertainment application messages need not be worried about confidentiality, but emergency messages, banking or safety messages should be protected from being unchanged from their originality by hackers. To maintain the message confidential encryption techniques are used. The messages which are forwarded by the sender will be encrypted into a form with a password, these encrypted messages will be decrypted by the receiver at receiving node by the usage of the same password which is known to only sender and receiver.

Non-Repudiation: This is another method to secure messages by digitally signing the documents. Before initiating a process, the documents are legally got signed by the sender to ensure only he is transferring the messages to the receiver node. The sender will take all the responsibility for whatever message is transmitted in the communication mode. The sender could not deny that he has not forwarded the message to any node [14, 15]. Table 1 classifies the various types of attacks based on the security mechanism.

Digital certificates: Digital certificates are used for sharing the public key which is used for encryption in the authentication process. These certificates are legally signed and kept up to date by the certification authority [5, 6].

Digital signatures: Digital signatures are used in the authentication process to verify that only the trusted sender has sent the messages, and he is only responsible for whatever message being transmitted. In symmetric cryptography, one can deny the message sent by him/her, the digital signature remains as proof of one's activity in the digital world.

5 Cryptographic Techniques

Various cryptographic techniques are used for securing message which is being transmitted. These techniques are used to maintain the confidentiality, integrity of data that are transmitted.

Table 1 Classification of various types of attacks based on the security mechanism

Security mechanism [14]	Attacks	Solution
Authentication	Tunneling	RSU cooperation-based approach [16]
	Sybil attack	PKC, PVM, RTM [15]
	GPS spoofing	RSU cooperation-based approach [16]
	Node impersonation	Kalman filter and watermarking [17]
	Replay attack	Pseudo identity-joining process with the (RSU) [18]
	Message tampering	The watchdog-based scheme, AWF [19]
Availability	Denial of service	MMPDA [20]
	Malware attack	Updating software regularly
	Jamming	The measure of correlation among the error and the correct reception times [21]
	Broadcast tampering	Threshold-based trust [22]
	Black, gray hole attack	AODV [23], dynamic trust-based method [24]
	Spamming	Updating of software
Integrity	Greedy behavior attack	GDVAN [25]
	Masquerading	Trust models are necessary
Confidentiality	Illusion attack	Plausibility validation network (PVN) [26]
	Traffic analysis attack	Homomorphic encryption operation on Global Encoding Vectors (GEVs) [12]
	Eavesdropping	Strong firewalls
Non-repudiation	Man in the middle attack	
	Repudiation attack	Digital signature

Public-key Infrastructure (PKI): This PKI is used for authenticating the communication between people. It is the framework for encryption that protects the communication between source and destination node. The major components of PKI are digital signatures, certificate authority and registration authority. This public key Infrastructure plays a major role in being trusted third parties for users.

Encryption/Decryption: Encryption and decryption are prominently used authentication technique which is used in the wireless mode of communication. The messages which are being transmitted are encoded into a suitable format such that malicious user cannot read it. So, the messages which are being transmitted will be an encrypted message. Decryption is the process of extracting the original information from an encrypted message by using a key. Therefore, in this phase, the encrypted messages are decrypted by using a secret key and password.

Digital certificates: Digital certificates are used for sharing the public key which is used for encryption in the authentication process. These certificates are legally signed and kept up to date by the certification authority [5, 6].

Digital signatures: Digital signatures are used in the authentication process to verify that only the trusted sender has sent the messages and he is only responsible for whatever message being transmitted. In symmetric cryptography, one can deny the message sent by him/her, and the digital signature remains as proof of one's activity in the digital world.

6 Trust Models in VANET

The trust-based model has been categorized into three categories as Entity based trust models, data-oriented trust models and hybrid trust-based models [13, 20, 21]. These trust models are infrastructure-less and can be easily deployed because of the movement of nodes. Figure 1 describes the classification of trust target.

Entity-based trust model. The entity-based trust model is based on the trust which is provided by the reputation of vehicles in VANET [24]. Hu H et al. [27] have proposed a paper on entity-based trust model “REPLACE” A Reliable Trust-based Platoon Service Recommendation Scheme in VANET to neglect or reject the platoon vehicle which behaves as a malicious hacker [8, 27]. This trust-based platoon service head provides a rank for head vehicles by establishing a trust and reputation system. Chuang and Lee have proposed a paper on “TEAM” Trust Extended Authentication Mechanism entity-based trust-based authentication to reduce several attacks in vehicle-to-vehicle communication but this technique uses a shared key leading to vulnerabilities for attackers. Marmol et al. proposed a model on Trust and Reputation Infrastructure-based [1, 9] called TRIP to analyze egotistical and greedy nodes, which tries to send bogus information to increase their network utilization.

Data-Oriented trust model. The data-oriented trust model [1] is based on the trustworthiness of data that are transferred during communication between vehicles. Rawat et al. proposed a paper on Trust on the Security of Wireless Vehicular Ad-hoc Networking which is totally dependent on the data-oriented trust approach. In this approach, the trust level of the received message is calculated using RSS and the vehicle's location [3, 28].

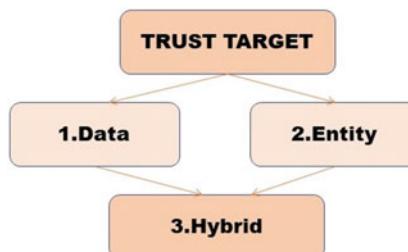


Fig. 1 Trust model in VANET

Hybrid model. In this, both vehicle and data trust are taken into considerations. Li and Song introduced ART Attack Resistant Trust management [29] which concentrated on both vehicle nodes and data trust for designing a trust model framework. Sugumar et al. proposed a method called TBAT [30] Trust based authentication technique which clubs both direct and indirect trust degree of nodes and it also verifies the message at the end using verifiers. Hasrouny et al. proposed a paper on Trust Model for a secure communication Group Leader (GL) [19] approach for communication in VANET. This model is used to classify vehicles based on their trustworthiness and elect potential GLs. The author proposes a risk evaluation methodology for the trust calculation of VANET. This methodology is used for identifying a malicious attack, assessing the imminence involved, and approaches are defined to alleviate them [3, 20, 31].

7 Trust Based Authentication Scheme in VANET

Authentication of messages is essential to ensure secure communication between nodes. It prevents unauthorized access of messages by hackers. Lotfi. A. Zadeh professor of UC Berkeley introduced fuzzy logic concepts. These fuzzy logic concepts are used in parametrization or it seems to be human decision-making methodology to deal with imprecise information. If the trust factor is 1 the node is considered to be heavily trusted if the trust factor is 0 the node is said to be the malicious node. This trust factor is either provided by the third-party trusted authority, direct node users or indirect node users. Shaikh et al. proposed a paper on, Fuzzy Risk-based Decision Method for Vehicular Ad Hoc Networks [11] to evaluate the risk before accepting unwanted or fake messages. Risk is evaluated by considering three elements application type and sensitivity level, vehicle context and driver's attitude [11, 15]. Figure 2 shows the risk elements of VANET. Application type messages are classified into three types as Infotainment application, traffic efficiency



Fig. 2 Risk elements

and safety application. Vehicle context is classified into its operation and characteristics of road and climate conditions. Driver's attitude is depended on experience and age. Messages which are received from safety application are given higher priority than other messages. In this paper, the priority of messages is given by characterizing application into two cases as when the sensitivity level of application is known and when the sensitivity of application is unknown [11, 32]. This work was again, extended by introducing a greater number of vehicles in surroundings, apart from accelerating and deaccelerating feature constant speed was introduced in the article Risk-Based Decision Methods for Vehicular Networks. Risk is evaluated by incorporating the decision-making process in fuzzy logic which improves the reliability of the system [33]. Soleymani et al. proposed a paper on a secure trust model based on fuzzy logic in VANET with Fog computing which not only detects malicious nodes but also overcomes uncertainty and imprecision data [3] in VANET involving both line of sight and non-line of sight communication [2, 34]. In this method, trust factor is calculated by calculating experience and plausibility which verifies information whether it is believable and reasonable. In direct interaction of vehicles, trust increases value 1 is given for absolute trust and 0 for untrusted nodes.

The plausibility model is evaluated based on line-of-sight communication which involves location based on distance GPS, RSS (Received signal strength), and using time calculation of occurring events like delivery of packet time, propagation speed is taken into considerations. In this proposed method, 55% of a malicious node was found when nodes are connected with fog node and 52% was found when connected with fog node. This fog node provides high accuracy of occurrence of events by calculating the node's location. The major drawback of this method was that it was not able to identify malicious attackers but tries to take action on the imprecision of data in VANETS Line of sight and Non line of sight communication. Saleem et al. proposed a paper on Cluster Head Expansion Stability technique using Fuzzy in Cognitive Radio (CR)-VANET to select a more trusted node for cluster head and to provide stability in the selection of cluster head nodes in CR VANET. For selecting a cluster head, fuzzy logic is used. A couple of fuzzy logic systems are utilized in this approach, one is used for enhancement of the probability of true detection of processing unit in cognitive radio with signal energy, signal to noise ratio is used as fuzzy input and the other for cluster head selection. Speed, probability of true detection, distance and lane weight are considered as inputs [10, 35].

Xia et al. discussed a paper on a Novel Trust-Based Multicast Routing for VANETs in which a novel trust-based multicast routing protocol is proposed to reduce malicious hacks and enhance routing techniques to deliver a packet at high accuracy. Fuzzy logic techniques are used to calculate the trust values of vehicle nodes. Trust has been divided into direct trust and indirect trust value of nodes. Direct trust is analyzed by using Bayesian theory and indirect trust value is obtained by evaluation and activity of nodes [1]. The total value of trust is obtained by the defuzzification process. The trust value which is obtained by total computation ensures a highly secured transmission of packets. The trust of a node is synthesized by using a fuzzy system. In the initial stage transformation of true value to a fuzzy set by a fuzzifier then fuzzy if-then rules are designed to calculate trust of a node, then this system is

followed by the fuzzy inference engine to find the degree of trust in a node. Defuzzifier is used to convert the output to real trust value. After the calculation of trust of a node, a secured route is discovered to deliver the packet. A trust-based protocol known as multi-cast trust-based ad hoc on-demand distance routing protocol was proposed to calculate neighbor's trust value and select a secured path for data communication [27, 36] Souissi et al. studied on a multi-level trust management model. The layers of trust are divided into three layers, to a sense a high quality of raw input of messages being transmitted. This data perception trust is done in the physical layer.

The physical layer consists of a dynamic vehicle node, Road Side Unit and Trusted authority. The vehicle nodes are equipped with the onboard unit, sensors and event data recorder to gather information from all surroundings; hence, data perception trust is done in the physical layer to ensure the quality of high standard of data is transmitted to nodes. The second layer is the network layer which is again divided into the transport and processing layer. The transport layer focuses on communication trust and the processing layer focuses on data fusion trust. In the transport layer, messages are exchanged between vehicles, roadside units, vehicles to a pedestrian. Each node should make a decision to trust the sender node or not based on the degree of trustworthiness. In this trust relationship, acquisition entity trust and message trust are evaluated then communication trust is established in the transport layer. Data fusion trust is established in the processing layer. The data received from the transport layer are analyzed then processed, and data fusion trust is established in the processing layer of the network. The application layer of VANET focuses on safety applications, traffic management applications and comfort applications vehicular node. New trust management may be defined in future which connects all levels of trust to provide a highly accurate message transmission. Figure 3 shows all levels of trust mentioned in this paper [37]. Pu, Cong presented a paper on Blockchain-based trust Management using multi-criteria decision-making model for VANETs [16]. In this, Blockchain technology is used to provide high-rate security to the messages being transmitted. Blockchain technology act as a digital ledger of transactions [17, 18]. It is a digital system that is difficult to hack by malicious hackers. In this paper, a Blockchain trust management is proposed by using a multi-criteria decision model to face the challenges which are related to the hacking of safety application messages [16, 38]. Kamran Ahmad Awan proposed a paper on stab trust; A stable and centralized trust-based clustering mechanism [21, 22] for IoT-enabled vehicles. Each node communicates with other nodes either by direct mode or by multi-hop mode of communication. The direct mode of communication causes high bandwidth problems [9, 16], high power consumption and utility of resources. To overcome these issues, clustering techniques are introduced. Clustering means the grouping of vehicle nodes under certain criteria to communicate with vehicle nodes. Each group of vehicles consists of a cluster head to direct all nodes for communication. To have this communication more stable without any malicious attacks, the author has proposed a stable and trust-based clustering technique. In stab trust, roadside unit plays a major role in recognizing trusted vehicles and selection of a trusted cluster head. All this work is carried out within the range of the roadside unit. The major trust component used by RSU is Knowledge, reputation and experience.

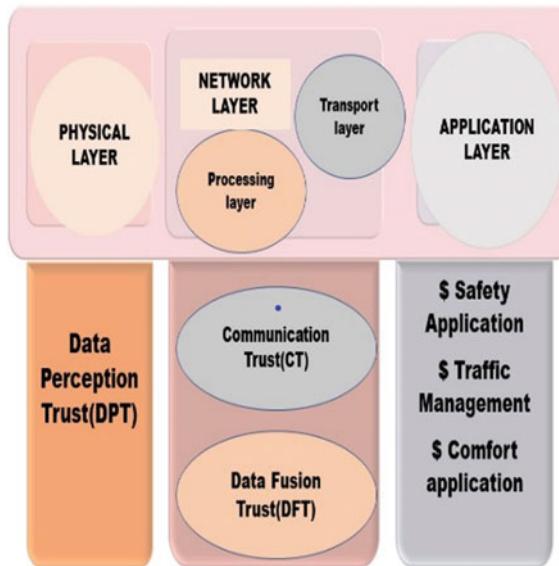


Fig. 3 Layers of trust in VANET

Knowledge trust has its subparameter as integrity and cooperativeness of observations, reputation has subparameter as honesty and behavior of nodes, experience trust has subparameter as competence and end-to-end packet delivery [39]. Tripathi et al. proposed a paper on Entity-Centric Combined Trust (ECT) Algorithm to detect a packet attack that harms a packet delivery ratio in Vehicular Ad Hoc Networks [28]. In this attack, the hacker can delay or change the information in the communication medium. This paper proposes a trust-based security algorithm based on the cooperation of vehicles. In this approach, both direct and indirect trust values of nodes are calculated. Direct trust [5] is calculated based on two parameters as satisfaction value of the node and weight factor of the vehicle node. Satisfaction value has its subparameter as reliability, position closeness and frequency. After calculation of direct value, indirect trust value is calculated based on recommendation trust value.

After computation of direct and indirect trust factor, both values are combined to get the Entity centric combined trust model to tackle the issues which are related to packet dropping in broadcast communication and to detect infected node and remove it from the communication platform [40]. Nandy et al. proposed a paper on T-BCIDS: Trust-Based Collaborative Intrusion Detection System [14, 15] for VANET in which every node is monitored in real-time to analyze and track the behavior of node. This real-time tracking is based on local identities agent which is encapsulated with k-nearest neighbor KNN nonlinear classifier [26, 41].

8 Conclusion

In the growing generation, the whole world is moving forward to see a digital world. All the communication which are done among users is the wireless mode of communication. Vehicular Ad Hoc network is one of the major applications in the wireless mode of communication which pays a way to give an intelligent and smart transportation system. When there is a growth in the advantageous effect of wireless communication there is also a disadvantageous effect of information stealing from malicious attackers or hackers to divert users. Hence, trust plays a significant and prominent role in the wireless mode of communication. The trust-based authentication technique drastically reduces and controls the information being hacked. In this paper, different types of VANET Architecture, cryptographic techniques and various trust models are discussed. A brief study on different types of trust-based authentication schemes is learned. A trust authentication can be implemented by using Fuzzy logic theory, Blockchain technology, and so on. A comparison has been made to determine the different types of trust authentication techniques used, parameters used to calculate the trust of a vehicle and tools used for this study. This study provides a way to work in trust factors in VANET to improve security mechanism in vehicular communication leading to a smart transportation system which will be useful in emergency issues like a precrash warning, traffic clearance, and safety application.

References

1. Ghori MR, Zamlı KZ, Quosthoni N, Hisyam M, Montaser M (2018) Vehicular ad-hoc network (VANET): review. In: 2018 IEEE international conference on innovative research and development (ICIRD), Bangkok, Thailand, 2018, pp 1–6. <https://doi.org/10.1109/ICIRD.2018.8376311>
2. Sheikh MS, Liang J (2019) A comprehensive survey on VANET security services in traffic management system. ID 2423915, 23 pages. <https://doi.org/10.1155/2019/2423915>
3. Devangavi AD, Gupta R (2017) Routing protocols in VANET—a survey. In: 2017 International conference on smart technologies for smart nation (SmartTechCon), Bengaluru, India, pp 163–167. <https://doi.org/10.1109/SmartTechCon.2017.8358362>
4. Gayathri M, Gomathy C (2021) A deep survey on types of cyber attacks in VANET. JCR 8(1):1029–1039. <https://doi.org/10.31838/jcr.08.01.114>
5. Mirsadeghi F, Kuchaki M, Gupta RB (2020) A trust infrastructure-based authentication method for clustered vehicular ad hoc networks. Springer Science+Business Media, LLC, part of Springer Nature 2020
6. Rasheed A, Gillani S, Ajmal S, Qayyum A (2017) Vehicular Ad Hoc network (VANET): a survey, challenges, and applications. In: Laouiti A, Qayyum A, Mohamad Saad M (eds) Vehicular Ad-Hoc networks for smart cities. Advances in intelligent systems and computing, vol 548. Springer, Singapore. https://doi.org/10.1007/978-981-10-3503-6_4
7. Wang Y, Ding Z, Li F, Xia X, Li Z (2017) Design and implementation of a VANET application complying with WAVE protocol. In: 2017 international conference on wireless communications, signal processing and networking (WiSPNET), Chennai, India, pp 2333–2338. <https://doi.org/10.1109/WiSPNET.2017.83001>

8. Chuang MC, Lee JF (2014) TEAM: trust-extended authentication mechanism for vehicular ad hoc networks. *IEEE Syst J* 8(3):749–758
9. Márml FG, Pérez GM (2012) TRIP, a trust and reputation infrastructure-based proposal for vehicular ad hoc networks. *J Netw Comput Appl* 35(3):934–941
10. IEEE guide for wireless access in vehicular environments (WAVE)—architecture. IEEE Std 1609.0–2013. IEEE, New York, NY, USA, 5 Mar 2014, pp 1–78
11. <https://www.arib.or.jp/english/arib/overview.html>
12. Fan Y, Jiang Y, Zhu H, Shen X (2009) An efficient privacy-preserving scheme against traffic analysis attacks in network coding. *IEEE INFOCOM 2009*:2213–2221. <https://doi.org/10.1109/INFCOM.2009.5062146>
13. Yao X, Zhang X, Ning H, Li P (2017) Using trust model to ensure reliable data acquisition in VANETs. *Ad Hoc Netw* 55. ISSN 1570-8705
14. Sheikh MS, Liang J (2019) A comprehensive survey on VANET security services in traffic management system. *Wireless Commun Mobile Comput* 2019, 23 pages. Article ID 2423915. <https://doi.org/10.1155/2019/2423915>
15. Pattanayak BK, Pattnaik O, Pani S (2021) Dealing with Sybil attack in VANET. In: Mishra D, Buyya R, Mohapatra P, Patnaik S (eds) Intelligent and cloud computing. Smart innovation, systems and technologies, vol 194. Springer, Singapore. https://doi.org/10.1007/978-981-15-5971-6_51
16. Kabbur M, Kumar A (2021) MAR_Spoof: securing VANET against spoofing and tunneling attack with cooperative assistance from RSU. In: ICASISSET 2020, 16–17 May 2020, Chennai, India
17. Jagadeesan DK (2019) Impersonation attack detection in VANET using Kalman filter and watermarking. Masters thesis, Dublin, National College of Ireland.
18. Alazzawi MA, Lu H, Yassin AA, Chen K (2019) Efficient conditional anonymity with message integrity and authentication in a vehicular Ad-Hoc network. *IEEE Access* 7:71424–71435. <https://doi.org/10.1109/ACCESS.2019.2919973>
19. Li Z, Chigan C, Wong D (2008) AWF-NA: a complete solution for tampered packet detection in VANETs. In: IEEE GLOBECOM 2008—2008 IEEE global telecommunications conference, New Orleans, LA, USA, 2008, pp 1–6. <https://doi.org/10.1109/GLOCOM.2008.ECP.377>
20. Kumar S, Mann KS (2018) Detection of multiple malicious nodes using entropy for mitigating the effect of denial-of-service attack in VANETs. In: 2018 4th international conference on computing sciences (ICCS)
21. Hamieh A, Ben-Othman J, Mokdad L (2009) Detection of radio interference attacks in VANET. In: GLOBECOM 2009–2009 IEEE global telecommunications conference. <https://doi.org/10.1109/glocom.2009.54253>
22. Daza V, Domingo-Ferrer J, Sebé F, Viejo A (2009) Trustworthy privacy-preserving car-generated announcements in vehicular ad hoc networks. *Veh Technol IEEE Trans* 58(4):1876–1886
23. Lachdhaf S, Mazouzi M, Abid M (2017) Detection and prevention of black hole attack in VANET using secured AODV routing protocol. In: Meghanathan N et al (eds), NeTCoM, CSEIT, GRAPH-HOC, NCS, SIPR—2017, pp 25–36, © CS & IT-CSCP. <https://doi.org/10.5121/csit.2017.71503>
24. Bhalaji N, Shanmugam A (2012) Dynamic trust based method to mitigate Greyhole attack in mobile Adhoc networks. *Proc Eng* 30:881–888. ISSN 1877-7058. <https://doi.org/10.1016/j.proeng.2012.01.941>
25. Mejri MN, Ben-Othman J (2017) GDVAN: a new greedy behavior attack detection algorithm for VANETs. *IEEE Trans Mobile Comput* 16(3):759–771. <https://doi.org/10.1109/TMC.2016.2577035>
26. Lo N, Tsai H (2007) Illusion attack on VANET applications—a message plausibility problem. *IEEE Globecom Workshops 2007*:1–8. <https://doi.org/10.1109/GLOCOMW.2007.4437823>
27. Hu H, Lu R, Zhang Z, Shao J (2017) REPLACE: a reliable trust-based platoon service recommendation scheme in VANET. *IEEE Trans Veh Technol* 66(2):1786–1797

28. Rawat DB, Yan G, Bista BB, Weigle MC (2015) Trust on the security of wireless vehicular ad-hoc networking. *Ad Hoc Sens Wireless Netw* 24(3–4):283–305
29. Li W, Song H (2016) ART: an attack-resistant trust management scheme for securing vehicular ad hoc networks. *IEEE Trans Intell Transp Syst* 17(4):960–969
30. Sugumar R, Rengarajan A, Jayakumar C (2016) Trust based authentication technique for cluster based vehicular ad hoc networks (VANET). *Wirel Netw* 22(5):1–10
31. Hasrouny H, Bassil C, Samhat AE, Laouiti A (2017) Security risk analysis of a trust model for secure group leader-based communication in VANET. In: Laouiti A, Qayyum A, Mohamad Saad M (eds) Vehicular Ad-Hoc networks for smart cities. advances in intelligent systems and computing, vol 548. Springer, Singapore. https://doi.org/10.1007/978-981-10-3503-6_6
32. Shaikh RA (2016) Fuzzy risk-based decision method for vehicular Ad Hoc networks. *Int J Adv Comput Sci Appl (ijacsa)* 7(9). <https://doi.org/10.14569/IJACSA.2016.070908>
33. Shaikh RA, Thayananthan V (2019) Risk-based decision methods for vehicular networks. *Electronics* 8:627. <https://doi.org/10.3390/electronics8060627>
34. Soleymani SA et al (2017) A secure trust model based on fuzzy logic in vehicular Ad Hoc networks with fog computing. *IEEE Access* 5:15619–15629. <https://doi.org/10.1109/ACCESS.2017.2733225>
35. Saleem MA et al (2019) Expansion of cluster head stability using fuzzy in cognitive radio CR-VANET. *IEEE Access* 7:173185–173195. <https://doi.org/10.1109/ACCESS.2019.2956478>
36. Xia H, Zhang S, Li B, Li L, Cheng X (2018) Towards a novel trust-based multicast routing for VANETs. *Secur Commun Netw.* Article ID 7608198, 12 pages. <https://doi.org/10.1155/2018/7608198>
37. Souissi I, Azzouna NB, Said LB (2020) A multi-level study for trust management models assessment in VANETs. *Int J Comput Intell Stud* 9:1–2, 107–127
38. Pu C (2020) Blockchain-based trust management using multi-criteria decision-making model for VANETs. *TechRxiv.* Preprint. <https://doi.org/10.36227/techrxiv.13106876.v1>
39. Awan KA, Ud Din I, Almogren A, Guizani M, Khan S (2020) StabTrust—a stable and centralized trust-based clustering mechanism for IoT enabled vehicular Ad-Hoc networks. *IEEE Access* 8:21159–21177. <https://doi.org/10.1109/ACCESS.2020.2968948>
40. Tripathi KN, Jain G, Yadav AM, Sharma SC (2021) Entity-centric combined trust (ECT) algorithm to detect packet dropping attack in vehicular Ad Hoc networks (VANETs). In: Deshpande P, Abraham A, Iyer B, Ma K (eds) Next generation information processing system. Advances in intelligent systems and computing, vol 1162. Springer, Singapore. https://doi.org/10.1007/978-981-15-4851-2_3
41. Nandy T, Noor RM, Idris MYIB, Bhattacharyya S (2020) T-BCIDS: trust-based collaborative intrusion detection system for VANET In: 2020 national conference on emerging trends on sustainable technology and engineering applications (NCETSTEA), Durgapur, India, 2020, pp 1–5. <https://doi.org/10.1109/NCETSTEA48365.2020.9119934>

Design of High-Speed Latched Comparator Used in Analog to Digital Converters by Using 180 nm Technology



Krishna Mohan Pandey, Yogendra Narayan Prajapati, and Naresh Kumar

Abstract In this paper, high-speed latch comparator has been designed for the application of analog to digital converter (ADC). The circuit's speed has been improved by a proposed comparator. It is designed with a supply voltage of 3.3 V at 180 nm CMOS technology at Cadence Virtuoso. By using the differential amplifier and latch design, a complete design for comparator is obtained.

Keywords Latched comparator · Differential amplifier · Speed · CMOS technology

1 Introduction

In this scenario because of increased order for convenient battery-powered tools, the need arises for high-speed latched comparators. Comparators have a significant part with commonly deployed ADCs, e.g., flash and SAR ADCs. Comparator considers as a chief building block in the design of any analog to digital converter. The comparator affects total ADC efficiency in certain forms of ADCs like successive ADC approximation. Moreover, in several instances, the failure of the comparator may be adjusted for. Figure 1 displays the comparator's design sign. A comparator could be viewed as a judgment producing circuitry. Generally speaking, Comparator defines as a circuit which differentiates an analog signal to some other analog or reference signal which generates a binary signal depending on differentiation. Consider the figure shown; if positive signal (V_{in+}), the comparator's input is larger than the negative (V_{in-}) input, the comparator's output is logic 1, wherever if positive, input is smaller than the negative input potential, then the comparator's output is logic 0. In reality because of limited gain, a comparator outcomes in single situation; (1). $V_{in} > V_{ref} + V_{IH}$

K. M. Pandey · Y. N. Prajapati (✉)

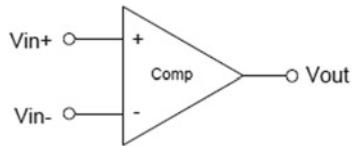
Department of Computer Science and Engineering, DVSIEET, Meerut, U.P., India

N. Kumar (✉)

Department of Computer Science and Engineering, Quantum University, Roorkee, U.K., India

e-mail: naresh.cse@quantuneducation.in

Fig. 1 Comparator's symbol
[1]



becomes zero (2). $V_{in} < V_{ref} + V_{IL}$ becomes one. Technically, a binary signal in a digital structure should only have single situation at a moment in period, and such a form of binary signal is good for actual-world conditions, anywhere here is constantly a transition area running among two binary conditions. Passing through this transition area very quickly is very necessary for the comparator. Under the method of varying analog signals into digital signals, comparator often utilized sampling the output signal is really the early stage in analog-to-digital transformation. The sampled signal will formerly be put to a mixture of comparators for measuring the digital equivalent of that analog signal. It can be described as a 1-bit ADC in simplest form.

2 Comparator's Principle

The basic principle for high-speed comparator is to have a preamplifier phase for generating a reasonably significant volume of output adjustment and then transfer it to trigger as an input. It combines the best elements of the circuit. The two results are mixed for generating single answer map, a negative exponential result (preamplifier) from circuit through a positive exponential result (latch).

3 Time Constant for Latch Mode

The time constant for latch circuit is in its positive feedback and when it can be located in the latch cycle by examining a simpler circuit composed of two consecutive inverters. When we presume as inverters in their true value and that output voltages of inverters are similar to each other in the start of latch cycle, formerly each of inverters could be demonstrated as a product of voltage-controlled current stirring a RC load. Low-frequency gain of each inverter is represented by A_v that has a transconductance represented as [2];

$$G_m = \frac{A_v}{R_L}$$

We have for this linear structure;

$$\frac{A_v}{R_L} V_y = -C_L \left(\frac{dV_x}{dt} \right) - \left(\frac{A_v}{R_L} \right) \text{ and}$$

$$\frac{A_v}{R_L} V_y = -C_L \left(\frac{dV_x}{dt} \right) - \left(\frac{A_v}{R_L} \right)$$

Consequence is that the above two formulas are combined by R_L and rearranged as follows:

$$\tau \left(\frac{dV_x}{dt} \right) + V_x = -A_v V_y \text{ and}$$

$$\tau \left(\frac{dV_y}{dt} \right) + V_y = -A_v V_x$$

$\tau = R_L C_L$ represents time constant at the output node of every inverter. By doing the subtraction of above two terms and then reordering them gives,

$$\Delta V = \left(\frac{\tau}{A_v - 1} \right) \left(\frac{d\Delta V}{dt} \right)$$

$\Delta V = V_x - V_y$ shows the voltage variation among the output voltages of inverters. The upper mentioned term is a first-order differential equation. The explanation for this is as follows;

$$\Delta V = \Delta V_0 e^{(A_v - 1) \frac{t}{\tau}}$$

Here, ΔV_0 is in starting of latch phase in voltage variation. So, this voltage variation is exponentially rises with time by a time constant which represents below;

$$\tau(\text{latch}) = \frac{\tau}{A_v - 1} = \frac{R_L C_L}{A_v} = \frac{C_L}{G_m}$$

And $G_m = \frac{A_v}{R_L}$ shows the transconductance of every inverter. $\tau(\text{latch})$ is equivalent to the opposite of unity-gain frequency of every inverter.

In MOS structures, the output load is generally proportional to a single-transistor's gate-source capacitance, or in general,

$$C_L = K_1 W L C_{\text{ox}}$$

Here, K_1 shows proportionality constant whose range lies in 1 and 2. The transconductance of the inverter is proportional to the transconductance of a single transistor, as follows;

$$G_m = K_2 g_m = K_2 u_n C_{\text{ox}} \frac{W}{L} V_{\text{eff}}$$

K_2 is a constant having range 0.5–1. By putting all the values, it gives;

$$\tau(\text{latch}) = \frac{K_1}{K_2} \frac{L^2}{u_n V_{\text{eff}}} = K_3 \frac{L^2}{u_n V_{\text{eff}}}$$

K_3 lies in between 2 and 4. Remember that term above, which primarily relies on the engineering. In order to safely identify the appropriate output value in the successive logic circuitry, it is necessary to obtain a voltage difference; then we note that the time needed to do this is given by;

$$\tau_{\text{latch}} = \frac{C_L}{G_m} \ln\left(\frac{\Delta V_{\text{logic}}}{\Delta V_0}\right) = K_3 \frac{L^2}{u_n V_{\text{eff}}} \ln\left(\frac{\Delta V_{\text{logic}}}{\Delta V_0}\right)$$

If ΔV_0 its tiny then this latch period may be huge, maybe it's bigger than the latch process time allowed. We can say that the latch's differential output voltage does not rise sufficient to be accepted by subsequent circuitry as the right logic quality because its early rate is low. Occasionally, circuit interference may trigger the initial difference in voltage to be low enough to induce metastability even when the initial difference in voltage is significant enough.

4 Practical Implementation

The phase has the task of saturating down output voltage to V_{SS} or V_{DD} point, depending upon its inputs. In particular, a phase has the comparator's main purpose, it can only react correctly if its differential input voltage is sufficiently high to confirm correct operational precision. To order to overcome this restriction, a preamplifier phase is typically used earlier the trigger, whose prime aim is of amplifying voltage of the input signal. It is important to use in our case LSB has a value of 2 milli Volt, and the appropriate preamplifier gain need to be minimum value of 7, that implies almost 17 dB. As the comparator's conceptual interpretation is stated above, the full implementation of the circuit is performed in cadence window together with the simulation outcome. The two major elements to be built as follows:

- I. Differential amplifier
- II. Latch circuit.

We know that we need to model our circuit on some requirements. So I'm naming the parameter as needed. I choose the template folder for the SCL 180 nm CMOS software. This directory includes vector information like u_n , u_p , V_{tn} , V_{tp} , t_{ox} etc. All these parameters are defined by the range of voltage we are working on. I have been worked on 3.3 V, and I pick all the parameters by keeping the value of voltage accordingly.

5 Design of Differential Amplifier

An electric implement or a digital amplifier which rises a signal's strength (a current or voltage that changes in time) is known as differential amplifier. An amplifier mainly use the electrical power from the power supply to rise the strength of the applied input signal. The quantity of strengthening given by an amplifier is generally calculated by its gain (known as the input-to-output ratio). An amplifier is a circuit that can typically produce more than one energy gain. An optimal amplifier is an electric two-port system that connects the output signal with an input signal copy as long as it rises in magnitude. Electronic amplifiers use either a current or a voltage as one parameter. We may assume that either voltage or current can be used as output.

Differential amplifier seems to be an essential circuit to be invented, which dates back to the era of the vacuum tube. Throughout today's times, differential process forms the prevailing option, providing several beneficial things. A single ended signal determines how a defined potential and floor are calculated. Likewise, a differential signal is the one which determined among the two nodes with identical and contradictory excursions nearby a defined potential. Center potential is referred as the Common Mode (CM) level in a differential signaling. We must be able to reject specific mode interference, which is why they have better resistance to environmental noise. In my model, I used the diode-connected load to create a differential amplifier. My initial criterion is to rises the comparator's speed. Therefore, I need to enlarge the differential amplifier's unification gain distance. If the gain has to be reduced, then a range of amplifier steps can be cascaded, and the gain is boosted to a greater extent as a result.

The UGB represents as;

$$\text{UGB} = A(0) * P_1$$

$$P_1 = \text{UGB}/A(0)$$

Given below shows time constant;

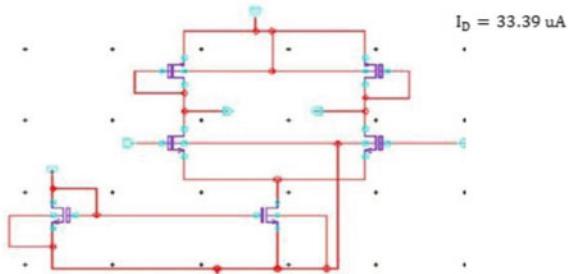
$$\tau = \frac{1}{2\pi f - 3\text{dB}}$$

6 Generation of Differential Amplifier's Schematic

Here, the generation of differential amplifier's schematic is done by cadence software tool (Fig. 2).

Using CMOS principle and model files, the values of different parameters are calculated. I built the above differential amplifier for 1 GHz bandwidth, V_{tp} , V_{tn} and

Fig. 2 Differential amplifier's schematic in Cadence



$u_n C_{ox}$, $u_p C_{ox}$ values were picked from the template folder. The overdrive voltage for the attached diode load is determined;

$$V_{0v3} = 0.9875 \text{ V}$$

As the gain is preferred to be 7, we can measure overdrive voltage for the transistors at the bottom of the pair by using the equation below;

$$\frac{V_{0v3}}{V_{0v1}} = A$$

$$\frac{0.9875}{V_{0v1}} = 7$$

Then,

$$V_{0v1} = 0.1412 \text{ V}$$

Use the relation of GB and g_m , gives;

$$\text{GB} = \frac{g_m}{2\pi C_L}$$

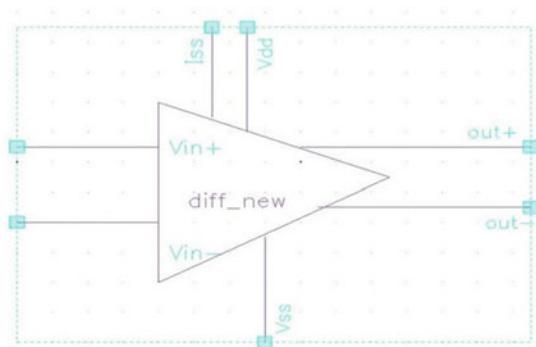
By substituting the value of load of next stage and GBW, the transconductance becomes;

$$g_m \geq 473.07 \text{ uA/V}$$

The correlation among overdrive voltage, drain current and transconductance represent as follows;

$$g_m = \frac{2I_D}{V_{0v1}}$$

Fig. 3 Differential amplifier's symbol



$$473.07 \mu = \frac{2I_D}{0.1412}$$

Drain current has a value;

Therefore, by the use of these parameters and using the saturation area definition, and using drain current formulation, W and L can be determined for each transistor by setting the value of the template collection's system parameters. After the schematic has been developed, the circuitry is checked by adding the similar output at both the input terminals and the response is correct to the predictable range, so that the symbol is produced and tested shown in Fig. 3.

Offset Calculation:

Upon making the differential amplifier phase sign, the next move is measuring offset voltage for this point & I used the calculator to test the offset in the spectrum simulator. I used the 'price' function to verify that the output is null at what stage when both outputs are zero. This point in the differential amplifier provides offset value. I used the study of monte-carlo, which gave the best answer.

7 Latch Design

The next move is to model the latch stage after positive testing of differential amplifier. Essentially, this stage is inserted to provide the performance with a specific level. The performance can be either logic 0 or logic 1 indicating V_{DD} . There is very little output for differential amplifier. As the scope of the full scale is given;

$$V_{FSR} = 0.65 \text{ V} - 2.65 \text{ V}$$

Therefore, the value of 1 LSB is observed, which is supplied by this value;

$$1 \text{ LSB} = \frac{V_{\text{REF}}}{2^N}$$

where $N = 10$ bit Hence,

$$V_{\text{LSB}} = \frac{2.65 - 0.65}{1024} = 1.9 \text{ mV}$$

It implies that a comparator will have to handle so many performance shifts. A preamplifier stage is required to strengthen this signal to a definite level. After enough amplification, the latch stage is cascaded to produce unique production resulting from applied signals (Fig. 4).

That transistor's W & L are selected to hold all the transistor in linear area. To offer the stability, second latch stage is inserted. The stage of the latch serves as the loop regenerating the output to a point. The schematic diagram of latch along with its symbol represents as (Fig. 5).

It takes the output differential and the formula given.

$$V_{\text{out}} = \Delta V_0 \left(1 - e^{-\frac{t}{\tau}} \right)$$

$$\tau = \frac{1}{2\pi \text{GBW}}$$

There are switches that initialize the circuitry for the first time and then execute the process. After successfully checking the latch circuit, the latch symbol is generated and supplied (Fig. 6).

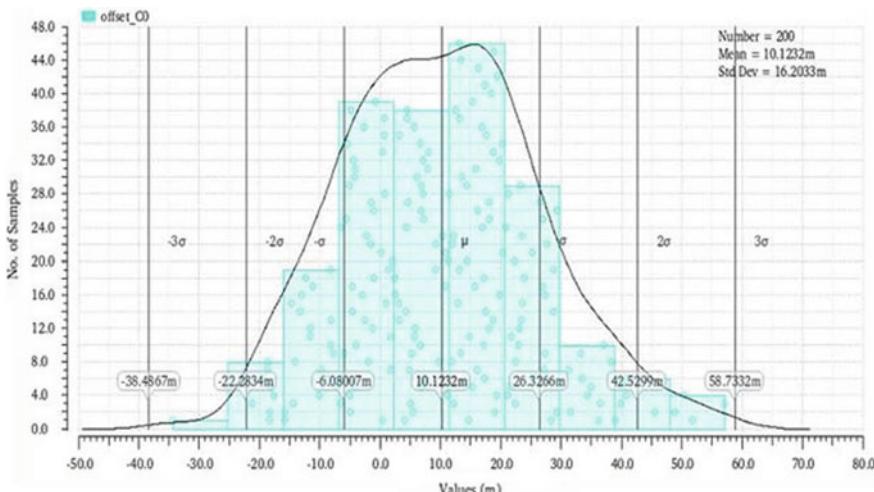


Fig. 4 Monte-Carlo simulation window

Fig. 5 Latch's schematic diagram

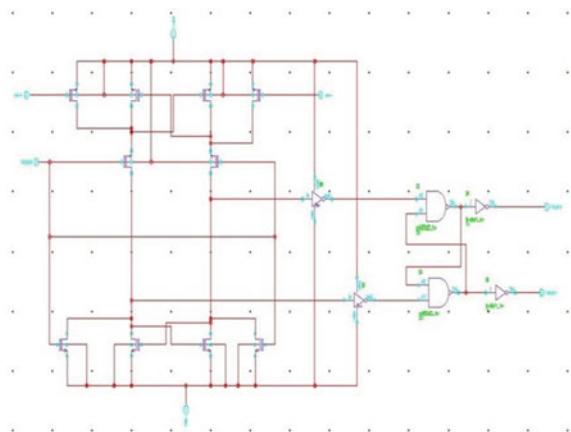
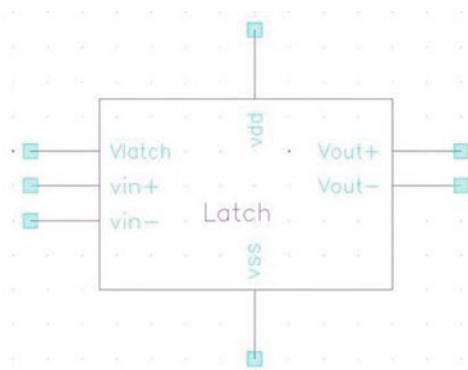


Fig. 6 Latch symbol

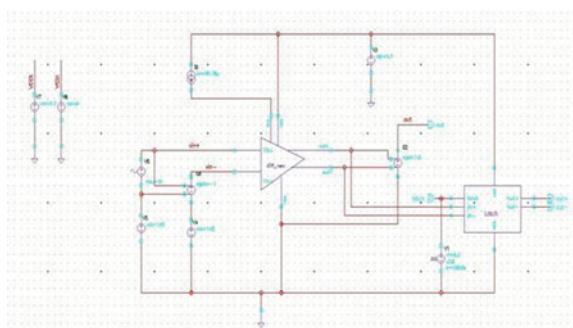


8 Comparator's Complete Schematic

After the formation of symbol, both stages are combined for testing comparator's usability. When the proper biasing is done, the collective circuit is represented in Fig. 7. The preamplifier and lock symbols are used in this setup, and the voltage of the standard mode DC is assumed to be 1.66 V.

The comparator describes the degree of voltage equivalent to 1 LSB. When voltage varies at the source, the output is created based on the higher voltage added to the terminal.

Fig. 7 Comparator's complete schematic



9 Simulation Results of Comparator Design

The ac and dc analysis are verified after the schematic is produced by applying the clock signal to latch. ADE L is completed for this schematic circuitry. It also indicates the outcome that represents phase and gain. The transient analysis can be found in Fig. 8.

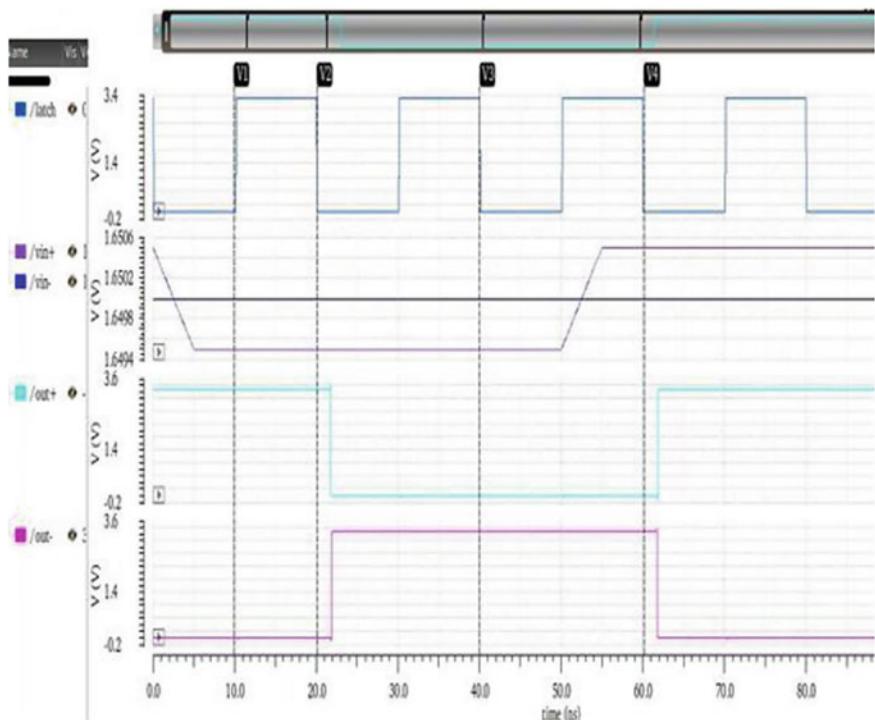


Fig. 8 Final output of the comparator

10 Comparator's Mismatch Analysis

When the complete circuit is applied, the succeeding move is to perform the mismatching test, which implies that in the comparator we must verify the offset. This offset would say nearly the output voltage variance. In the Hspice simulator, I did this research. To evaluate the offset cost, I used the ‘measure’ order. I measured the shift from $-1/4$ LSB to $+1/4$ LSB randomly. At the two interfaces, I have to add two feedback signals ranging from -50 to 50 mV.

The meaning of the offset voltage is at what stage we can display the difference in output. Also, I did a minor mathematical computation just to test the scheduling.

$$1 \text{ LSB} = 1.9 \text{ mV}$$

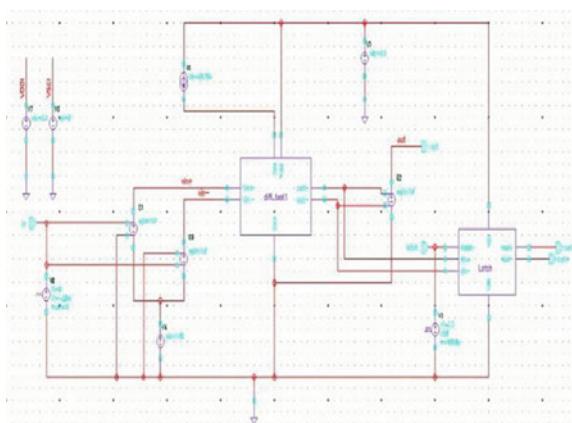
$$\frac{1}{4} \text{ LSB} = \frac{1.9}{4} \text{ mV} = 0.447 \text{ mV}$$

Because my circuitry is built for 1 MSPS, I select frequency of 12 MHz upon which circuit runs. As per this analysis, the timing necessary to operate the waveform is provided as the period must be 4.384 u-sec for the 100 mV signal. The outcome of the histogram plot together with the schematic shown in the diagrams below.

Figure 9 represents the schematic diagram in which I implemented the equivalent and opposite magnitude of the input voltage. Thus, both the signals differ in a variety and voltage range at which the improvements in output are regarded to be voltage offset.

There are two methods to determine offset using measure control, either by generating net list and applying command to net list, or by saving a new file in which command has been already written and by manually inserting it. Therefore, below the histogram plot is represented which provides the consequence of the calculation of offset (Fig. 10).

Fig. 9 Schematic window for offset calculations



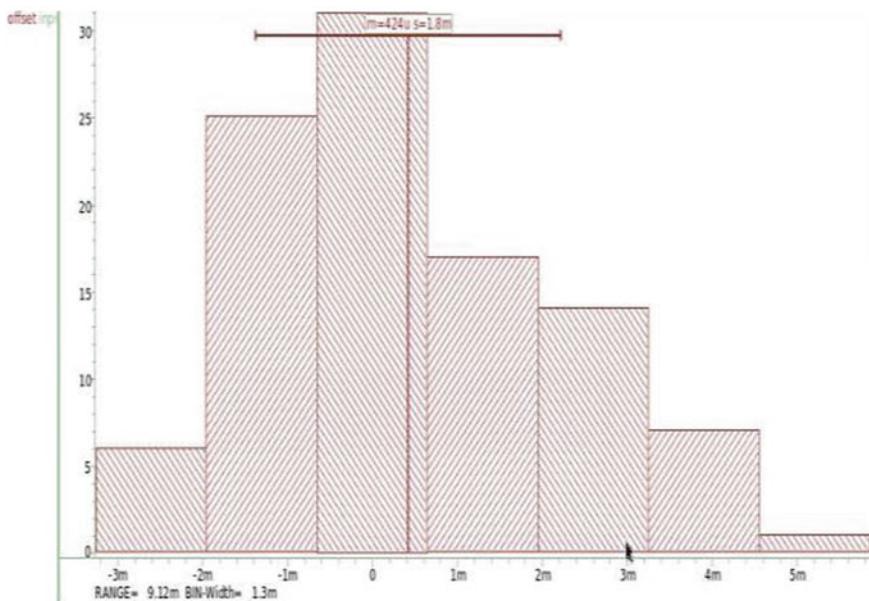


Fig. 10 Offset calculation in the form of histogram

The value of offset voltage and standard variation is represented under (Table 1).

$$V_{\text{offset}} = 0.424 \text{ mV}$$

$$\text{Std_Var} = 1.8 \text{ mV}$$

And, by using different ways, one can strengthen certain principles and this is the activity one can do in future to maximize the comparator's output.

Table 1 Comparator's performance

S. No	Parameters	Values
1	Gain	17.69 dB
2	Supply voltage	3.3 V
3	Standard deviation	424 uV
4	UGB	1.13 GHz
5	Standard deviation	1.8 m V

11 Conclusion

A high-speed latched comparator is designed. The speed of the comparator is being improved and various parameters for comparator is calculated in table. The calculation and performance of comparator are designed so that it proves beneficial to be used in various applications of ADC.

References

1. Behzad R, Wooley BA (1992) Design techniques for high-speed, high-resolution comparators. Solid-State Circ IEEE J 27(12):1916–1926p
2. Cho TB, Gray PR (1995) A10 b, 20 Msample/s, 35 mW pipeline A/D converter. Solid-State Circ IEEE J 30(3):166–172

Healthcare Assistant—A Tool to Predict Disease Using Machine Learning



Sujata Joshi, Harish Kumar, Jagadish Babu, Akhil Raju, and Mohammad Nihaz

Abstract The rapid spread of Internet technologies and mobile devices has created new opportunities for online health care. There are times when consumers can absorb Internet medical help or healthcare advice more easily than in-person assistance. People prefer to look for a solution or therapy online for mild ailments rather than going to the hospital or seeing a doctor. And people in rural areas sometimes ignore or try to ignore mild symptoms until the sickness has progressed to the point where it is no longer treatable. In many cases, however, these modest symptoms might lead to serious health problems. People ask health-related questions on a variety of healthcare forums (e.g., what kind of ailment they might be suffering from). Another group of people responds to those messages by predicting diseases that may or may not occur. These forecasts, however, may not always be accurate, and there is no guarantee that users will always receive a response to their posts. Furthermore, some posts are manufactured or made up, which can lead the sufferer astray. According to a CNN poll, 25% of users on social networking sites lie. As a result, trustworthiness is a major concern. In recent years, there has been a lot of study toward automated disease prediction, but the accuracy and capacity to process user input have been a key worry. Our proposed application will probably be able to predict the diseases more accurately based on the query or symptoms provided by the patient and also be able to predict the future threats to the health based on the patient's history. For the newly graduated doctors, this application will initially be able to provide medical prescription as an assistance.

Keywords Prediction · Machine learning · Cosine similarity · Query vectorization

S. Joshi (✉) · H. Kumar · J. Babu · A. Raju · M. Nihaz
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: sujata.joshi@nmit.ac.in

1 Introduction

The problem of providing high-quality, affordable health care is becoming increasingly challenging. Investigating and interpreting the use, costs, quality, accessibility, and delivery of healthcare services due to the complexities of healthcare services and systems. In India, the doctor-to-population ratio is 1:1456, compared to the WHO's recommended ratio of 1:1000. According to a national poll performed by the Pew Internet Project, 72% of Internet users in the USA have sought health information online. On several healthcare forums, people publish their health-related questions. Consumers are increasingly using handheld devices to access the internet, which leads them to search for an online solution or remedy (typically on Google or online forums). There is no guarantee that the anticipated data will be accurate, and people may not receive a response to their messages.

Furthermore, some posts are manufactured or made up, which can lead the sufferer astray. People living in poverty or in rural areas with a lack of excellent healthcare facilities may live longer if they have access to a proper hospital or doctor. People frequently overlook little symptoms until they become something that cannot be ignored, but it is sometimes too late to cure due to a lack of knowledge about the disease. Once the condition has been identified, appropriate lifestyle adjustments and treatment options can be recommended for improved health and longevity.

The major goal is to create a healthcare assistant that will be able to forecast a patient's ailment using the patient's symptoms as input, as well as make lifestyle recommendations for better health. To forecast major health risks based on a person's medical history [1]. To aid newly graduated doctors in recommending medical prescriptions.

2 Literature Survey

The value of ancestor medical history has been described in [1] with the statement that ancestor medical history has been an important tool in developing risk profiles for offspring. These medical histories can be utilized to develop a risk profile for the family's current and future generations. The majority of doctors rely on the patients' family medical history. Most patients, on the other hand, are unaware of their parents' and grandparents' comprehensive medical histories. Predictions based on family medical history may aid in predicting an individual's distinctive characteristics. Based on the symptoms, the author proposes a revolutionary method for recognizing diseases and predicting their cure time [2]. This is accomplished by assigning distinct coefficients to each disease symptom and filtering the dataset using the severity value supplied by the user to each symptom. The diseases are detected using a numerical value calculated in a particular way [3]. People nowadays suffer from a range of diseases as a result of the environment and their lifestyle choices. As a result, predicting sickness at an early stage becomes a critical task. Doctors,

on the other hand, find it difficult to make precise predictions based on symptoms. Over the last few years, there has been a massive growth in medical data. As a result, reliable predictions may be made, and patients can be treated sooner [4]. This research study looks into a variety of data mining methods that have been utilized to forecast diabetes disease. As a result, for each subsequent primary component in the system, the variance decreases [5]. The author uses three different disease datasets to apply multiple categorization methods, each with its own set of benefits (heart, breast cancer, diabetes). The most prevalent type of BN is the Naive Bayesian network, which has the highest accuracy values for diabetes and CVD classification, respectively, 99.51 and 97.92%. However, when mean accuracy was calculated using ANN, the findings were better, indicating that when ANN is employed, there is a greater chance of getting more accurate findings in diabetes and/or CVD categorization [6, 7]. Most medications have the same impact on the majority of patients; however, a few medications may have a minor or severe adverse effect on a small number of people. With PH, a patient can get a medicine that suits him, lowering costs and improving care quality. For instance, not all diabetic patients react to food and insulin intake in the same way. As a result, a Bluetooth-enabled blood sugar monitor can assist the patient in determining when and how much food to consume. On a mobile device, same advice can be given. Qualified doctors are few in distant and semi-remote places. Patients who are unable to visit the hospital on a regular basis may benefit from remote monitoring. Doctors can keep track of their patients' health utilizing cloud-based technologies that benefit both the doctor and the patient [8, 9]. Malaria outbreaks in Maharashtra were forecasted using a neural network using SVM. It predicted this based on rainfall, temperature, previous recorded cases, and other medical information. Here, both the needs of the consumer and the privacy of the patient are respected. If the consumer wants to see the patient's details, he will have to pay a price. This data also includes the patient's meeting with all of the doctors. The Counterfeit Medicines Project was created with the goal of preventing drug counterfeiting. To prevent counterfeiting, they utilized blockchain since data cannot be changed and no one can change the date when the medicine was inserted [10].

3 Methodology

In this section, the proposed methodology is discussed. The system architecture can be divided into three important components, namely feature extraction, disease prediction, and risk prediction. The system architecture is shown in Fig. 1.

The dataset record has been created in Excel sheet and later converted into the CSV format.

The data used for predicting disease using symptoms is summarized in the above table. And, the dataset used for risk prediction consists of medical history of the patient. Medical History consists of PatientId, Name, Age, Gender, Location, and Diseases, and date of diagnosis of disease. The CSV format consisting of the above data is given as input to cosine nearest neighbor risk prediction model. The ML model is interested in the PatientId, disease, location and date of diagnosis. Then the dataset is processed, and the method for classification considers the main part of the dataset by using classification algorithm and cosine the nearest neighbor algorithm.

3.2 Feature Extraction and Vectorization

The aim of this process if to extract all the features from the human-readable language and then convert into a machine-readable form. We want the output of the current process to be in a format which is desirable to the next component of the process, i.e., after the conversion from human-readable form to machine-readable form which is format that can be directly fed into the next process as an input without any additional data processing to increase the performance by eliminating any other intermediate task required.

Here in this process when a person enters a query into the application, this query needs to be processed into binary data as in the upcoming process is an expecting a data in binary format. The data is converted into binary format so that the processing time and the data storage are significantly reduced. To convert human readable query to machine understandable code, we have used a bag of words method to extract all the features from the query and then produce an array containing data about the symptoms in the query. The features that are extracted in query are the symptoms of the patient. Suppose in the query, if the user has entered all the information about the symptoms he is suffering, then using bag of words we extract all the necessary data and ignore all other information. Once the features are identified, then a binary array containing symptoms data (each position in an array represents a symptom in the database) will be populated with 1 in respective position of symptom in the array if a user enters a symptom keyword representing the presence of the symptom in the query else the default value (i.e., 0 will be entered, for all other missing symptom in the query) will be kept as it existed. Then the array containing all the necessary information from the user query will be fed into the trained classifier. The process is depicted in Fig. 2.

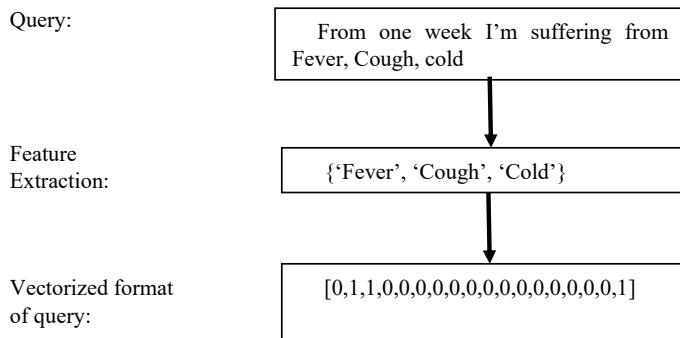


Fig. 2 Feature extraction

3.3 Disease Prediction

In this process of the application, the output from the previous stage is fed into the trained classifier. The vectorized query which contains all the necessary data from the query is provided to the machine learning model. Here decision tree classifier is used to classify/predict the user query and provide the possible disease behind all the symptoms. The classifier is trained using the binary tabulated table data containing all the information of the symptoms and respective disease data. This data is periodically updated, and relatively the model is also trained for necessary changes to be updated. Here according to the test conducted on database using various ML technique, it was noticed that decision trees has been consistent and provides results with high accuracy compared to other ML models that we tested.

After training and developing the model using the database (here in our case tabulated binary database), the vectorized query is feed into the model, then the classifier provides a value (i.e., respective disease). Suppose if the model provides a less confident value, then the user is again asked to provide more information, so that a better and accurate information can be given out. These query results are stored for future usage and risk prediction.

3.4 Risk Prediction

In this process, patients or user's risk is predicted using the health records/history of the patient based on the geographical location. As we know, immunity of a person varies from region to region due to climate, food, lifestyle, and various other factors. So, it is ideal to generate an individual's risk based on the geographical location rather than other factors. Firstly, we collect all the patient history of a particular region, then we find the similarity of the patient living with similar kind of disease. Then a pattern of disease occurrence of the individual is analyzed. Using this disease

pattern, we match with various another patient with similar disease pattern. Using this data, our model predicts the future risk a person can face using the collaborative patient's database.

$$\text{Cosinesimilarity}(|x.y|) = \frac{x.y}{||x||||y||} \quad (1)$$

where $x \rightarrow$ vector of target patient., $y \rightarrow$ vector of all other patients in same region.

To analyze the patient similarity, cosine nearest neighbor is used to find out the measure of similarity of the patient with the similar condition living in the same or similar region. By using cosine nearest neighbor, it is known that the lesser the value of theta the closer the user is related to each other, and higher the value of theta then the users are dissimilar. Hence using the similarity of the user, a pattern can be recognized and the next risk/disease can be generated. This helps the patient or the user to well aware of the future threats and can take necessary precautions to prevent the diseases.

4 Results

In this part, we have used a variety of input parameters to evaluate the classification algorithm's performance for the dataset. Here we have used two testbed cross-validation processes: a percentage split, which utilizes 66% of the dataset for training and 34% for testing and outputs, and a tenfold cross-validation process, which uses nine folds for training and onefold for testing and outputs. The results are tabulated in Table 3.

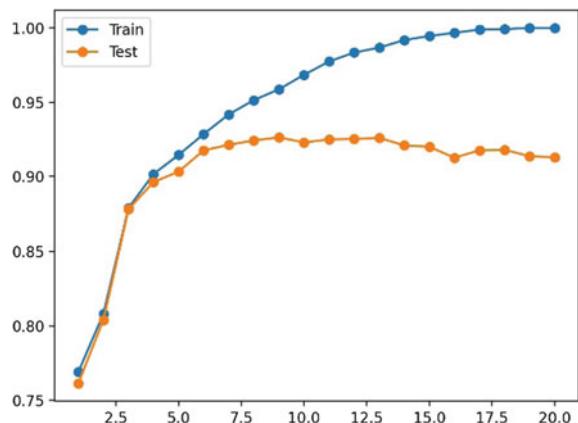
In this work, the proposed model uses decision tree model to classify the disease the symptom, because the classification accuracy for the given input was quite higher than all the other classifier we have tested. The output in our implementation matched the predicted results very well, and the accuracy we obtained with the dataset we utilized was rather good and as shown in Fig. 3.

This work cannot be an immediate replacement to the existing medical procedures, but after high training and constant observation, this application can be used to eliminate the problems faced by patients currently faced by patients in the real world.

Table 3 Accuracy comparison of different algorithm

Classifier	Accuracy (%)
Decision tree	100
Naïve-Bayes	86.41
KNN	84.09
SVM	79.67

Fig. 3 Accuracy of decision tree



5 Conclusion and Future Work

Based on current tools such as My Family Health Portrait and Microsoft Health Vault, where users must enter details, these technologies capture the family history. These particulars are shared with medical professionals for their review. If two diseases have the same symptoms, the tool will be unable to anticipate the exact condition that a person is experiencing. The system designed here mainly helps for people living in rural areas to predict their disease in early stage. Due to this, sudden death of people can be avoided. When a highly experienced doctor died due to age, his experience cannot be wasted because he can feed his experience periodically to the system or model so that newly graduated doctors can make use of this. Nowadays, Internet health advice is readily available so that this model can provide users with a significant advantage by integrating it with modern social media chat facilities. It can be further developed to include data from multiple locations and prediction results separately, exclusively for that particular region. The proposed system is capable of predicting disease by taking symptoms of a disease. Hence, the diseases prediction system can be optimized to predict and produce accurate results by taking symptoms of multiple diseases at once.

References

1. Agrawal R, Suleiman M, Seay C, Gloster C (2020) Dynamic disease forecast network using family medical history—IEEE conference publication. [Online] Ieeexplore.ieee.org.
2. Shankar M, Pahadia M, Srivastava D, Reddy G (2013) A novel method for disease recognition and cure time prediction based on symptoms—IEEE conference publication. [Online] Ieeexplore.ieee.org
3. Dahiwade D, Patle G, Meshram E (2019) Designing disease prediction model using machine learning approach—IEEE conference publication. [Online] Ieeexplore.ieee.org.

4. Kanchan B, Kishor M (2017) Study of machine learning algorithms for special disease prediction using principal of component analysis—IEEE conference publication. [Online] Ieeexplore.ieee.org.
5. Singh Kohli P, Arora S (2019) Application of machine learning in disease prediction—IEEE conference publication. [Online] Ieeexplore.ieee.org.
6. Alić B, Gurbeta L, Badnjević A (2017) Machine learning techniques for classification of diabetes and cardiovascular diseases—IEEE conference publication. [Online] Ieeexplore.ieee.org.
7. Ahamed F, Farid F (2019) Applying internet of things and machine-learning for personalized healthcare: issues and challenges—IEEE conference publication. [Online] Ieeexplore.ieee.org.
8. Vyas S, Gupta M, Yadav R (2019) Converging blockchain and machine learning for healthcare—IEEE conference publication. [Online] Ieeexplore.ieee.org.
9. Ahmed Neloy A, Alam S, Alif Bindu R, Jahan Moni N (2019) Machine learning based health prediction system using IBM cloud as Paas—IEEE conference Publication. [Online]. In: Ieee Lan Z, Zhou G, Duan Y, Yan W (2018) AI-assisted prediction on potential health risks with regular physical examination records—IEEE conference publication. [Online]
10. Hong W, Xiong Z, Zheng N, Weng Y (2019) A medical-history-based potential disease prediction algorithm. [online] Ieeexplore.ieee.org.

A Comprehensive Survey on Topic Modeling in Text Summarization



G. Bharathi Mohan and R. Prasanna Kumar

Abstract Topic modeling is the statistical model for discovering hidden topics or keywords in a collection of documents. Topic modeling is also considered a probabilistic model for learning, analyzing, and discovering topics from the document collection. The most popular techniques for topic modeling are latent semantic analysis (LSA), probabilistic latent semantic analysis (pLSA), latent Dirichlet allocation (LDA), and the recent deep learning-based lda2vec. LDA is most commonly used in extractive multi-document summarization to determine whether the extracted sentence reflects the concept of the input document. In this paper, we will try to explore various multi-document summarization techniques that use LDA as a topic modeling method for improving final summary coverage and to reduce redundancy. Finally, we compared LDA and LSA using the Genism toolkit, and our experiment results show that LDA outperforms LSA if we increase the number of features considered for sentence selection.

Keywords Topic modeling · Text summarization · LDA · LSA

1 Introduction

Most people nowadays rely on the Internet for most of the social and economic activities. The amount of data produced ever-increasing on daily basis. People do not find the time or spend time reading all the articles produced on the web. The need for an automatic text summarizer grows more as the reader is not interested in all the content feed to him. One of the powerful techniques to discover the underlying thematic structure of the document is topic modeling. The topic modeling

G. B. Mohan · R. P. Kumar

Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Chennai, India

e-mail: g_bharathimohan@ch.students.amrita.edu

R. P. Kumar

e-mail: r_prasannakumar@ch.amrita.edu

technique is used to find hidden topics from the document, and it is applied in Bioinformatics, software engineering, and natural language processing. Topic modeling technique latent Dirichlet allocation (LDA) [1] is widely used in automated text summarization, especially in multi-document text summarization. Multi-document text summarization generates the generalized summary from multiple documents. Based on the output of the summarizer, we can split it into extractive and abstractive text summarization. Extractive summarization creates the summary from existing sentences in the original documents. In abstractive summarization, new sentences are created by analyzing the semantics of the original document and added to the summary. The summarizer produces many sentences that can potentially be part of the output summary. Sentence ranking approaches based on semantic or statistical approaches have to be used to rank the sentence. The top k sentence can be selected to be part of the summary. The LDA technique can be used to find whether rank the sentence is based on semantics, whether the selected sentence reflects the hidden topic present in the document.

In this paper, we get motivated to provide a comprehensive survey on topic modeling techniques for extractive multi-document, abstractive, and hybrid text summarization. The organization of the paper is as follows: Sect. 2.1 explains the structure of the LDA model which can be viewed as a distribution of topics over documents. In Sect. 2.2, LSA we focus on analyzing distributional semantics for term-document matrix and how SVD is used to reduce the matrix. Section 2.3 explains how the topic modeling method is used in text summarization. Section 3 contains a literature survey of various topic modeling methods used in three categories of text summarization, namely extractive, abstractive, and hybrid model. In Sect. 4, we presented a table format of comparing topic modeling and summary to improve coherency and to reduce redundancy. Section 4 contains the simple implementation details to compare LDA and LSA using Gensim. Finally, a summary of this paper is presented as the conclusion in Sect. 5.

2 Related Concepts

2.1 Latent Dirichlet Allocation

Latent Dirichlet allocation is a generative probabilistic model to represent corpus or collection of discrete data [2]. The peer idea is to represent the document in the mixture of latent topics, and topics are the distribution of words. LDA is unsupervised learning where documents are viewed as bags of words. LDA assumes that documents are generated by formulating a set of topics, and in turn, topics are identified by using a set of words. To understand how LDA works, let us assume the following assumption. Figure 1 shows the structure of the LDA topic model; assume there are M topics across D documents. By assigning each word a topic, M topics can

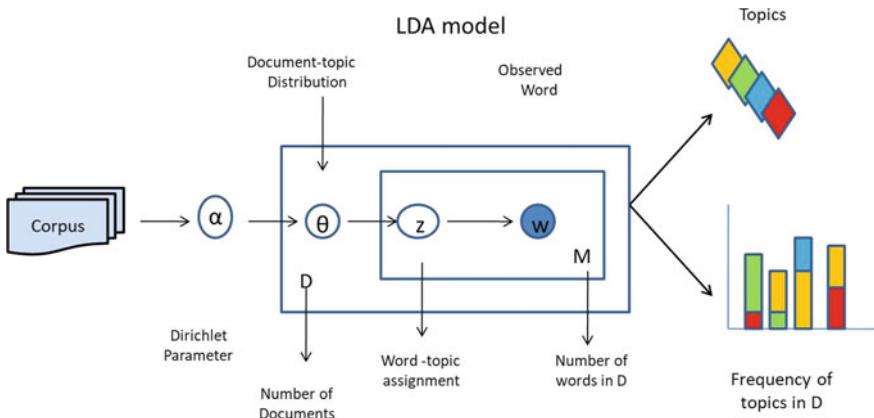


Fig. 1 Structure of LDA topic model

be distributed over D document, this distribution is also called per-document topic distribution (α). Assign each word to a topic based on what topics are present in the document and how many times the word w assigned to the topic across the entire document (per—topic word distribution β).

2.2 Latent Semantic Analysis (LSA)

LSA is one of the important topic modeling techniques in natural language processing (NLP). Mostly it is applied in distributional semantics for analyzing the relationship between the set of terms and documents. A matrix with rows as words and columns as the document is constructed from the large text. The resulting matrix may be large, so the mathematical technique singular value decomposition (SVD) is used to reduce the dimension of the matrix without altering the similarity structure of the document. LSA method learns latent topic information by a decomposing term-document matrix as shown in Fig. 2. Let us consider D as term-document matrix, it is decomposed into three matrices d , S , and t such that $\{D\} = \{d\} \{S\} \{t\}$.

2.3 Topic Modeling LDA in Text Summarization

One of the main problems in text analysis is how to determine the concept of the document. A human can understand the event or concept of the document, but the computer cannot do it. To enable a computer program to understand the concept in the document, topic modeling techniques are used. Topic modeling [3] is a popular technique for extracting hidden information from the collection of documents or

SVD of the LSA topic modeling method

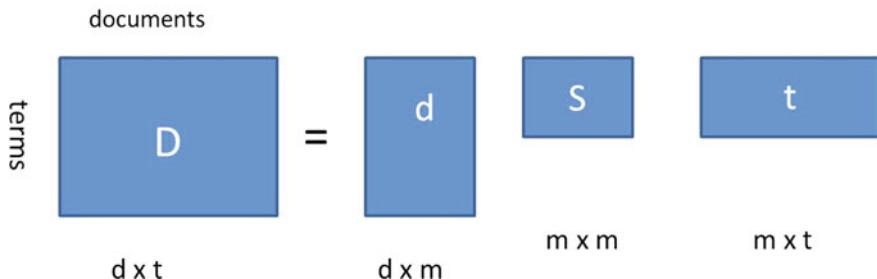


Fig. 2 LSA topic modeling using SVD

corpus. It is more suited for text analyzing bioinformatics, social, and environmental data. A detailed survey on standard topic model approaches [4–6] discussed various topic model approaches. One of the basic approaches is LDA, and it performs better in inferring the unseen documents compare to previous work [7]. The estimation of parameters for maximizing the log-likelihood of data and sparsity in handling large corpus is major drawbacks of LDA. In this paper, we will discuss how topic modeling is used in text summarization. Figure 3 explains how topic modeling is used for text summarization using the LDA approach. The input from multiple documents is preprocessed and converted to sentences to form a single large document. The topic modeling LDA method is applied to convert the sentence to clusters of sentences based on the latent topic generated from LDA. The sentence for each one of the topics is selected to improve coherency. The semantically related topics can be identified by performing clustering; sentences are selected from each one of the clusters to reduce redundancy. The sentences are ordered based on the weights to improve coherency and weights to reduce redundancy, and the top k sentence is selected to be part of the final summary.

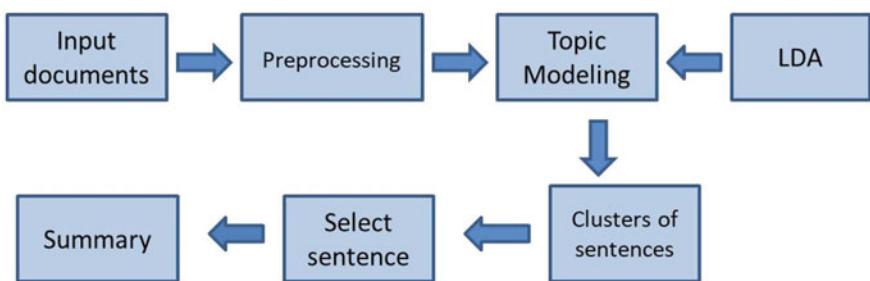


Fig. 3 Topic modeling LDA in text summarization

3 Literature Survey

Automatic text summarization approaches are divided into three categories based on their output summary. In an extractive text summarizer, important sentences are selected from the original document only and are displayed based on their sentence ranking. Summary with new content based on the existing document will be created in abstractive text summarizer. The hybrid approach combines the best of both approaches to overcome the shortcomings of extracting and abstracting methods.

3.1 Extractive Text Summarizer

Extractive text summarization is based on unsupervised graph-based approach [8, 9] where the author proposed the representation of sentences in the form of a graph. Sentences are represented as a node of a graph, and an edge is connected between two nodes if there is a similarity between the sentences. The task of assigning weights to the edge is difficult in the graph-based method. The author used the LDA method to find the topic words in the document. The weights are assigned based on the similarity between sentences, and how sentences reflect the important topics in the document. The author used the LDA approach to find the topic in the document. LDA is used to calculate the percentage of words in the document assigned to a topic. Finally, percentage of the times word is assigned to T in the overall document. Multi-document summarization is based on sentence clustering proposed by the author [10]. The author used the LDA method to select the representative sentence to be part of the final summary. Latent semantic indexing (LSI) is used to find the similarity between the sentences that form a cluster. The author used the similarity-based histogram clustering (SHC) method to improve the coherence of the clusters. The clusters are sorted based on the weights based on the number of words in the document. Finally, the weights are added to sentences based on the LDA method to select the representative sentence to be part of the summary. The probabilistic methods such as Hollinger's distance method, Jensen–Shannon divergence method, and KL divergence methods are combined with LDA which was proposed in [11]. The proposed LDA reduces the size of the large corpus in a step manner and preserves the important topics of the document. The classifier is used in the proposed system to select the important sentence from the generated summary. The author also suggested the formation of clusters based on the topic sentence and the selection of representative sentences from each cluster to improve cohesion.

An approach that combines K means clustering and LDA gives a better summary based on a specific topic which is proposed in [9, 12]. Here author used the K means clustering algorithm on the input dataset, and the inference LDA process is applied to each cluster of documents. Inference LDA is the process of retrieving information from a document. Finally, the sentence LDA approach is applied to make the document a topic representation, and the same is used to weigh the sentence

significance in the summary. One of the drawbacks of the clustering-based method is that it ignores informativeness of the words, and that of LDA is, it tends to extract long sentences and suffer from redundancy. The informativeness of the summary can be improved by using the LDA method which captures the theme of the document [13]. The author has proposed density peaks clustering method to represent clusters centers based on the density. The algorithm is based on the concept that sentences will be more relevant if it possesses higher density. They used an integrated sentence scoring method based on the four objectives: informative score, relevance score, diversity score, and length constraint. Some papers proposed, feature-based extractive text summarization [14], where features are chosen based on application and kind of data.

3.2 *Abstractive Text Summarizer*

An unsupervised framework for generating abstractive summaries was proposed by Alambo et al. [15]. Topic model LDA is used to find the optimal number of topics. Redundant keywords in topics generated are removed by using hierachal agglomerative clustering (HAC). The author introduced the concept of abstractive language unit (ALU) for generating the headline/title of the document. Summary from extractive language unit (ELU) used to generate ALU clusters, and then they used the abstractive score to select the ALU that gives the highest score. An approach that combines both traditional approach and abstractive was proposed in [16]. Multiple documents are converted into a single document, and a deep learning algorithm is applied to generate the final abstract summary. At first, an improved LDA algorithm is used to cluster sentences in all the documents. The Extended LexRank algorithm is used by the author to sort the sentence in each cluster. Then sentences are compressed using the Hedge Trimmer algorithm, and finally, integer linear programming is applied for selecting sentences to get a single document. Finally, the tensor flow algorithm is used for the final summary. Aspect-based extraction based on two-step LDA is proposed [12] for review summarization. LDA is used mainly to capture the aspect-related words in the reviews. The duplicate aspect can be removed by clustering approach using word2vec to capture the common words in the aspects. Abstract summary suffers from the problem of redundancy, incoherency, and inaccurate comprehensive summary. In multi-document LDA modeling [17], LDA has three layers of structure of words, topics, and documents to find the shallow topics. The sentences are ranked by combining word frequency, length and location characteristics, and sentence and title similarity. Sentence with high rank is extracted, and it is compared with abstract sentence, if the similarity is above a threshold, sentence will be included in the final abstract summary.

3.3 Hybrid Text Summarizer

Some authors [18] used topic modeling by extending the ARTM algorithm as it performs well for large corpus, and it does not require an increase in the parameter as the documents grow. The author used the topic model to build a dictionary with topics ranked on their weights by constructing a unigram model, and they extended it by adding the multiword terms. A hybrid architecture that covers the benefits of both extractive and abstractive text summarization was proposed by the author [19]. The author initially used the extractive method to generate an input feed summary for the abstractive method. The input feed summary has an original document combined with an output summary of the extractive method. In the abstractive method, recursive neural network (RNN) is used to predict final summary tokens. The SVM classification is used to classify the topics in a new article into one or more class labels such as crime and political entertainment. Text ranking algorithms such as Text rank, LDA, and Tf-idf are implemented in its library. A hybrid approach [20] used clustering and SVM to generate a summary by avoiding redundancy among words in the document.

4 Evaluation and Experiment Results

We conducted the experiment setup to compare the performance of LDA over LSA topic modeling using the mostly used NLP toolkit Gensim. In our experiment, we used 20 newsgroup data and evaluated the performance of topic modeling methods using statistical measures such as precision, recall, and F-measure.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (1)$$

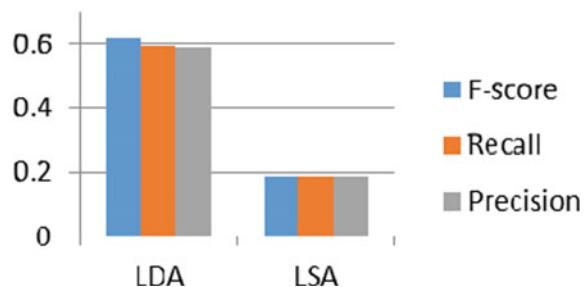
$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (2)$$

$$F\text{-score} = P.R/(P + R) \quad (3)$$

where TP termed as number of important words detected as topics, FP as number of unimportant words detected as topics, TN as number of unimportant words detected as non-topics, and finally FN as number of important words detected as non-topics.

In our experiment, we increased the number of features and computed the F-score for more iteration. Our experiment results show that if we increase the number of features, LDA performs consistently better compared to LSA. Figure 4 shows the *F*-score, precision, and recall for both LDA and LSA for 10,000 features. We formulated table as shown in Table 1 to discuss briefly how and what type of topic modeling used in different summary models such as extractive summarizer, abstractive summarizer, and hybrid summarizer in recently proposed papers.

Fig. 4 F -score, precision, recall with features $f = 10,000$



5 Conclusion and Future Scope

The amount of information available on the Internet makes it a natural requirement for any reader to get a summary from it regarding the topic they search for. So there is a need for an automatic text summarizer which is inevitable. Summarizer will not always produce a relevant and coherent summary for the topic searched by the user. The topic modeling method can be used to identify topics in a document, and sentences can be selected in case of extractive multi-document summarization or it can be generated in case of abstractive summarization. In this paper, we try to give a detailed survey on topic modeling methods used in text summarization. Most of the paper used the LDA method to identify the topics in the collection of documents and used it as one of the parameters to add weight to sentence selection. Our experiment result also shows that LDA consistently performs well with the increased number of features and topics. In the future work, we try to use LDA topic modeling to build an automatic text summarizer by combining the features of both extractive and abstractive summarizers.

Table 1 Comparison of topic modeling techniques in text summarization

Years	Authors	Topic modeling	Summary model
2020	Tatiana et al.	ARTM algorithm unigram model built and it is extended with multiword expression	Hybrid text summarizer Rhetorical analysis and text transformation are applied to form quasi-abstract
2020	Ramesh et al.	LDA—used to identify important topics in the document. Used in sentence score estimation to be part of the summary	Graph-based Extractive text summarizer Weights are assigned to the edge of the graph based on sentences reflecting the topics of the document and similarity between sentence nodes

(continued)

Table 1 (continued)

Years	Authors	Topic modeling	Summary model
2019	Rajendra Kumar et al.	LDA—to identify optimal no of topics To reduce sentence length	Extractive text summary—Summary generated using fifteen important features of a sentence
2020	Ruby Rani et al.	Lexical LDA—to identify the structure of sentence	Extractive text summary—sentence weights are calculated based on lexical features. Higher ranked sentences are selected to be part of the summary
2019	Junail et al.	Fuzzy topic modeling- used to generate global and local term frequency PCA used to remove negative impact global term weights	Extractive summary-generated from a short text
2017	Zongdo Wa et al.	LDA—probability of word in a novel document is calculated to find whether the word is important topics related to documents	Extractive summary-sentence to be part of the summary is calculated based on diversity and redundancy
2020	Thir Ahmed et al.	Topic modeling based on SVM—SVM algorithm is used to classify the documents based on topics	A hybrid approach for text summarization. The extractive method is based on sentence features. Abstractive method train RNN uses LSTM with pointer generation attention mechanism
2020	Amanuel Alambo et al.	LDA—used to generate topics and improve coherence Topical Hierarchical Agglomerative clustering—used to remove a semantically redundant keyword	Hybrid model—extractive phase followed by an abstractive phase Extractive phase to maximize coverage and abstractive phase to generate ALU for the abstract summary
2018	Wang et al. [21]	Topic embedding using LDA LDA is used in pre-training phase to assign input text with topics LDA and Gibbs sampling is used to train corpus	Abstractive summary based on convolutional sequence to sequence framework Topic aware attention mechanism based on probability-based mechanism to generate coherent summary

References

1. Blei DM, Ng AY et al (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3(2003):993–1022
2. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichelt allocation. *J Mach Learn Res* 3:993–1022
3. Jelodar h, Wang Y et al (2018) Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. [arXiv:1711.04305v2](https://arxiv.org/abs/1711.04305v2) [cs.IR]
4. Vayansky I, Kumar SAP (2020) A review of topic modeling methods. *Inf Syst PII* S0306–4379(20):30070–30073
5. Vayansky I, Kumar SAP (2020) A review of topic modeling methods. *Inf Syst.* <https://doi.org/10.1016/j.is.2020.101582>
6. Albalawi R, Yeap TH, Benyoucef M (2020) Using topic modeling methods for short-text data: a comparative analysis. *Front Artif Intell*
7. Kherwa P, Bansal P (2019) Topic modeling: a comprehensive review. *EAI Endorsed Trans Scalable Inf Syst*, 10 2019–01 2020
8. Belwal RC, Rai S, Gupta A (2020) A new graph-based extractive text summarization using keywords or topic modeling. *J Ambient Intell Human Comput*
9. Uma Shankari E, Krishna Rao NV et al (2020) Multi-document text summarization using genism. *Int J Adv Sci Technol* 29(12):1362–1370
10. Lisjana OA, Rini DP, Yusliani N (2019) Multi-document text summarization based on semantic clustering and selection of representative sentences using latent Dirichlet allocation. *Adv Intell Syst Res*, 172
11. Roul RK (2020) Topic modeling combined with classification technique for extractive multi-document text summarization. Springer-Verlag GmbH Germany, part of Springer Nature
12. Das SJ, Murakami R, Chakraborty B (2020) Development of a two-step LDA based aspect extraction technique for review summarization. *Int J Appl Sci Eng* 18(1):2020120K
13. Wang B, Zou Y et al (2018) Multi-document summarization via LDA and density peaks based sentence-level clustering. *CCIS* 873:313–323
14. Chiney RP, Prasanna Kumar R (2020) Extractive summarization approach for news articles based on selective features. *Int J Adv Sci Technol* 29(6):8215–8224
15. Alambo A, Lohstroh C, Madaus E et al (2020) Topic-Centric unsupervised multi-document summarization of scientific and news articles, arXiv.org. cs, [arXiv:2011.08072](https://arxiv.org/abs/2011.08072)
16. Zhong Y, Tang Z, Ding X, Zhu L, Le Y (2017) An improved LDA multi-document summarization model based on tensorflow. In: International conference on tools with artificial intelligence 2375–0197/17
17. Dan T, Yu S (2020) Multi-feature automatic abstract based on LDA model and redundant control. *J Phys Conf Ser* 1693:012211
18. Batura TV, Bakiyeva AM, Charintseva MV (2020) A method for automatic text summarization based on rhetorical analysis and topic modeling. *Int J Comput* 19(1):118–127
19. Shaik TA, Vikas A, Pradyumna GVN (2020) Hybrid approach for text summarization with topic modelling and entity extraction. *Int Res J Eng Technol*. p-ISSN: 2395-0072
20. Mab Shiva Kumar K, Soumya R, Text summarization using clustering technique and SVM technique. *Int J Appl Eng Res* 10:25511–25519, 201
21. Wang L, Yao J et al (2018) A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence (IJCAI-18)
22. Twinandillaa S, Adhya S, Surarso B, Kusumaningruma R (2018) Multi-document summarization using K-means and latent Dirichlet allocation (LDA)—significance sentences. In: International conference on computer science and computational intelligence

Design and Analysis of Microstrip High-Frequency Filter



V. Reji, R. Arthi, and C. T. Manimegalai

Abstract In this paper, a design and analysis of wideband bandpass filter with compact size are proposed. Chebyshev and elliptic approximation model is considered for the filter design. The filter is designed, and the analysis is carried out by three, five and seven parallel coupled lines. The model can be used for GSM, WiMAX and WiFi applications for both licensed and unlicensed frequency bands. The filter is designed to give a high gain and less insertion loss, while operating in the bandwidth range of 2–5.2 GHz. The proposed filter is fabricated on an FR4 substrate with a relative dielectric constant of 4.4 and 1.6 mm of substrate thickness and fed by two $50\ \Omega$ microstrip lines.

Keywords Low pass · High pass · Band pass · Microstrip

1 Introduction

In recent years, designing a high-frequency bandpass filter has become quite a challenge. Filters like Butterworth and Chebyshev have become a bit easy nowadays as the filters are already being used for telecommunication and military applications. Filters like elliptic and Bessel have never been popular as they have a complicated design structure. As we research more, we can see that though these filters are more complicated they have a better response when compared to other filters. The elliptic filter has ripples in both passband as well as stopband. It has a very fast transition and also has a high level of rejection. Thus, elliptic filters can be made to work inside a specified range of frequency. Elliptic filters can be modified to act as both Butterworth as well as Chebyshev filter. If the ripples in stopband approaches zero

V. Reji (✉) · R. Arthi · C. T. Manimegalai

Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India

e-mail: rejiv@srmist.edu.in

R. Arthi

e-mail: arthir2@srmist.edu.in

it acts as Chebyshev I, if it approaches 1 in passband it acts as Chebyshev II and if ripples in both the band approach zero it acts as a Butterworth filter.

A wideband filter has been demonstrated with an extremely upper stopband for UWB band applications [1]. Fifth order Chebyshev approximation is used for suppressing four stopband harmonics in the above filter. Another one ultra-wideband bandpass filter has been demonstrated [2] with loaded open stub on dual-ring resonator. The next wideband filter is developed for millimeter wave applications. To reduce the size of the filter a complicated co-planar waveguide structure is used with cross-over in multilayer [3]. A GaAs substrate with nitride dielectric layer UWB filter realization was given in [4]. In this model, four inductors were used to design the filter by Chebyshev approximation. The gain of the filter is lower compared with other filters. The next paper described a compact ultra-wideband quasi-elliptical bandpass filter with a via-free multimode resonator [5]. This structure consists of a step impedance transmission line resonator that is loaded centrally with a folded open-circuited stub. A chained-elliptic function waveguide bandpass filter has been implemented with fewer pass band ripples and low tolerance [5]. Another miniaturized BPF with a notch band has been presented in this paper [6]. A compact filter with transmission line technology for harmonic suppression was given in [7]. This bandpass filter was designed by combining a low pass filter and a high pass filter. Next, a multi-mode resonator filter was presented in [8]. This filter is formed by open-circuited transmission line sections with coupled lines. Low pass filter and bandpass filter with ultra-wide stopband have been demonstrated in [9].

In this paper, a microstrip elliptic wideband bandpass filter is proposed with three, five and seven parallel coupled lines without any radial open stubs. The filter is fabricated on FR4 substrate having a dielectric constant of 4.6 and a thickness of 1.6 mm. The aim of the project is to reduce the size of the filter, increase the gain and reduce the insertion loss.

2 Filter Design

The system is designed as a bandpass filter. The reason that a bandpass filter [10, 11] is used because it is a circuit that allows signal between two specific frequencies to pass but discriminates against signal at other frequencies. Bandpass filter is used as it limits the bandwidth to the minimum necessary. The filter has to be designed in such a way that it operates only inside the given specified range of bandwidth and rejects the other frequencies. Hence, bandpass design is the most suitable design for such a system. Moreover, the bandpass filter also has fewer signals to noise ratio, more speed and more number of signals and a minimum interface which makes the filter more efficient. Constant k filters are the prototype filters that have the ideal filter characteristics.

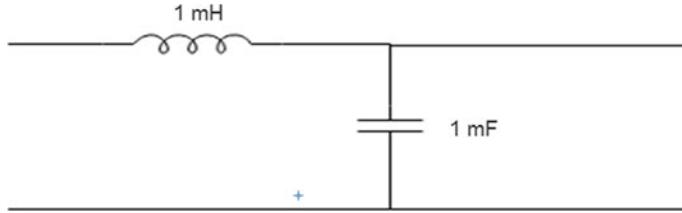


Fig. 1 Prototype filter

2.1 Prototype Filter Design

Every filter design starts with a prototype filter which is used as a reference for the required design. The ideal characteristics of the prototype filter [12, 13] are matched with the proposed system and the efficiency of the filter is calculated. Constant K filters are basically low pass filters that are converted to bandpass filters and then to the required RF design using Kuroda's identity (Fig. 1).

$$K^2 = \frac{Z}{Y} \quad (1)$$

$$K = \sqrt{\frac{iw_l}{iw_c}} = \sqrt{\frac{L}{C}} \quad (2)$$

2.2 Bandpass Filter Design

Frequency transformation is done to convert normalized frequency Ω into actual frequency ω and Impedance transformation is done to convert standard generator and load resistances g_0 and $g_{(N+1)}$ to actual resistances R_L and R_G using the following equations.

$$L = L/\omega_c \quad (3)$$

$$C = C/\omega_c \quad (4)$$

$$R_G = R_G \quad (5)$$

$$R_L = R_L R_G$$

To convert the actual bandpass filter into a high-frequency filter to make the filter work under the given specified range of 2.4–5.2 GHz and reject all other frequencies, the filter has to be converted into a high frequency filter. This can be done using Karuda's transformation. Karuda's identities are used to convert lumped elements into distributed elements as shown in Eq. (7).

$$N = 1 + \frac{Z_2}{Z_1} \quad (7)$$

The elliptic filter for order 5 is designed as a Pi (π) type filter with 2 capacitors and 3 inductors. The inductors are short circuited and the capacitors are open circuited. Unit elements are added on both sides to convert the series stub into shunt stub and vice-versa. The series inductance implementation is more complicated than the shunt capacitance design. Thus, in this project, the final result is implemented as shunt stubs.

The length of the microstrip line remains the same throughout but the width of the lines varies. The width is calculated using the below mentioned equations,

$$\frac{w}{h} = \frac{8e^A}{e^{2A} - 2} \quad (8)$$

The above element values are normalized values. These are de-normalized by scaling the element values to the 50Ω line impedance. Using $\lambda_o/8 = v_p/(8f_c)$, the length is found out to be 11.35 mm where $f_c = 3.66$ GHz. The final result is shown in Fig. 3. The overall length of the filter comes out to be 5.37 mm (Fig. 3b, c).

$$A_{Z_L} = 2\pi \frac{Z_L}{Z_f} \sqrt{\frac{\epsilon_r + 1}{2}} + \frac{\epsilon_r - 1}{\epsilon_r + 1} \left(0.23 + \frac{0.11}{\epsilon_r} \right) \quad (9)$$

2.3 Filter Design for $N = 3$

First, the antenna design considered for order 3. The filter is designed from the basic prototype design as shown in Fig. 2a. Frequency and impedance transformations are used for finding the length and width [14, 15] of the microstrip filter. The designed impedance values are given in Fig. 2b, c. Table 1 shows the calculated length and width of each element from the impedance value. The frequency band selected for this design is 2.4–5.2 GHz.

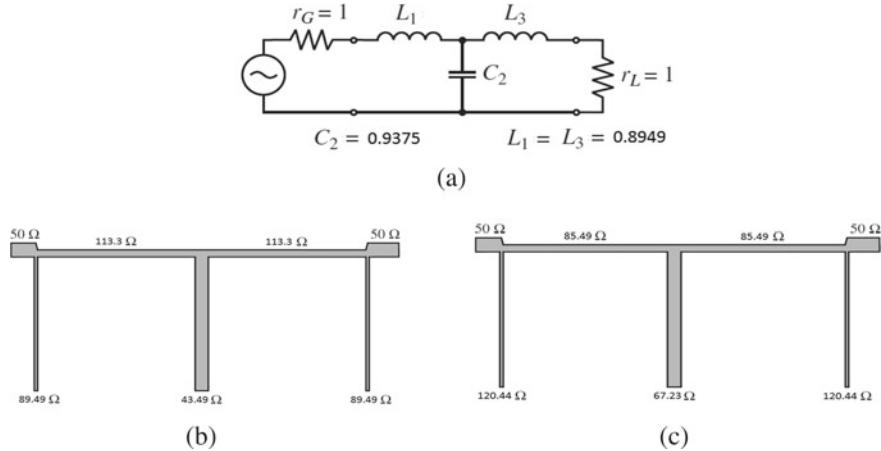


Fig. 2 Order of the filter $N = 3$. **a** Constant K filter. **b** Elliptic filter with transferred impedance value. **c** Chebyshev filter with transferred impedance value

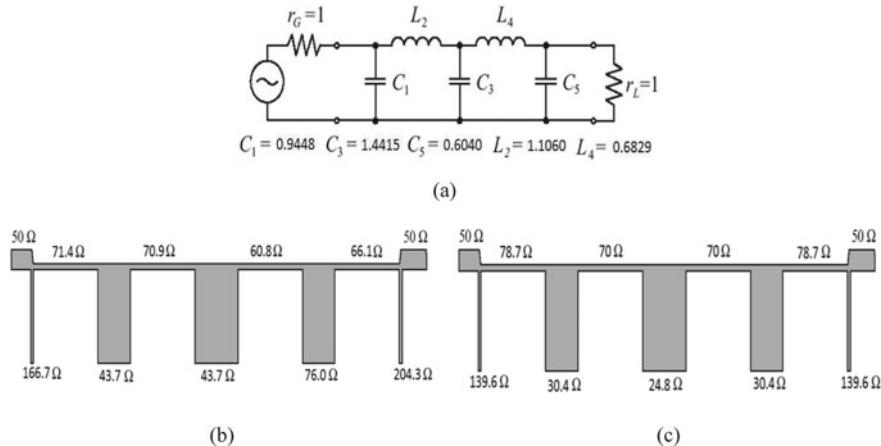


Fig. 3 Order of the filter $N = 5$. **a** Constant K filter. **b** Elliptic filter with transferred impedance value. **c** Chebyshev filter with transferred impedance value

Table 1 Element width and Length of a filter ($N = 3$)

Impedance (Ω)	Width (mm)	Length (mm)	Impedance (Ω)	Width (mm)	Length (mm)
Elliptic filter			Chebyshev filter		
Z120.4	0.74	11.1	Z89.4	0.89	11.48
Z85.4	1.00	11.49	Z113	0.456	11.68
Z67.23	1.704	11.25	Z43.4	3.66	10.91
Z85.49	0.34	10.2	Z113	0.456	11.68
Z120.4	0.372	11.733	Z89.4	0.89	11.48

Table 2 Element width and length of a filter ($N = 5$)

Impedance (Ω)	Width (mm)	Length (mm)	Impedance (Ω)	Width (mm)	Length (mm)
Elliptic filter			Chebyshev filter		
Z166.7	0.10	11.96	Z139.6	0.09	12.4
Z71.4	1.50	11.30	Z78.7	1.74	11.67
Z43.7	3.64	10.91	Z30.4	3.75	11.2
Z70.9	1.52	11.29	Z70	1.77	11.34
Z43.7	3.64	10.91	Z24.8	3.87	11.11
Z60.8	1.58	11.30	Z70	1.99	11.67
Z76	1.31	11.35	Z30.4	1.52	11.54
Z66.1	1.76	11.23	Z78.7	1.82	11.44
Z204.3	0.03	12.07	Z139.6	0.05	12.34

2.4 Filter Design for $N = 5$

Next filter design for order five. In this filter, three capacitors and two inductors are used. Initially, the filter is designed from prototype design, and then the impedance and frequency are transferred to a bandpass filter design. The prototype filter and their impedance values are driven in Fig. 3 and the width and length of the calculation is shown in Table 2.

2.5 Filter Design for $N = 7$

Next filter design for order seven. In this filter, three capacitors and four inductors are used. Initially, the filter is designed from prototype design, then the impedance and frequency are transferred to a bandpass filter design. The prototype filter and there impedance values are given in Fig. 4, and the width and length calculation are shown in Table 3.

3 Results and Discussion

It is clearly seen that the output response of the filter with order 3 for elliptical filter design is reached nearly 2.4–4.8 GHz and Chebyshev filter design is nearly 2.4–4.7 GHz as shown in Fig. 5b. For $N = 5$, maximally flat response is given from elliptical filter compared with Chebyshev filter as illustrated in Fig. 6. The elliptical filter gives the required band response of 2.4–5.2 GHz. For $N = 7$ almost elliptic and Chebyshev filter response are the same as shown in Fig. 7, but the length and width of the elliptic filter design is less compared with the Chebyshev design. The group

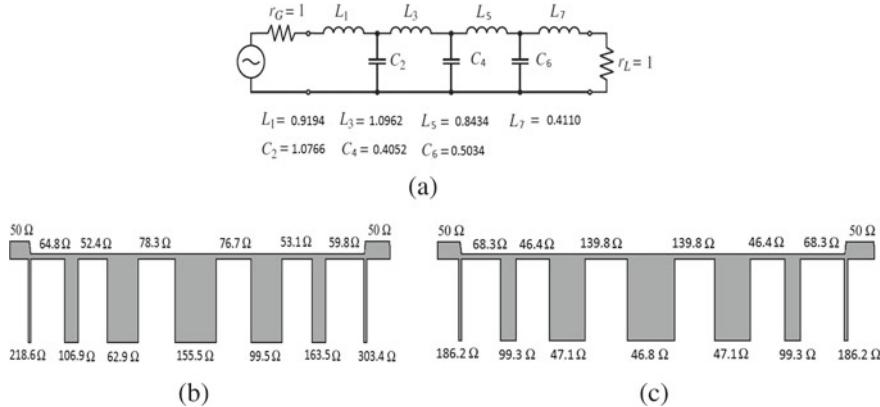


Fig. 4 Order of the filter $N = 7$. **a** Constant K filter. **b** Elliptic filter with transferred impedance value. **c** Chebyshev filter with transferred impedance value

Table 3 Element width and length of a filter ($N = 7$)

Impedance (Ω)	Width (mm)	Length (mm)	Impedance (Ω)	Width (mm)	Length (mm)
Elliptic filter			Chebyshev filter		
Z218.6	0.024	12.10	Z186.2	0.059	12.0
Z64.8	1.833	11.22	Z68.3	1.65	11.2
Z106.9	0.54	11.63	Z99.3	0.677	11.5
Z52.4	2.699	11.05	Z46.4	3.307	10.9
Z62.9	1.94	11.19	Z47.1	3.228	10.9
Z78.3	1.231	11.37	Z139.8	0.217	11.8
Z155.5	1.14	11.92	Z46.8	3.26	10.9
Z76.7	1.29	11.35	Z139.8	1.00	11.5
Z99.5	0.673	11.57	Z47.1	0.77	11.7
Z53.1	2.638	11.06	Z46.4	3.304	11.3
Z163.5	0.112	11.95	Z99.3	0.322	11.9
Z59.8	2.135	11.15	Z68.3	2.81	11.3
Z303.4	0.002	12.20	Z186.2	0.002	12.3

delay comparison is also given in Figs. 8 (a) and (b). The overall length of the elliptic filter is 12.2 mm (maximum length), and width of the filter is 16.2 mm. The overall length of the Chebyshev filter is 12.3 mm (maximum length), and the width of the filter is 20.6 mm. From the previous two filter comparison, we concluded the elliptic filter is giving a good response and less dimension compared with Chebyshev filter. So the filter is fabricated for the elliptic filter design with order five shown in Fig. 9.

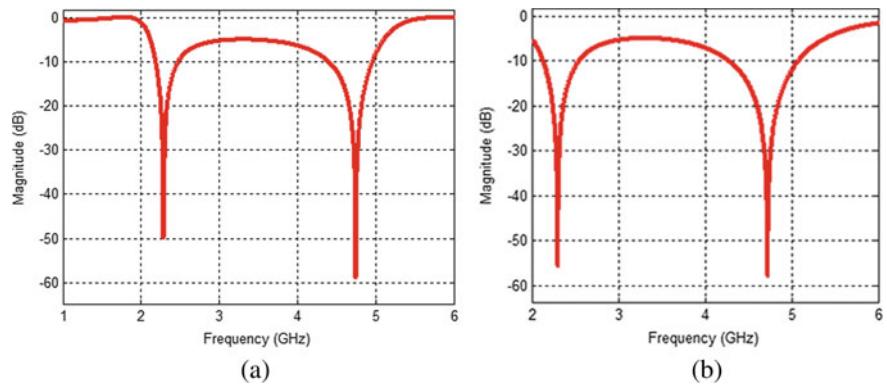


Fig. 5 Order of the filter $N = 3$. **a** Elliptic filter response. **b** Chebyshev filter response

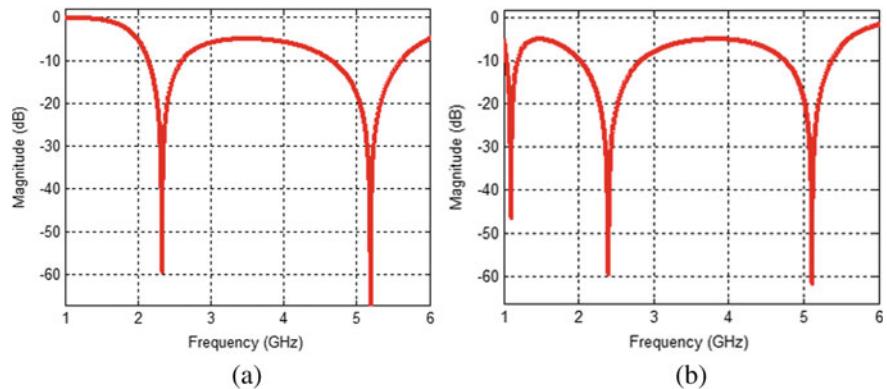


Fig. 6 Order of the filter $N = 5$. **a** Elliptic filter response. **b** Chebyshev filter response

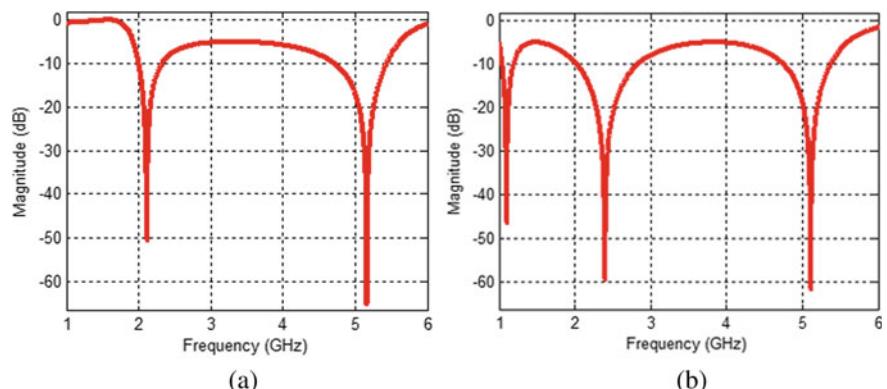


Fig. 7 Order of the filter $N = 7$. **a** Elliptic filter response. **b** Chebyshev filter response

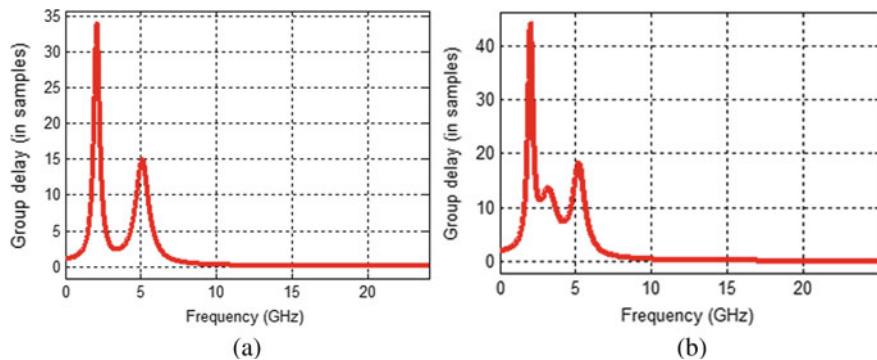


Fig. 8 Group delay of elliptic filter (order of the filter $N = 7$)

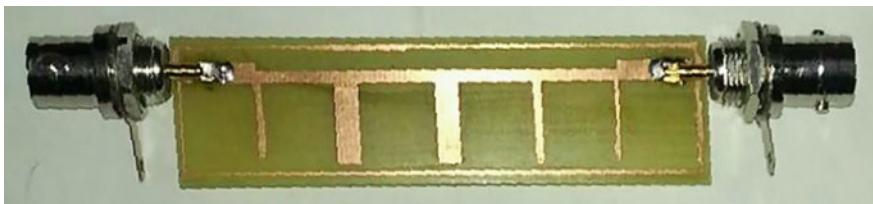


Fig. 9 Top view of the designed filter

4 Hardware Design

The filter is designed using the FR4 substrate. Both the ends of the filter are joined using a metallic via and at the end with connectors for giving input to the filter.

5 Conclusion

Thus, a microstrip wideband bandpass filter consisting of three, five and seven parallel lines with reduced insertion loss and a high gain is developed. The filter was designed as an elliptic filter. The response of the elliptic filter is high when compared with the Chebyshev filter. The filter response for $N = 5$ of the elliptic filter is equal and better than the $N = 7$ order of the Chebyshev filter. So the elliptic filter is fabricated for the order of $N = 5$.

References

1. Malherbe JAG (2018) Wideband bandpass filter with extremely wide upper stopband. *IEEE Trans Antenna Propag* 66(11):5943–5954, 0018–9480
2. Long Z, Tian M, Zhang T (2020) High temperature superconducting multimode dual-ring UWB bandpass filter. *IEEE Trans Appl Superconduct* 30(2)
3. Jia H, Mansour RR (2019) Millimeter-wave ultra wideband multilayer superconducting filter. *IEEE Trans Appl Superconduct* 29(5)
4. Hu S, Hu Y, Gao Y, Zhang X, Zhang X, Wang Z, Zhou B, Cai Z (2020) A compact UWB bandpass chip filter on a GaAs substrate with modified Chebyshev structure. 10.3390-Electronics9020313 www.11
5. Riaz M, Virdee BS, Shukla P, Onadim M (2019) Quasi-elliptic ultra-wideband bandpass filter with super-wide stopband. *Int J Electron Commun.* 10.1016, April-2019
6. Ng GS, Cheab S, Wong PW, Soeung S (2020) Synthesis of Chained-Elliptic function waveguide bandpass filter with high rejection. *Prog Electromagn Res C* 99:61–75
7. Reji V, Kavitha Sudha R (2018) Microstrip UWB bandpass filter with dual notch frequencies. *Int J Pure Appl Math* 120(6):10133–10145
8. Shaman HN, Almughamis AM, Alamro AM, Alharthi YS (2016) Compact ultra-wideband (UWB) bandpass filter with wideband harmonic suppression. In: Proceedings of 21st international conference on microwave radar and wireless communications (MIKON), May 2016, pp 1–4
9. Zhang T, Xiao F, Tang X, Guo L (2016) A multi-mode resonator based UWB bandpass filter with wide stopband. *Int J Microw Wireless Technol* 8(7):1031–1035
10. Xu J, Ji Y-X, Wu W, Miao C (2013) Design of miniaturized microstrip LPF and wideband BPF with ultra-wide stopband. *IEEE Microw Wireless Compon Lett* 23(8):397–399
11. Shi SY, Feng WJ, Che WQ, Xue Q (2013) Novel miniaturization method for wideband filter design with enhanced upper stopband. *IEEE Trans Microw Theory Techn* 61(2):817–826
12. Zhang R, Luo S, Zhu L, Yang L (2017) Synthesis and design of miniaturized wideband bandpass filters with scaled transmission line for spurious-response suppression. *IEEE Trans Microw Theory Techn* 65(8):2878–2885
13. Malherbe JAG (2014) Pseudo-elliptic function ultra wideband bandstop filter with stepped impedance stubs. In: Proceedings of 44th European microwave conference, Rome, Italy, Oct 2014, pp 536–539
14. Marimuthu J, Bialkowski KS, Abbosh AM (2015) Compact bandpass filter with multiple harmonics suppression using folded parallel-coupled microstrip lines. In: Proceedings of Asia-pacific microwave conference (APMC), vol 2, Dec 2015, pp 1–3
15. Chen F-C et al (2016) Design of wide-stopband bandpass filter and diplexer using uniform impedance resonators. *IEEE Trans Microw Theory Techn* 64(12):4192–4203

Arduino UNO-Based Smart Hand Gloves for Physically Challenged People



B. V. Santhosh Krishna, L. Sowmya, Neetha Nataraj, Nivedita Salimath, and ShivaniYadav

Abstract Smart hand gloves assist impaired individuals with living with typical individuals. This paper aims to aid auditory and speech impaired people by providing them a means to communicate basic lines without having to learn sign language, or as a temporary communication device until they can learn sign language. The glove has three flex sensors that run along the length of any three fingers. The flex sensors detect when a finger has been moved due to a change in their bend angle and produce an accordingly varying output resistance. This change is given as input to Arduino UNO that in turn, activates a pre-determined sentence to be displayed on a 16×2 LCD module. The pre-determined sentences can be changed by changing the Arduino source code accordingly, and thus, the device can be easily customized to the user's needs.

Keywords Arduino · UNO · Sign · Gloves · Sensors

1 Introduction

Smart hand gloves assist genuinely tested individuals with carrying on with an ordinary existence with typical individuals. As speech disabled individuals can't talk, these gloves enable them to change over their hand signal into text and pre-recorded voice [1, 2]. Gloves likewise assist ordinary individuals with understanding what they are attempting to state and reply or act appropriately. This smart glove has the advantage of simple control from which a disabled individual becomes free to live. The main goal of the executed venture is to build up a solid, simple to utilize, weightless keen hand glove framework is useful in limiting the impediments for debilitate individuals where they can remain with the race. As impaired individuals can't talk, these smart gloves enable them to change over his hand signal into text and if required, pre-recorded voice [3]. This additionally assists typical individual with understanding what he is attempting state and answer likewise. In this paper,

B. V. S. Krishna (✉) · L. Sowmya (✉) · N. Nataraj (✉) · N. Salimath (✉) · ShivaniYadav (✉)
Department of Electronics and Communication Engineering, New Horizon College of
Engineering, Bengaluru, India

we attempt to aid users who are not very sharp. Around nine billion individuals on this earth are physically challenged like deaf and dumb people. The correspondence between a hard of hearing ordinary visual individuals is very cumbersome [2].

Physically challenged people can talk rarely by many artificial methods such as showing gestures and postures using typical sign language [4] which helps them in day-to-day life. To make their life easy this is induced gesture-based communication is widely used in society. It is a non-verbal type of communication which is found very difficult among all of hearing networks on the planet [5]. The dialects don't have a basic beginning and subsequently difficult to decipher. This venture plans to encourage individuals by introducing method of glove-based correspondence which acts as mediating framework for communication purposes and etc.

The glove is fitted inside with 3 bend sensors. For every particular motion, the bend sensor produces a relative change in opposition. The handling of these hand motions is in Arduino UNO Board which is a development rendition of the Arduino microcontroller and the Arduino IDE programming [4]. It contrasts the information signal and predefined voltage levels put away in memory. As indicated by that necessary sound is created which is put away in memory with the assistance of speaker. In such a manner, it is simple for not too sharp to speak with ordinary individuals.

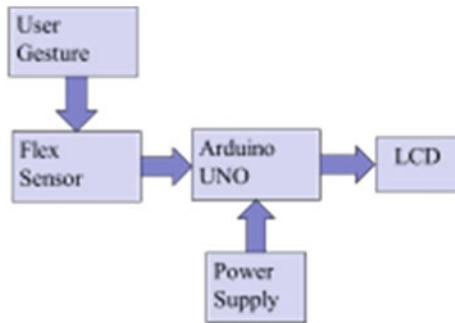
In our life, we as a whole go over many crippled individuals, some of them are incompletely and some are totally impaired. The somewhat debilitated individuals like who are imbecilic, hard of hearing, loss of motion in single leg or hand deals with their existence with troubles and feel confined from others [2]. Consequently, with the assistance of this undertaking, the obstruction looked by these individuals in speaking with the general public can be diminished by and large.

2 Proposed Method

The proposed glove houses an electronic circuit that certain hand movements for the deaf and dumb in order to eradicate the information transmission barrier between the mute and the general public. The cruxes of this paper lies in the Arduino UNO chip and flex sensors. The proposal is to rely on a system that can effectively, within acceptable error, translate gesture-based communication to linguistic communication. The flex sensors convert motion caused by gesturing into a varying output voltage that can be interpreted by the Arduino and converted into visible text that can be displayed on the LCD (Fig. 1).

3 Hardware Components

1. Arduino UNO: It is an ATmega328P or microcontroller board and can be accessed by anyone. This board contains both digital and analog input/output (I/O) pins that are useful to be interfaced to various external boards and other

Fig. 1 Block diagram

circuits[4]. Consists of 14 digital I/O pins (6 pins are capable of PWM output), 6 analog I/O pins, and is programmable with the Arduino IDE, through USB cable of type B. It is powered by 9v battery or USB cable.

2. Flex Sensor: It is a 2-terminal device and also termed as bend sensor[3], useful in measuring the amount of bending and deflection. This bend sensor is attached to the surface and its resistance of sensor element is varied when the surface is bent. No terminals like diode are present, and hence, there is no existence of positive and negative.
3. LCD: This LCD is mostly basic module. It mainly uses liquid crystal to manufacture a visible image. In this 16×2 display the information is present in two lines and each line holds 16 characters. It uses 5×7 -pixel matrix to display every character.

4 Algorithm

Working Principle

1. When the user bends one of the three fingers-thumb finger, index finger or middle finger, for more than 45° , a stimulus is sent to the Arduino UNO board.
2. The Arduino UNO board, based on the program loaded on it, converts it into a text message output.
3. The Arduino UNO board send its output to the LCD Display.
4. The LCD display displays the pre-determined message for that particular finger movement (Fig. 2).

5 Results and Discussion

The flex sensors attached to three fingers had three individual stimuli, which could be converted to one of three lines of basic communication. The circuit was realized on a breadboard, and the Arduino was programmed on Arduino IDE. It was observed

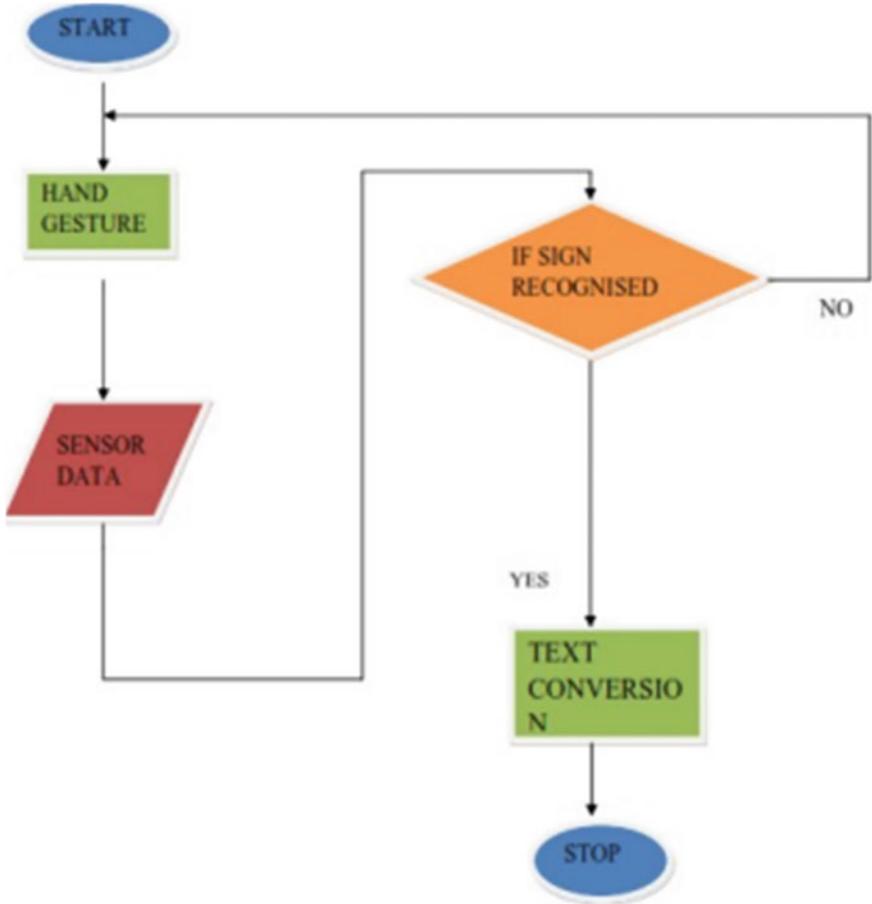


Fig. 2 Flow diagram

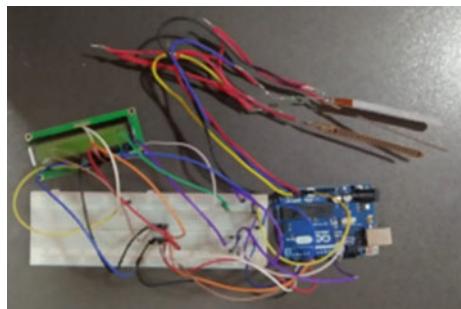
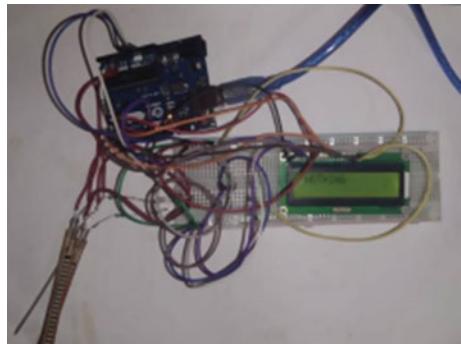
that finger motion successfully generated the pre-determined messages assigned to it. The circuit and the results of the paper are as shown in Fig. 3.

The circuit of the smart glove, after attaching glove is shown in Fig. 4.

Figure 4 depicts the circuit that has been realized, before being attached to a glove. It can be made more compact by replacing the Arduino UNO with Arduino Nano and changing the source code accordingly, and by using a PCB instead of a breadboard (Fig. 5).

Advantages

1. The gloves enable auditory and vocally disabled people to carry out basic communication that is very much required to function in a society.
2. The gloves do not require either the user or the person the user intends to communicate with to know American Sign Language.

Fig. 3 Glove model**Fig. 4** Circuit of the method**Fig. 5** Output circuit

3. The gloves save time by converting a single gesture into a line of text.

6 Conclusion

The main aim of this paper is to reduce the gap of communication between the physically challenged people and the normal world. On usage of data gloves, physically challenged people can also develop in the community and make the country better

place, as the number of physically challenged people are billions in count. It makes the future of them better and automatically leads in development of the country.

References

1. Jamdal A, Al-Moflehi A (2018) On design and implementation of a sign-to speech/text system. In: 2017 international conference on electrical, communication, computer and optimization techniques (ICEECCOT), 08 Feb 2018
2. Suri A, Singh SK, Sharma R, Sharma P (2020) Development of sign. language using Flex sensors. In: 2020 international conference on smart electronics and communication (ICOSEC), 07 Oct 2020
3. Praveen N, Karanth N, Megha MS (2015) Sign language interpreter using a smart glove. In: 2014 international conference on advances in electronics computers and communications, 08 Jan 2015
4. Ahmed SF, Ali SMB (2010) Electronic speaking glove for speechless patients, a tongue to a dumb. In: 2010 IEEE conference on sustainable utilization and development in engineering and technology
5. Sengupta A, Mallick T, Das A (2019) A cost-effective design and implementation of Arduino based sign language interpreter. In: 2019 devices for intergrated circuits (DevIC)

An Enhanced Haze Removal: Using DCP and Enriched-Invariant Features



Reenie Tanya, Faaiz Akhtar, and Mahipal

Abstract Removal of atmospheric turbulence from an image is a task that has been under study for some time now. When a picture is captured, it is corrupted due to the suspended light particles in the air. Fog-free or haze-free images are needed for better visualization and image information retrieval. This paper reports different mist evacuation calculations on how haze is taken out and improved perceivability. The defogging procedure we have further utilized is the dark channel prior algorithm framework which has proved to be a better option in obtaining the features from a hazed image. We have used the Laplacian-based gamma correction for the haze thickness estimation, which refines the generated transmission map. The next phase utilizes the White Patch Retinex technique that restores the image's white balance for visibility restoration. This paper has introduced an additional step of identifying and enhancing the critical features in the resulting dehazed image using the SIFT descriptors. Analysis exhibits that the current methodological analysis creates dehazed outcomes that have an improved ocular calibre with sharpened and contrast features compared to the simple frameworks available.

Keywords Dark channel prior · Fog expulsion · Image improvement · Performance assessment · Image enhancement

1 Introduction

Pictures taken outdoor are corrupted due to low lighting and unfavourable climatic conditions. Images taken in such atmospheric conditions are affected before arriving

R. Tanya (✉) · F. Akhtar · Mahipal
SRM Institute of Science and Technology, Ramapuram, Chennai, India
e-mail: eeniet@srmist.edu.in

F. Akhtar
e-mail: fs7972@srmist.edu.in

Mahipal
e-mail: mr6097@srmist.edu.in

at the camera because of vaporizers like residue, mist and water-drops mixed with the sealed shut, surrounding light that ponders the view. The pictures taken of cloudy scenes, this interaction substantially affects the taken a photograph, prompting the deficiency of difference and ocular calibre that takes obstructions in some PC ocular appeals in reconnaissance, insightful automobiles, and outside object acknowledgement and so on. Haze expulsion or fogging has subsequently concentrated broadly in the PC ocular area.

The climatic dissipating model regularly portrays the picture arrangement with cloudy views and literature; more methodologies have been proposed as of late dependent on this model. By and large, because of the equivocalness of the dehazing issue, those strategies can be partitioned into three classifications: investigating priors or requirements on the foggy picture, learning a model of picture highlights and scene transmission utilizing extra data of the picture scene.

Removing the haze from the images has gotten attention in various fields like autonomous driving [1], video surveillance in smart city setups, computer vision and applications, and the aviation industry and National Security [2, 3]. This has led to the introduction of more profound learning-based dehazing strategies, including neural models for better haze-related element extraction and transmission assessment.

A description of the existing dark channel prior framework followed by a study of some contributing research works that have utilized such frameworks for dehazing. The following section gives the workflow of the proposed idea followed by the obtained results and observations.

1.1 Dark Channel Prior

Haze removal of images has attracted much research and has gained importance in computational photography. The emphasis is to increase the visibility [4, 5] and the naturalness of the scene at hand. Another objective is to remove the effects of artificial lights degrading the quality of the images further. The dark channel prior algorithm has garnered much attention since it was first introduced [6] and is considered a breakthrough in this field since it has proved to give better-dehazed images than its predecessors.

The DCP algorithm works on the principle that a foggy image formed can be mathematically formulated as:

$$I(x) = J(x)t(x) + A(1 - t(x))$$

Here: $I(x)$ —hazed image component

$J(x)$ —haze-free image component

$t(x)$ —transmission map component

A —atmospheric light component.

Here, the transmission map $t(x) = e^{-\beta d(x)}$, which is directly related to the efficiency of the haze-free image. The β refers to the coefficient related to the particles'

scattering, whereas the $d(x)$ refers to the distance between the camera and the object. In clear climatic conditions, β is negligible, unlike the misty climate.

The dark channel was framed [7] on the idea that in a patch in a haze-free image, at least one colour channel in the RGB channel will hold intensity values that are negligible or approximately close to zero, inferring a very dark pixel. Based on the above conception, a dark channel was defined as

$$J^{\text{dark}}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} J^c y \right)$$

J^c is the intensity of the colour channel in consideration from RGB, and $\Omega(x)$ is the spot predetermined with the pixel ' x ' as its centre. The inference here is that a dark channel is nothing but the lowest value between the three colour groups is in all the pixels in the selected group of pixels as shown in Fig. 1.

The low-intensity values might be due to shadows of the objects or highly colourful objects like bright red or bright yellow or dark colours like black or dark brown.

However, in hazed images captured outdoor, the dark channels hold nowhere close to zero. This is caused due to the atmospheric light particles that cause scattering of the light particles, thus increasing the captured pixel values. This, in turn, aids in determining the haze density. Hence, based on this concept, the DCP framework for dehazing images is built as shown in Figs. 1 and 2.

The atmospheric light model and transmission map are gathered from the DCP construction. The transmission map is subject to refinement from which the dehazed image is recovered. Further, discussion on this is done in the succeeding sections about the proposed system.

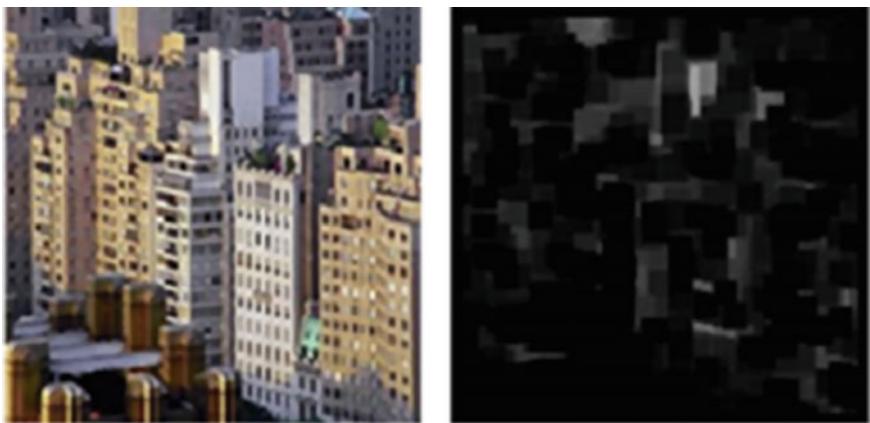


Fig. 1 Hazed image (left) and its dark channel (right)

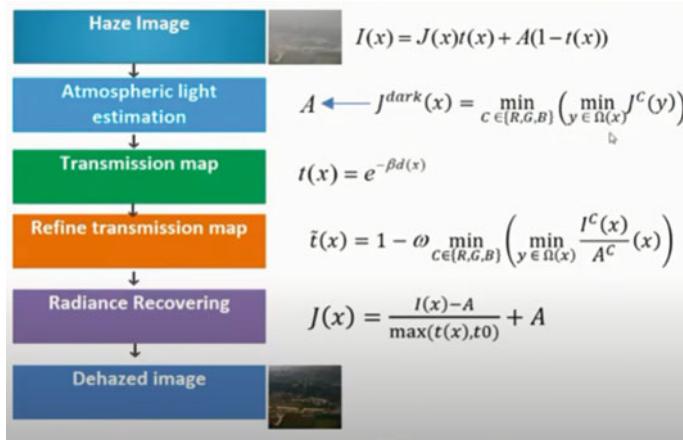


Fig. 2 DCP algorithm framework for dehazing images

Related Work and Challenges

Image dehazing can be done in three ways: (1) image enhancement or by improving the quality of the given image, (2) image fusion and (3) image reconstruction from the degraded image. Each of these methods has its algorithms suggested producing efficient results. Various works have been published with some great contributions [8, 9] under these classifications.

An application-oriented work suggested by Ji et al. [10] has employed the DCP algorithm for the incomplete haze removal and then has introduced the histogram equalization method to increase the brightness and bring out the contrast in the images. This was created to use in a traffic video surveillance system with fast acquisition of images.

Another work [11] addressed the importance of estimating the haze's depth and estimating the saliency measures of the hazed image. A saliency map is drawn to highlight the salient features, which are in turn used to refine the transmission map. This draws our attention to the concept of drawing out elements that are important and enhancing them for better visualization.

As we shift our focus on night images restoration, there are more efficient solutions that pop up. Night image restoration brings in many new challenges, which need more than just haze removal. It involves brightening of images due to overall low light conditions, low contrast and requires colour transfer [12]. Various combinations of daylight haze removal are added up like a bilateral filter in local contrast correction is used on top of the DCP framework [13] to alleviate the 'Blueshift' airtight. Another notable work [14] that caught our attention was the Retinex theory used to remove the halo effect in the transmission map. The halo effect is the distortion of the illumination between neighbouring pixels, which reduces the efficiency of the resulting image [15]. The Retinex theory is defined to remove the unwanted illumination effects from the image based on the concept that any photograph is a culmination of the

reflected image R and its illumination image L. Hence, if the illumination image can be obtained, then the reflected image can be derived. The same Retinex theory was used in the bilateral filtering [16] approach to estimate the atmospheric illumination. A variant of the normal dark channel prior-adaptive dark channel prior was introduced for a better transmission map and hence a better-dehazed image in the end.

Another perspective was because of the efficiency of the haze-free picture. The incorrect estimation of the thickness of haze in the image degraded its quality [17]. Using Laplacian-based haze thickness estimation was suggested for efficient computation and to solve colour cast problems by generating vivid colours. The colour-coded areas are first identified within the image, and then the Laplacian distribution values are computed for the RGB channels. According to the Laplacian strategy, the image with a slight colour cast will have a high Laplacian distribution score because of a lower colour variance of the picture content. Still, on the other end, an image with a heavy colour cast will possess a low Laplacian distribution score due to the considerable colour variance. Using this concept, the transmission map is refined to produce a better haze-free image.

There are few other parameters [18] that influence the visual perception of the final dehazed image. The patch size that is worked on determines the effect of transmission. Likewise, the transmission parameter also affects the naturalness of the picture. It is maintained at a constant of 0.95 and is needed to restore colours and object detection in dim contrast conditions efficiently. The results observed were that the larger the patch size, the lesser the noise in the image. The transmission parameter has produced good results in detecting distant objects when set between 0.9 and 1.

2 Project Description

2.1 Proposed System

An unconventional DCP-based visual restoration method utilizes the suggested haze thickness estimation module and the suggested image visibility restoration module. It integrates to prevail over colour cast difficulties efficaciously. In demarcation on conventional DCP built methods, the recommended method's is based on a Laplacian strategy. This strategy suggests that technique can produce more haze-free pictures than conventional DCP-based methods. The main points of the proposed method are mentioned below:

The HTE component operates to circumvent inadequate approximation of the haze thickness in actuality. The component is built on the Laplacian-based gamma rectifying method and effectively approximates the haze development width, that later on purifies the transmission map.

The width of haze is successfully deliberated after the suggested HTE component; the recommended IVR component is appealed along with the Laplacian-based White

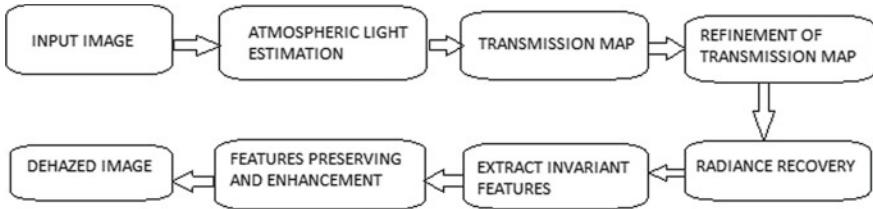


Fig. 3 Overview of the proposed system

Patch Retinex method [19], which retrieves actual colours in the scene effectively. Therefore, a haze-free picture is successfully created using this technique (Fig. 3).

2.1.1 Haze-Free Image Retrieval

Haze, an ariel occurrence that faints the lucidity of the noticed view because of the particles, namely fogs, smokes and dust particles. The hazy view distinguishes a salient thinning of colour that is conditionally corresponding to the interval of the object scenarios. The dehazing process controls parameters such as atmospheric veil inferences, smoothing, restorations and tone mappings [20, 21]. The model of haze is established by the following,

Here, $J(x)$ —Scene radiance, $T(x)$ —direct transmission, A —Atmospheric Light and $I(x)$ —Intensity of Scene Observed.

Atmospheric Light

The dark passages in the hazy picture approximate atmospheric light. The most intense 0.1% of pixels in the dark course, the highest intensity pixels are selected from the RGB plane of the haze picture as ariel light. Dark passage earlier is approximate by the minutest filter, which appeals to the picture fed in. It is grounded on the analytical idea that hazy-free pictures have a minimum of one colour passage with the fainter score. It finds the transference map and is indicated by $J_{\text{dark}} = \min(\min(I(x)))$. $\min(I(x))$ is used to find the lowest value amid every dot of RGB. The second low filter states the weakest of the local patch (Figs. 4 and 5).

Transmission Map

To recover visibility of an image, transmission map is determined from normalization of hazy image with atmospheric light. It is expressed as follows:

$$t(x) = 1 - w(\min(\min(I(x)/A)))$$

Here w is set to 0.95.

A —Atmospheric light, $I(x)$ —Hazy image intensities observer.

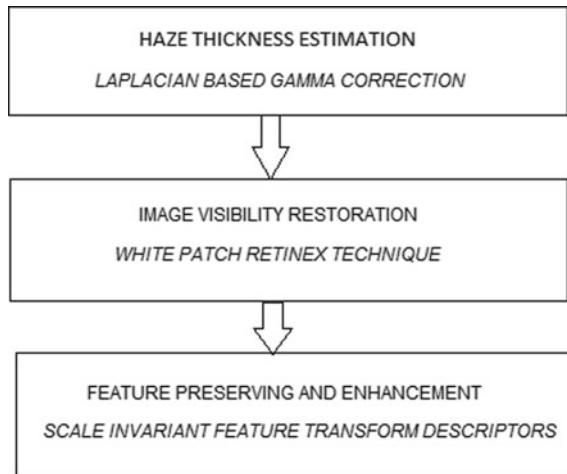


Fig. 4 Methodologies involved



Fig. 5 Hazed image and its dark channel

Inner min filter is determined as minimum value of colour channel at each point, and outer min filter finds the minimum of local patch centred at each pixel. A transmission map is further utilized to find refined transmission using detailed edge information, and it is enhanced by adaptive gamma correction to get better visibility from hazy images.

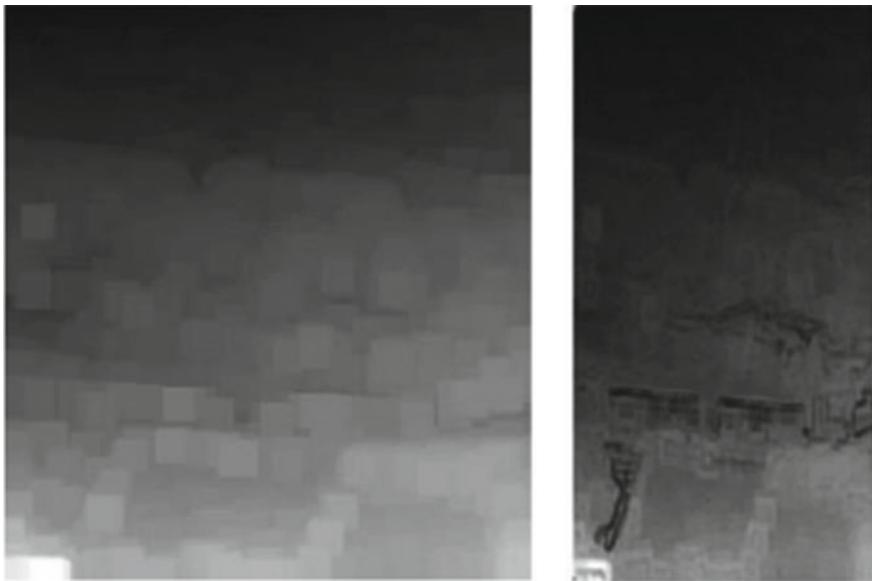


Fig. 6 Transmission map and refined transmission map

The transmission map is refined using the Laplacian-based gamma correction technique. This technique is primarily used to increase the contrast of an image. The Laplacian operator in itself increases the local contrast of an image. Gamma correction works on an image with extreme grey values [22] to improve the contrast features (Fig. 6).

Scene Recovery

The visibility of an image will be restored by estimating scene radiance from observed scene intensity values [23] using atmospheric light and an enhanced transmission map. Scene radiance will be seen as follows,

$$J(x) = ((I(x) - A) / \max(t_e(x), \text{to})) + A$$

The ‘x’—enhanced transmission map is 0.1

A—atmospheric light, $I(x)$ —hazy image intensity

After the dehazing process, the robust-invariant features are extracted using the scale-invariant feature transform. These features are utilized to match the SIFT features of hazy input images. Finally, the depth map estimation is based on dehazing yields better performance in image matching.

SIFT Descriptors

The SIFT algorithm transforms an image into a group of local vectors. The uniform characteristic taken from images can reliably match various scenes for an object.

SIFT feature removal process involves steps such as difference of Gaussian pyramid creation, extrema detection, key points elimination, orientation assignment and critical points localization. Minima or maxima of DOG values compared with local neighbourhood helps find extrema, which yields vital points.

Advantages

- It reduces the atmospheric distortion with better accuracy.
- SIFT locates invariant feature points effectively.
- Dehazing increases the local features of an image for robust matching.

2.1.2 Haze-Free Reference Retrieval

Haze-free pictures are utilized as a kind of perspective for dehazing. The image recovery component shows us that we consume a significant SIFT descriptor. It packs limited scope scale-invariant feature transform descriptors. The assembling of spatially related without haze comments are retrieved and coordinated with SIFT gatherings, positioning with the coordination. The outcome and above K recovered consequences [24] choose the connected sans haze of the information picture.

2.1.3 Haze-Free Reference Registration

Albeit the related sans haze images are exceptionally connected to dim information images with the comparable scene design and view satisfaction, looking for coordinated with pieces of these pictures straightforwardly will diminish the blending with precision. They come in various perspectives, mid-lengths and enlightenment. The distinctions are enormous, and they still have a close relationship. Hence, the scene is generally unchanged, and the principle view form is constant.

3 Result and Discussion

Based on the domain application, the Laplacian-based gamma correction is used for haze depth estimation, followed by the White Patch Retinex technique to increase the efficiency of colour retrieval. The adequacy of the strategies, diverse subjective appraisals are assessed, and the exploratory outcomes exhibit the proposed design show excellent effects for haze corrupted visuals. The usage of SIFT descriptors after the infamous dark channel prior framework has produced better results than the former methodologies (Fig. 7).

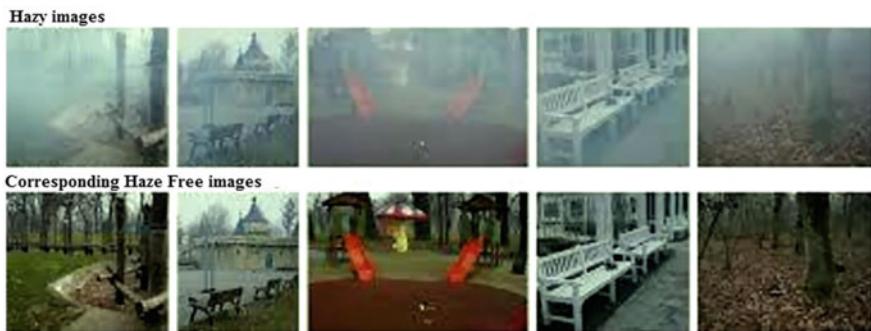


Fig. 7 Snapshots of our result

4 Gaps and Challenges

The advanced defogging calculation assumes a fundamental part in various vision applications, and the limits in the writing review are listed beneath.

1. The majority of the examined calculations have disregarded the utilization of delicate figuring methods to improve the adaptiveness of the automatic defogging expulsion calculation.
2. The majority of the paper has disregarded the issue of unpredictable light.
3. 85% of the current strategies have taken static reclamation esteem.

5 Conclusion and Future Work

In the not-so-distant future, the issue of lopsided light of computerized haze evacuation must be figured out. To improve the permeability of pictures brought about by environment suspended particles like residue, haze and mist which causes disappointment in picture handling, for example, video observation frameworks, impediment location frameworks, outside object acknowledgement frameworks and savvy transportation frameworks. Furthermore, permeability rebuilding procedures ought to be created to run under different climate conditions. Just like adding up the SIFT descriptors module for features preserving, a depth estimation technique can be used for the derived features to enrich them using converting a 2D image to 3D.

References

1. Tarel J-P, Hautière N, Cord A, Guyere D, Halmaoui H (2010) Improved visibility of road scene images under heterogeneous fog. In: Proceedings of IEEE intelligent vehicles symposium, Jun 2010, pp 478–485

2. Kermani E, Asemani D (2014) A robust adaptive algorithm of moving object detection for video surveillance. *EURASIP J Image Video Process* 2014(27):1–9
3. Ozaki M, Kakimura K, Hashimoto M, Takahashi K (2011) Laser-based pedestrian tracking in the outdoor area by many mobile bots, given at proceedings of annual conference on IEEE industrial electronics society 2011 (IECON, Melbourne, 2011), pp 197–202
4. Nayar SK, Narasimhan SG (1999) Vision in bad weather. In: Proceedings of IEEE conference on computer vision, vol 2, Sep 1999, pp 820–827
5. Narasimhan SG, Nayar SK (2000) Chromatic framework for vision in bad weather. In: Proceedings of IEEE conference on computing vision and pattern recognition (CPR), vol 1, Jun 2000, pp 598–605
6. He K, Sun J, Tang X (2011) Single image haze removal using dark channel prior. *IEEE Trans Patt Anal Mach Intell*
7. Lee S et al (2016) A review on dark channel prior based image dehazing algorithms. *EURASIP J Image Video Process* 1:1–23
8. Meng G, Wang Y, Duan J, Xiang S, Pan C (2013) Efficient image dehazing with boundary constraint and contextual regularization. In: Proceedings of the IEEE international conference on computer vision, Dec 2013, pp 617–624
9. Shwartz S, Namer E, Schechner YY (2006) Blind haze separation. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition (CPR), vol 2, Jun 2006, pp 1984–1991
10. Ji X et al (2014) Real-time enhancement of the image clarity for traffic video monitoring systems in the haze. In: 2014 7th international congress on image and signal processing. IEEE
11. Zhang L, Wang X, She C (2017) Single image haze removal based on saliency detection and dark channel prior. In: 2017 IEEE international conference on image processing (ICIP). IEEE
12. Reinhard E, Adhikhmin M, Gooch B, Shirley P (2001) Color transfer between images. *IEEE Trans Comput Graph Appl* 21(5):34–41
13. Pei S-C, Lee T-Y (2012) Nighttime haze removal using colour transfer pre-processing and dark channel prior. In: 2012 19th IEEE international conference on image processing. IEEE
14. Zhao H, Xiao C, Zhao H (2015) Night colour image enhancement via statistical law and retinex. In: Emerging trends in image processing, computer vision and pattern recognition. Morgan Kaufmann, 2015, pp 249–261
15. Rahman Z, Jobson DJ, Woodell GA (2004) Retinex processing for automatic image enhancement. *J Electron Imag* 13(1):100–110
16. Pei T et al (2019) Nighttime haze removal using bilateral filtering and adaptive dark channel prior. In: 2019 IEEE 4th international conference on image, vision and computing (ICIVC). IEEE
17. Huang S-C, Ye J-H, Chen B-H (2014) An advanced single-image visibility restoration algorithm for real-world hazy scenes. *IEEE Trans Industr Electron* 62(5):2962–2972
18. Neha, Aggarwal RK (2017) Effect of various model parameters on fog removal using dark channel prior. In: 2nd IEEE international conference on recent trends in electronics information & communication technology (RTEICT)
19. Ebner M (2007) White Patch Retinex. Colour Constancy. Wiley, Chichester, West Sussex
20. Kopf J et al (2008) Deep photo: model-based photograph enhancement and viewing. *ACM Trans Graph* 27(5):116:1–116:10
21. Nishino K, Kratz L, Lombardi S (2012) Bayesian defogging. *Int J Comput Vis* 98(3):263–278
22. Fattal R (2008) Single image dehazing. *ACM Trans Graph* 27(3):72
23. Tarel J-P, Hautière N (2009) Fast visibility restoration from a single colour or gray level image. In: Proceedings of IEEE international conference on computer vision, Sep 2009, pp 2201–2208
24. Gao R, Wang Y, Liu M, Fan X (2014) Fast algorithm for dark channel prior. *Electron Lett* 50(24):1826–1828

Comparative Analysis on the Prediction of Road Accident Severity Using Machine Learning Algorithms



Manoj Kushwaha and M. S. Abirami

Abstract The present expansion ratio of the human population is a major concern, and it is ever growing. In peak hours, traffic congestion is high and may lead to road accidents. Most of the fatal road accidents are due to unruly driving, disturbing reflections from fellow vehicles and other factors. Such incidents cause loss of human lives and properties. This paper studies multilevel models such as prediction and classification algorithms which are analyzing the severity of the accidents. Further, this will help to minimize casualties and follow established road safety measures. Prediction algorithm is used for predicting the occurrence of road accidents, and classification algorithm is used for categorizing the severity of road accidents into fatal, severe and mild injury. The researchers have used different machine learning (ML) algorithms that would analyze the severity of road accidents and other vital risk factors. ML models are also developed with the help of Internet of things (IoT) for prediction and classification. In this work, comparative study is done on the basis of performance metrics as accuracy on different ML algorithms. It is found that the random forest (RF) algorithm performs better than other algorithms. These algorithms cater high-quality results for the accident database. This study may assist the researcher, Public Works Department (PWD), and be useful to build government policies for identifying vital risk factors and minimizing road accidents.

Keywords Prediction · Classification · Severity · Machine learning · Internet of things

M. Kushwaha

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, India
e-mail: mk3342@srmist.edu.in

M. S. Abirami (✉)

Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, India
e-mail: abiramim@srmist.edu.in

1 Introduction

The world population is rising as a result people necessitate additional vehicles to reach their destination point as a result more traffic, and it may end in accidents. In a road accident, we could lose property or human lives and severe injury. Due to the growing vehicle technology and road infrastructure, the transportation of people increases in an exponential way, which may come to an end with heavy traffic and cause road accidents [12]. Road accidents are a global problem and the ninth major cause of fatality [13]. In daily life nowadays, traffic is one part of life when we are traveling, and it is a big problem for city life [10]. Nowadays, one of the global issues is road accidents. Approximately, 70% of people who die are 18–45 years old and per day around 416 die in road accidents in India. We worked on several machine learning (ML) approaches such as prediction and classification algorithms on road accidents. We analyze the prediction methods used to predict road accidents and use classification methods to categorize the severity levels such as fatal, severe and slight injury.

These ML methods were also helpful to take necessary action regarding preventive and safety measures of road accidents and also detect critical risk factors of road accidents. Predictive analysis is mostly used as a regression model which plays a significant role to reduce road accidents by forecasting future accidents and helpful to activate related organizations to take decisions and protective measurements [19]. Prediction algorithms are also used in this paper to identify critical risk factors and try to reduce severe injuries. There are some techniques used in this paper.

Machine learning algorithm, ML, allows the computers to study from data and even improve themselves not including being unambiguously programmed. ML is also used in image recognition, speech recognition, medical diagnoses, video surveillance and social media platforms. The ML algorithms are classified into three kinds as supervised, unsupervised, and reinforcement.

Supervised Learning is a ML technique that deals with the labeled training dataset learned from experiences that are already flagged with precise data [10]. It is divided into two categories such as classification and regression. Supervised learning agrees to gather and generate data output from previous experiences. It helps to resolve several types of real-world problems [22].

Unsupervised learning is the training of a machine that does not provide any teacher. It works on algorithms without any guidance, and information is neither labeled nor classified. It works with unlabeled data. It is divided into two approaches as clustering and association. Clustering algorithms are used to find natural grouping in data and Association algorithms are used to find the relationships among the data items in large data set [10].

Reinforcement is a complex ML method that deals with software agents who should take actions in an environment to get the highest rewards; q-learning is one example of reinforcement learning [10].

Artificial neural networks are motivated by biological neural networks like our brain. The neuron structure is used for ML thus called artificial learning. It helped

several problems, especially where layering is required for refinement and granular details are required.

Deep Learning algorithms perform sophisticated computations on a big amount of data. It is also used in artificial neural networks and works based on the structure and function of the human brain. It has obtained huge fame in scientific computing and is extensively used by several industries which solve complex problems.

Internet of things (IoT) is the Internet network connectivity in which physical objects are connected with technology to make communication or sense with internal or external environments. Nowadays, more than 9 billion things are connected to the Internet, and in the future, it will cross 20 billion. In other words, we can say that IoT is the collection of trillions of sensors, billions of smart systems, and millions of applications.

2 Related Work

In this section, we represented related works on the road accident severity and identified crucial factors for road accidents including ML techniques and IoT used with several approaches.

Rajkumar et al. [17] suggested ML methods to reduce the probability of road accident occurrence by analyzing the road accident and factors involved in accidents. It was found that the lighting conditions were the main key factor for road accident severity. Working on a large variety of data and the analysis of this kind of large data helps to avoid road accidents and is also helpful for the government and its people to control accidents by taking preventive and safety actions against road accidents. The multi-label classification using Logistic Regression (LR) algorithm provided better mean of accuracy than other classification algorithms.

Geyik et al. [10] suggested a model to predict the accident severity levels. The main aim of this research is to extract data from the database to compare and predict the degree of severity and also try to reduce accident injuries. Datasets classified them as fatal, serious, and slight. Different classifier models were used and different attributes: road surface conditions, road type, light conditions, and weather conditions. Experimental results conclude that the Naive Bayes (NB) algorithm gives more efficient performance compared to other algorithms. This work is useful for a researcher who worked on this type of Stasts19 datasets.

Kumeda et al. [12] presented a model to detect the important factors of road traffic accidents and concern about prevention of road accident severity. The proposed system applied the ten cross validations technique and further datasets separated 90% of training and 10% of testing data and also considered the three classes fatal, serious and slight accident classes. Based on different attributes of road accidents, Fuzzy-FARCHD algorithm applied on ML tools for prediction and classification. It used six algorithms as Fuzzy-FARCHD, radial basis function (RBF) network, random forest (RF), hierarchical learning vector quantization (LVQ), multilayer perceptron (MLP), and NB.

Ting et al. [19] suggested on detecting predictors that could effectively lead to casualty. The main objective is to detect significant factors for accident severity and to develop a predictive model. Feature selection is a difficult task with a large number of features. It is the process of choosing important variables or attributes of given datasets and further classified into three files, accident, car driver, and injury Info files. The four families for classification are RF, support vector machine (SVM), neural network (NN), XGBoost, and for the predictive model used as NB and Cart, 30 crucial features selected by recursive feature elimination (RFE) which contributed to road accident severity.

Al Mamlook et al. [2] developed a model for classifying the fatality of injuries and choosing a group of factors. Traffic accidents cause loss of lives, severe injury, property damage, and also loss of economic and social levels. In peak hour, there are so many vehicles going so there is a traffic jam and it can end as traffic accidents. The prediction classification algorithm used such as AdaBoost, LR, NB, and RF. This type of study provides important and useful information for highway engineers and transportation designers to do safer road development.

Zong et al. [23] analyzed traffic crash severity levels with the help of Information entropy and Bayesian networks. The proposed model also surveys the relationship between crash severity level and different risk factors involved in traffic crashes. It developed a severity causation network that provides effective information related to collision events. Different methods were used such as the entropy weight method and Bayesian parameter learning to achieve an effective proposed model.

Labib et al. [13] analyzed the road traffic accident in Bangladesh closely by using ML algorithms to find the intensity of the accident and also identified the significant factors. Different supervised algorithms used as decision tree (DT), K-nearest neighbors (KNN), and NB. Data classified based on injury. Data preprocessing used data cleaning feature selection and set missing value. Proposed model is helpful to reduce the number of accidents and implemented ML techniques which gave high accuracy to forecast the accident severity.

Gumasing et al. [11] proposed a work on increasing road safety in metro manila and added to awareness for motorcycle riders regarding how to do safe driving. Also, analysis work was done on the severity level of motorcycle accidents. It helped to identify the main cause of risk and severity level of motorcycle accidents. The researcher can do work on these safety model measures which can help to reduce accidents.

Mokoatle et al. [16] proposed a model as multiple linear regression (MLR) and XGBoost, where XGBoost outperformed MLR. Road traffic accidents have an important impact on the daily routine of people. By adding distance features of nearby places of interest, namely public places and crowded places, they proposed the best classification model that forecasts severe injury of drivers. They used three various scenarios, in the first scenario to predict severity with the help of distance features, the second scenario used accident report (AR) data, and the third scenario used distance features combined with AR data.

Bharadwaj et al. [3] investigated that inattention driver was the main key risk factor involved in a crash and near-crash work zone and an odds ratio found as 29.06.

In this research for the first time, they narrowly investigated the work zone events and gave important clarification to understand the different risk factors which are involved during the crashes. It used naturalistic study (NDS) data with a binary logistic model to examine and predict the crash risk in the work zone and also performed main work to identify the risk factors and safety-critical events. The first time researcher works on NDS data to answer the risk factor contribution of crashes in work zones.

Almamlook et al. [14] suggested detecting the important causality factors of elder driver's crash and also predicted the traffic crash injuries severity. For data normalization and overcoming the missing data, they used a preprocessing procedure. SMOTE methods were used for balanced training data. The proposed ML model helps in a proactive approach to reducing the high risk of traffic severity of elderly drivers. Light-GBM models gave better performance than others with an accuracy of 87.97%.

Shen et al. [18] investigated factors involving road transport accidents to help to detect the cause of accidents and reduce the casualty and improve the traffic safety measure. Hazardous materials are very harmful to humans as well as to environments. The proposed method used XGBoost as a ternary classification model for a fatal accident, injury, and property damage. For data validation of proposed models, other four popular models were applied. Results found that the XGBoost method has better performance than others.

3 Methodology

In this section, several ML algorithms applied on road accidents to predict and classify road accident severity. Some of the techniques are described below which are helpful to identify the road accident severity and risk factors.

3.1 Multilayer Perceptron

It is used for classification states and helpful in solving XOR problems by neural networks model [10]. It is one kind of feedforward artificial neural network and refers to a multilayer of the perceptron. It consists of three layers of nodes as input, hidden, and output layers. Each layer is created by units. Incoming data accepted by the first layer and forwarded to the hidden layer. The nature of the hidden layer varies according to the problem. MLP is also used in back-propagation for training and can distinguish nonlinearly separable data (Table 1).

Table 1 Background study of different methodologies related to road accident

Authors	Methodologies	Data source	Result
Ghandour et al. [8]	Combination of the bagging of J48 DT and voting SMO techniques	Lebanese Road Accidents Platform (LRAP)	Demonstrative machine learning-based model developed to detect main risk factors which are used in severe road injuries
Wang et al. [20]	Linear regression, ANOVA	Traffic Administration Bureau and National Statistics Bureau of PRC	Protection worn by victims and releasing way of impact energy are the main attributes which cause the road accident severity
Chen et al. [4]	Bayesian network (BN)	Website in china, State and Provincial Work Safety administrations	Categorize the severity level according to the priority, which is helpful to reduce the traffic accidents damage and expose critical reasoning
Choi et al. [5]	Decision tree (DT), random forest (RF), logistic regression (LR), AdaBoost	Collected from the Republic of Korea, (MOEL)	Predicted the fatal accidents and handled the system of manpower control at the construction site
Febres et al. [7]	Bayesian networks	Department direction General (DGT), Spain	Investigated that the cause of increasing the fatal and severe injury is to failure to wear the seat belt by drivers
Eboli et al. [6]	Logistic regression (LR)	The Italian National Institute of Statistics (Istat)	Suggested that road safety increased by enhancing one way roads use, urban design development and improving road intersections
Alkheder et al. [1]	Decision tree, Bayesian network, linear support vector machine	Abu Dhabi police departments	Compared to other approaches, the Bayesian network was good at predicting the variables
Mansoor et al. [15]	KNN, decision tree, adaptive boosting, support vector machine, feedforward neural network	Great Britain's Department of Transport online database	From readily available features such as crash-location got predicted crash severity with high accuracy

(continued)

Table 1 (continued)

Authors	Methodologies	Data source	Result
Xi et al. [21]	support vector machine (SVM)	Traffic Accident Research Institute of China Automotive Technology and Research Center	The model improved the recognition accuracy and reduced the computational workload
George et al. [9]	Lognormal regression	Database of Hellenic Statistical Authority	Approaches are helpful for enhancing road safety and avoidance severity

3.2 Decision Tree

One kind of algorithm is used mainly in predictive modeling, and it is also used in statistics and data mining etc. [10]. It consists of leaves and branches. Branches indicate making observations about patterns and leaves are indicated as the pattern target variable. It is a supervised algorithm used for classification problems and it contains nodes as the leaf [13]. DT methods are useful in avoiding over-fitting problems and processing datasets with missing values. It is simple and easily visualized and readable by users, able to deal with bigger datasets with large memory and high computational time [14].

3.3 Random Forest

Popular ML algorithms are used for both regression and classification [17]. It is a collection of different DT. RF consists of several DT and these trees without suffering from over-fitting and overtraining [19]. RF is a simple and classification algorithm that develops a huge outcome without hyper-parameter tuning [12]. For splitting a node, it requires a minimum number of samples to be set to two and for a leaf; it requires one [2]. To solve the problem of DT algorithms, the RF algorithm developed several trees from training datasets [14].

3.4 Naive Bayes

It is one of the suitable fast classification algorithms for a large amount of data in accident data and usage probability of Bayes theorem for prediction of unknown classes which all are independent of its attributes that classes called as Naive [17]. It comes under supervised algorithms, easy to apply, and easy to understand. The names come from Thomas Bayes and is good for ML analysis. The NB algorithm

is attractive with its elegance, simplicity, and durability. It is the oldest algorithm that is used in classification and is used widely on text and spam filtration. NB classifiers used other than the expensive iterative approach and simple model for creating classifiers.

3.5 XGBoost Classification

eXtreme gradient boosting algorithm is the best algorithm for structured data for performing classification and regression [17]. XGBoost will construct DT one at a time and be made by the previously trained tree, which helps to correct the errors [19]. It is more scalable and ten times faster in comparison to previous solutions on a single device [16].

3.6 AdaBoost

It is a boosting algorithm that effectively uses the short decision trees and training datasets containing the weight of instances [13]. AdaBoost is a simple classification method; by using a series of rounds, it invokes a given weak learner algorithm, and in this classification, algorithms include milestones of other classification such as logitboost, LP boost, and boost by majority [2].

3.7 K-Nearest Neighbor (KNN)

K-nearest neighbor is one kind of classification algorithm based on feature similarity works based on *K*-values [13]. Operations like analysis of data, measurement of distances, clustering, and similarity between the datasets can be performed by using KNN algorithm. Euclidean distance measurement can be used for the calculations related to distance.

3.8 Logistic Regression (LR) Algorithm

LR helps in the prediction of the probability of the occurrences of classes, here as probability of non-severe injury crash is (0) or severe injury crash is (1) [14]. It is also a classification and simple mathematical model and uses the linear function to sigmoid functions, which is easy to implement [2].

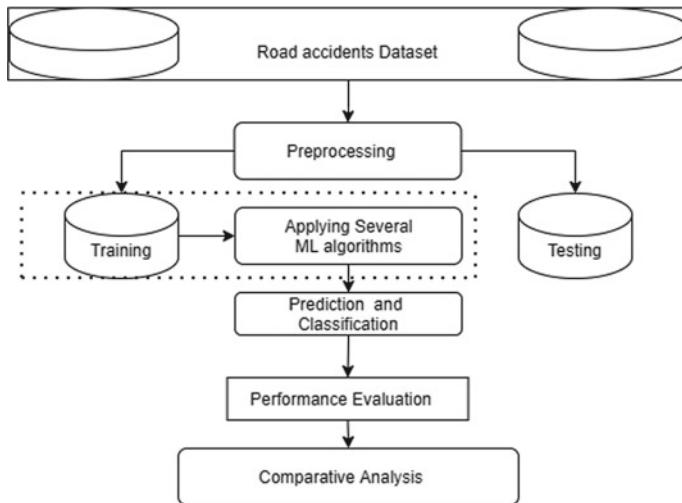


Fig. 1 Proposed methodology for road accidents prediction classification

3.9 Fuzzy-FARCHD

It is a ML classification algorithm that can give a descriptive model of a user. It is also used in data mining [12]. It is a rule-based classification that is used in data mining methods and also develops unknown knowledge from databases.

4 Proposed Methodology

This section presented the proposed methodology, which takes road accident datasets from different resources. Collecting datasets perform preprocessing operations and implement the different ML algorithms such as RF, DT, NB, LR, XGBoost, etc. on training datasets. Training datasets contain 80% of overall datasets and testing data contain 20%, and after evaluating performance, get the best result and finally get the prediction and classification result to analyze the severity levels of road accidents (Fig. 1).

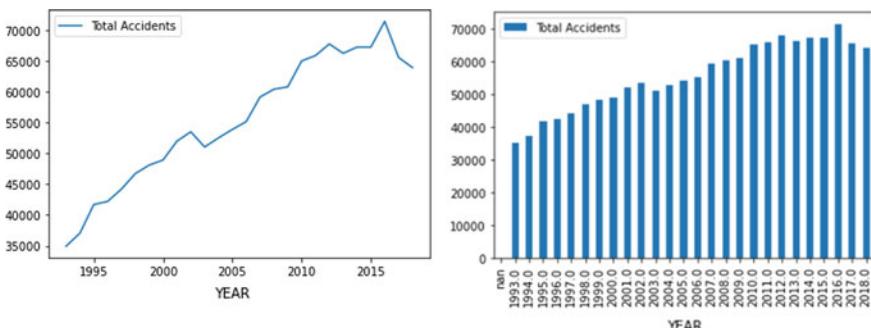
5 Comparative Analysis

In the previous section, various ML algorithms were used and compared with their performance evaluations. The systems are compared in terms of accuracy, their methodologies. There is no 100% accuracy; there is a chance for changing up and

Table 2 Comparative analysis of accuracy parameter for different machine learning algorithms

Source		Performance evaluation accuracy (%)									
	RF	NB	DT	FF	NN	SVM	MLP	CT	XB	AD	
[25]	✗	84.4	✗	✗	✗	✗	✗	✗	84.5	✗	
[14]	85.1	83.4	80.7	✗	✗	✗	✗	86.6	✗	✗	
[17]	83.4	✗	✗	85.9	✗	✗	✗	90.2	✗	✗	
[31]	95.4	52.0	✗	✗	46.8	58.8	56.7	✗	94.8	✗	
[3]	75.5	73.1	✗	✗	✗	✗	✗	✗	✗	74.5	
[20]	✗	✗	79.4	✗	✗	✗	71.4	✗	✗	75.6	

FF: Fuzzy-FARCHD, CT: CART, NN: Neural Network XB: XGBoost, AD: AdaBoost

**Fig. 2** Number of accidents happened in Tamil Nadu from 1993 to 2018

down in performance. So we compared the accuracy with the use of performance metrics. Significantly, some parameters like good weather conditions and night accidents help in increasing severity [9]. A summary of different types of methodologies, data sources, and accuracy, for road accident severity detection and avoidance is presented in Table 2. Conclude that most of them use RF, NB, and XGBoost algorithms for classification and prediction of road accident severity.

Figure 2 represents the line chart and bar chart specifying the number of accidents that happened from the year 1993–2018. It consists of the attributes such as year, fatal, grievous injury, minor injury, non-injury, total accidents, and total number of the persons involved. The graph reached a peak level in 2016 and slowly came down.

6 Conclusion and Future Work

Road accidents are serious problems for the public, and grievous accidents cause loss of valuable human lives. Different prediction and classification ML algorithms are used for road mishaps. A comparative study is done on applying different ML

algorithms for road accident datasets. Also, the vital factors involved in severe road accidents are discussed. RF, XGboost, and NB algorithms are given better results than other algorithms with respect to the accuracy parameter. Early warning systems and alert message services could be sent to medical units with the help of IoT to avoid deaths from accidents. The main aim of this study is to develop the ML-based multilevel models on the road accident database. This will facilitate accident rescue with the help of identified critical risk factors.

The study has identified critical parameters such as weather conditions, light conditions, which are typically important factors of finding accidents severity. The researchers have analyzed several datasets collected from different resources and the performance of different ML algorithms. This study discussed the severity of accidents which will assist the researchers to work in road safety measure and minimizing casualties. Developing an IoT-based ML model for predicting road accidents severity will be the future work.

References

1. Alkheder S, Alrukaibi F, Aiash A (2020) Risk analysis of traffic accidents' severities: an application of three data mining models. *ISA Trans* 106:213–220
2. Almamlook RE, Kwayu KM, Alkasisbeh MR, Frefer AA (2019) Comparison of machine learning algorithms for predicting traffic accident severity. In: 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT)
3. Bharadwaj N, Edara P, Sun C (2019) Risk factors in work zone safety events: a naturalistic driving study analysis. *Transp Res Rec J Transp Res Board* 2673(1):379–387
4. Chen H, Zhao Y, Ma X (2020) Critical factors analysis of severe traffic accidents based on Bayesian network in China. *J Adv Transp* 2020:1–14
5. Choi J, Gu B, Chin S, Lee J (2020) Machine learning predictive model based on national data for fatal accidents of construction workers. *Autom Construct* 110:102974
6. Eboli L, Forciniti C, Mazzulla G (2020) Factors influencing accident severity: an analysis by road accident type. *Transp Res Proc* 47:449–456
7. Febres JD, García-Herrero S, Herrera S, Gutiérrez JM, López-García JR, Mariscal MA (2020) Influence of seat-belt use on the severity of injury in traffic accidents. *Euro Transp Res Rev* 12(1)
8. Ghanda AJ, Hammoud H, Al-Hajj S (2020) Analyzing factors associated with fatal road crashes: a machine learning approach. *Int J Environ Res Public Health* 17(11):4111
9. George Y, Athanasios T, George P (2017) Investigation of road accident severity per vehicle type. *Transp Res Proc* 25:2076–2083
10. Geyik B, Kara M (2020) Severity prediction with machine learning methods. In: 2020 international congress on human-computer interaction, optimization and robotic applications (HORA)
11. Gumasing MJ, Magbitang RV (2020) Risk assessment model affecting the severity of motorcycle accidents in metro manila. In: 2020 IEEE 7th international conference on industrial engineering and applications (ICIEA)
12. Kumeda B, Zhang F, Zhou F, Hussain S, Almasri A, Assefa M (2019) Classification of road traffic accident data using machine learning algorithms. In: 2019 IEEE 11th international conference on communication software and networks (ICCSN)
13. Labib MF, Rifat AS, Hossain MM, Das AK, Nawrine F (2019) Road accident analysis and prediction of accident severity by using machine learning in Bangladesh. In: 2019 7th international conference on smart computing & communications (ICSCC)

14. Mamlook RE, Abdulhameed TZ, Hasan R, Al-Shaikhli HI, Mohammed I, Tabatabai S (2020) Utilizing machine learning models to predict the car crash injury severity among elderly drivers. In: 2020 IEEE international conference on electro information technology (EIT)
15. Mansoor U, Ratnout NT, Rahman SM, Assi K (2020) Crash severity prediction using two-layer ensemble machine learning model for proactive emergency management. *IEEE Access* 8:210750–210762
16. Mokoatle M, Marivate DV, Bukohwo PM (2019) Predicting road traffic accident severity using accident report data in South Africa. In: Proceedings of the 20th annual international conference on digital government research
17. Rajkumar AR, Prabhakar S, Priyadharsini MA (2020) Prediction of road accident severity using machine learning algorithm. *Int J Adv Sci Technol* 29(6):116–120
18. Shen X, Wei S (2020) Application of XGBoost for hazardous material road transport accident severity analysis. *IEEE Access* 8:206806–206819
19. Ting C, Tan NY, Hashim HH, Ho CC, Shabadin A (2020) Malaysian road accident severity: variables and predictive models. In: Lecture notes in electrical engineering computational science and technology, pp 699–708
20. Wang D, Liu Q, Ma L, Zhang Y, Cong H (2019) Road traffic accident severity analysis: a census-based study in China. *J Safety Res* 70:135–147
21. Xi J, Wang L, Ding T, Sun W (2017) Classification and recognition model for the severity of road traffic accidents. *Transp Infrastruct Syst* 1037–1044
22. Yadav P, Jung S, Singh D (2019) Machine learning based real-time vehicle data analysis for safe driving modeling. In: Proceedings of the 34th ACM/SIGAPP symposium on applied computing
23. Zong F, Chen X, Tang J, Yu P, Wu T (2019) Analyzing Traffic Crash Severity with Combination of Information Entropy and Bayesian Network. *IEEE Access* 7:63288–63302

Automatic Sickle Cell Anaemia Detection Using Image Processing Technique



V. Kiruthika , A. L. Vallikannu , and G. Vimalarani

Abstract Sickle cell anaemia is a congenital red blood cell disorder which depicts an absence of sufficient red blood cells that are fit for transporting oxygen all over body. Usually, ordinary red bloods are disc shaped. But in case of sickle cell anaemia (SCA), the red blood cell is formed in crescent shape instead of disc shapes. A novel methodology is introduced in this study to detect, classify and count the sickle cells available in the red blood cell images to start early identification and treatment. In the developed methodology, the sample images of blood containing both normal and sickle cells are collected. The pre-processing consists of grey scale image conversion and noise filtering using median filter. Watershed segmentation aids in partitioning the images distinctly so that normal cells and sickle cells can be detected. Region properties are calculated to find the number of sickle cells in the given image. Amongst the various region properties, centroid plays an important role for the detection and counting the number of sickle cells available in the image. This automated recognition will especially be useful to the medical experts as a decision support system.

Keywords Sickle cell anaemia · Median filtering · Watershed segmentation · Feature extraction · Region properties

1 Introduction

Blood is an important fluid where red blood cell (RBC) is the main factor. The structure of RBC is biconcave. The life of RBC is for 120 days. Generally, RBCs

V. Kiruthika · A. L. Vallikannu · G. Vimalarani
Hindustan Institute of Technology and Science, Chennai 603103, India
e-mail: vallikannu@hindustanuniv.ac.in

V. Kiruthika
e-mail: vkiruthika@hindustanuniv.ac.in

G. Vimalarani
e-mail: gvrani@hindustanuniv.ac.in

are circular and move compliantly through blood vessels. The presence of sickle cell anaemia, transforms the shape of RBCs into a C shape. It is a blood disorder that changes haemoglobin in the RBCs. These cells are not flexible and stick to each other. The life of sickle cells is about 10–20 days. These abnormal cells usually block or reduce the flow of blood to the arteries and carry a reduced amount of oxygen to various body parts. Globally, millions of people are affected by sickle cell anaemia. By 2050, this disorder is expected to increase by almost 30% worldwide. Several adults around the world are sufferers of sickle cell anaemia. Every year, thousands of children are born with sickle cell anaemia.

Sickle cell anaemia (SCA) is usually diagnosed in infancy through new-born screening programmes. People with sickle cell anaemia have an increased risk of serious infection, with fever to be its initial indication. The indications of sickle cell anaemia varies with respect to the individual. The main symptoms are heavy pain in the chest, abdomen, bones and joints. The other symptoms include swelling in the hands or feet, abdominal swelling, chronic contamination, difficulty in breathing, pale skin, eyesight problem and signs or symptoms of stroke. Detection of sickle cell anaemia at an early stage will help to prevent complicated problems.

Traditional methods of testing consume more time and require skilled interpretation. Manual analysis may lead to misinterpretation and false diagnosis. Interobserver variation is also possible. An ambiguity may arise if a sickle cell is predicted as normal. Moreover prompt diagnosis is a serious necessity in detection of SCA in both infants and adults. So, there is a need for an automated method for detection and analysis of sickle cell anaemia. It will facilitate easy diagnosis and prompt treatment. Hence, development of a novel algorithm for automatic detection and analysis of sickle cells is proposed in this study.

2 Literature Survey

Hajara et al. [1] performed Otsu thresholding method and it was used for segmentation, followed by the feature extraction. The drawback of this method is detection, and classification of normal and sickle cells was not achieved accurately as thresholding method was used. Extraction of proper features was not carried out. Chy and Rahaman [2] carried out threshold based segmentation for detection, followed by morphological operations. The drawback of this algorithm was that it was highly sensitive to noise, and the method did not achieve the desired accuracy. Mohamad et al. [3] performed Laplacian of Gaussian (LoG) edge detection algorithm to detect sickle cells diseases at the early stage. The drawback of this technique was that it was very sensitive to noise and the edges were eventually degraded.

Bhagavathi and Niba [4] performed an automated method to predict the number of red and white blood cells using fuzzy logic and the drawback of this method was that it did not scale well to large or complex problems. Fuzzy logic is not always accurate, and the results are supposed as a deduction and so it may not be as widely trusted. Sahu et al. [5] used fractals for predicting the shape of sickle cells.

Edge-based segmentation and clustering techniques were used for segmentation. The drawbacks of this technique were that the edge detection algorithm did not work well for images with lot of edges. At certain instances, it predicted incorrect edges and ended up in computational complexity. Likewise, clustering technique is sensitive to noise and highly relies on choosing an appropriate number of the cluster. Athira and Ashok [6] performed Circular Hough transform for feature extraction. The drawback of circular Hough transform is that it requires numerous calculations. Its tolerance towards noise is less and reduces the ability of computation. Manuel Gonzalez-Hidalgo et al. developed a method using ellipse adjustment for separating the red blood cluster from the images [7]. Circular hough transform was used for feature extraction [8]. Fractal dimension was used for sickle cell detection [9, 10]. Mean radius was used for computing the shape of the cell [11].

3 Methodology

The developed system uses image processing techniques to detect and to count the numbers of sickle shaped RBC cells. The blood smear containing the sickle shaped RBC's are segmented, followed by the application of region properties on the segmented image to count the number of sickle cells. Figure 1 illustrates the flow process.

The input image comprises the images of the sickle cells and normal cells. It is collected in png or jpeg format and are available in RGB colour space.

a. Pre-processing

Conversion of RGB image to Grey scale. The images are transformed to grey scale. Various shades of grey are available here. For grey images, the three primary colour components are of the same intensity.

Image Denoising. Image denoising is carried out using a median filter. It is nonlinear in which each output sample is computed as the median value of the input

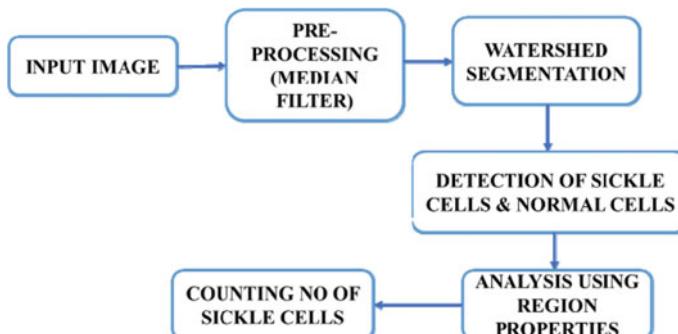


Fig. 1 Flow diagram of the proposed system

samples under the window. This filtering procedure is accomplished by descending a window upon the image.

In this study, it helped to enhance the image quality so that the normal and sickle cells were clearly differentiated. It preserved the sharp edges of the sickle cells and important information in the image. It removed only the noise without disturbing the edges of the normal and sickle cells.

Conversion of Grey scale image to binary image. Greyscale images are transformed into binary images with the thresholding technique. A particular threshold is selected. The grey level values which fall below the selected threshold is set to category 0(background) and the grey level values which fall above the selected threshold are set to category 1(foreground).

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The threshold pixel at (x, y) is indicated by $g(x, y)$ and greyscale pixel at (x, y) is indicated by $f(x, y)$.

b. Watershed Segmentation

Segmentation using watershed transform is applied on an image to classify or distinguish foreground objects and positions in the background. It is widely used in the medical image processing. Images are perceived to be a geological landscape, and the boundaries that segregate the regions of the image are determined by watershed lines.

It helps to detect unvarying objects from a homogeneous background. Watershed transform can differentiate the ridges or watershed boundaries amid the local minima adjacent to it. The local minima regions are diluted and all the watershed boundaries are united. The united regions will possess distinct identity which is definite to the initially selected seed points.

In this study, it clearly helped to differentiate the sickle and the normal cells from its background. It also helped to predict the exact shape and contour of the normal and sickle cells which were incredibly useful in differentiating both.

c. Region Properties

Region properties measure a variety of image quantities and features in an image. It automatically determines the properties of each contiguous region. There are different region properties such as area, convex area, major axis length, bounding box, minor axis length, eccentricity, centroid, perimeter, extrema and extent.

In this study, all the region properties were computed for analysis of sickle cells and centroid property helped to analyse the normal and sickle cells clearly. Centroid, also called as the centre of mass works such that for each object that is seen in the image, it calculates the centre of mass for that object. This is the (x, y) locations of where the middle of each object is located. In this study, the centroid for each normal and sickle cell available in the image is computed so that they could be differentiated accurately. Centroid helped in counting the number of sickle cells.

4 Results

The proposed novel sickle cell anaemia detection is implemented in MATLAB 7.12. The images are taken from the website sicklecellanaemia.org which contains image datasets. Fifty images with different grading was used for the study. However, for presentation in the thesis, three images are represented. Figure 2 shows the sample image database.

The given input image is in RGB colour space, and it is transformed into grey scale image for reducing the computational complexity. The presence of noise in the input image degraded quality of recognition and created ambiguity in detection of normal cell and sickle cells. So, there was a necessity to remove the noise in the given input images. Image denoising is carried out using the median filter, and the filtered image is given in Fig. 3.

For segmentation to be better and precise, the greyscale image is again converted to binary image so that the segmentation of sickle cells could be more effective. All the pixels available in the greyscale image are replaced for transforming it into a binary image. This is because watershed segmentation worked better on a binary image. Here, the white regions are assigned a value of 1 and the black regions are assigned a value of zero. The resultant binary image is shown in Fig. 4.

Region of interest is fixed using regionfill. Regionfill helps in efficient smooth interpolation that takes place from the boundary pixel values towards the inward

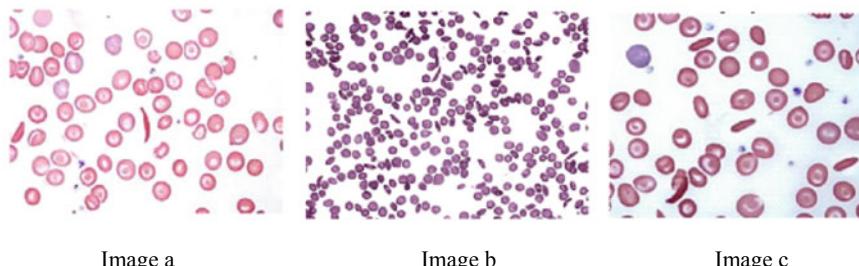


Fig. 2 Original images

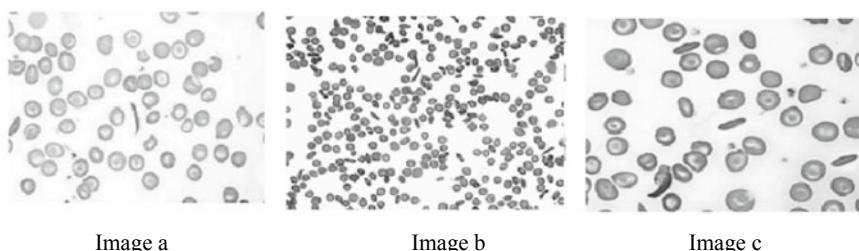


Fig. 3 Denoised images

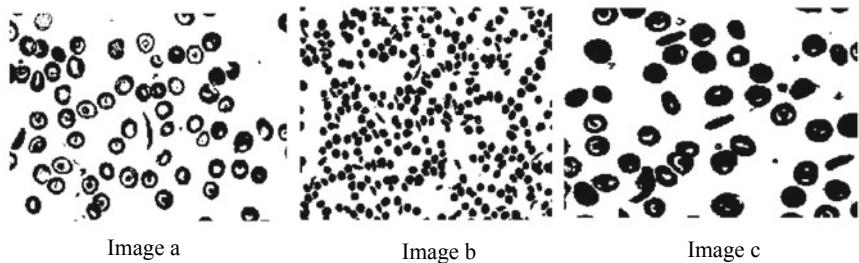


Fig. 4 Binary images

regions. It is also used for removal of extraneous and unwanted details in the image. Region filling process substitutes value in the specific area with the value which blends with the region in the background. It creates a mask image in order to indicate the region of interest which requires filling. Filled image is shown in Fig. 5.

Watershed segmentation is performed on the filled image, and all the cells are marked clearly with exact borders.

During watershed segmentation the unwanted small cells are not detected. This helps in finding the sickle cells very clearly. Labelling of the objects is done in order to find the region properties of the segmented image. Labelled image is show in Fig. 6.

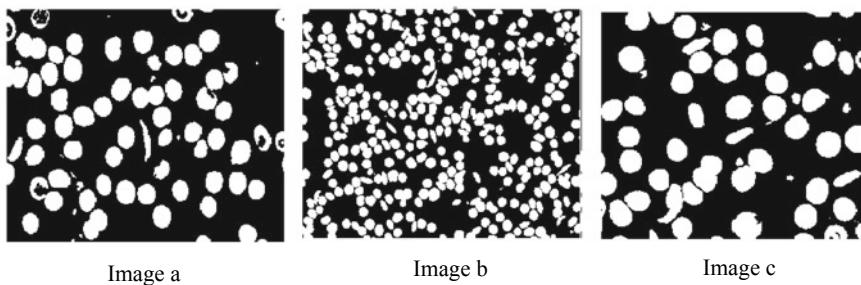


Fig. 5 Filled images

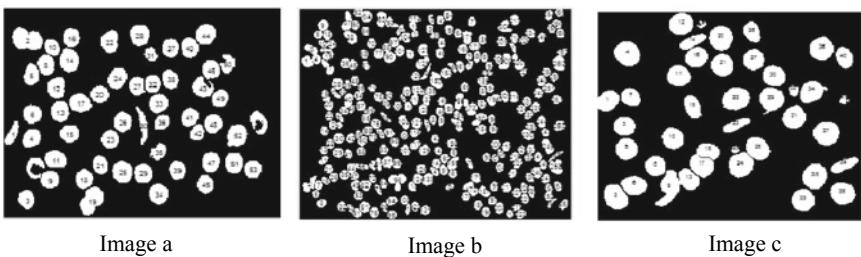


Fig.6 Labelled images

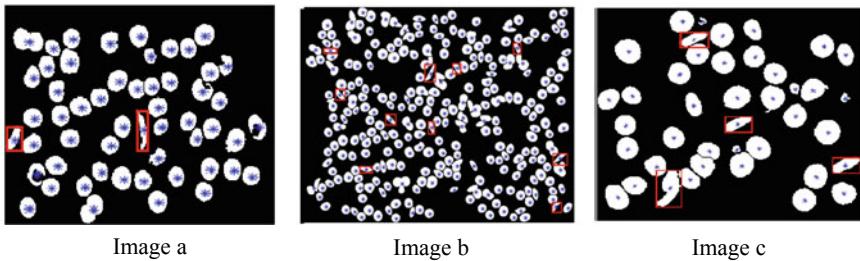


Fig. 7 Detection of sickle cells

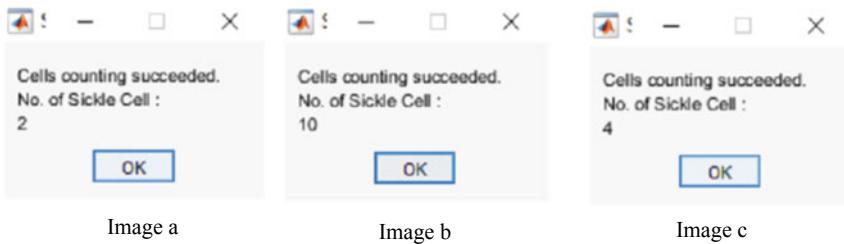


Fig. 8 Count of sickle cells

Various region properties such as area, centroid, convex area, Minor Axis Length, eccentricity and Major Axis Length are calculated to find number of sickle cells. Centroid played an important role in finding sickle cells. It helped in differentiating the normal cells and sickle cells. Sickle cells were clearly detected and marked in red colour with a box in red using the region properties as shown in Fig. 7.

Finally, followed by the segmentation and classification of the sickle cells in the given image, the quantity of sickle cells was also predicted in the resultant segmented image.

The numbers of sickle cells were counted and displayed using a box as shown in Fig. 8. This facilitates easy analysis of the input image of the patient. It also helps to know the size, shape, location of the sickle cells in the given image. It aids the medical expert enable quick medication to the patient to reduce severity of disease.

5 Conclusion

In sickle cell anaemia, the most difficult task is accurately detecting RBC sickle cells within the shortest possible period. An automated algorithm was developed in this study to detect the sickle cells and count the number of sickle cells in the given image. Pre-processing was done to enhance better segmentation. Segmentation was done using watershed algorithm and the region properties were used to identify

the number of sickle cells. The proposed method has rendered promising results by giving the accurate count of sickle cells in the given image. This methodology will be of great support to the medical experts who deal with sickle cell anaemia disorders. It helps them to take a decision regarding the presence of sickle cells and give appropriate treatment to the patients accordingly. It will also facilitate easy and prompt diagnosis.

6 Future Work

This work shows the progress in the detection of sickle cells and had more accurate results, but there is a dire need to employ some classification methodologies using machine learning to enable automatic classification of normal and sickle cells.

References

1. Hajara A, Abdul RM, Sudirman R (2019) Segmentation and detection of sickle cell red blood image. In: AIP conference proceedings. 2173, 020004, 1-10
2. Chy TS, Rahaman MA (2018) Automatic sickle cell anemia detection. In: International conference on advancement in electrical and electronic engineering, IEEE, Gazipur, Bangladesh, pp 1–4
3. Mohamad AS, Halim NSA, Nordin MN, Hamzah R, Sathar J (2018) Automated detection of human RBC in diagnosing sickle cell with Laplacian of Gaussian filter. In: IEEE conference on systems, process and control, IEEE, Malaysia, pp 214–217
4. Bhagavathi SL, Niba ST (2016) An automatic system for detecting and counting RBC and WBC using fuzzy logic. ARPN J Eng Appl Sci 11(11):6891–6894
5. Sahu M, Biswas AK, Uma K (2015) Detection of sickle cell anemia in red blood cell: a review. Int J Eng Appl Sci 2(3):45–48
6. Athira S, Ashok B (2014) Identification of sickle cells from microscopic blood smear image using image processing. Int J Emerg Trends Sci Technol 1(5):783–787
7. Gonzalez-Hidalgo M, Guerrero-Pena FA, Herold-Garcia S, Jaume-i-Capó A, Marrero-Fernández PD (2015) Red blood cell cluster separation from digital images for use in sickle cell disease. IEEE J Biomed Health Inf 19(4):1514–1525
8. Hegde RB, Keerthana P, Harishchandra H, Sandhya I (2018) Peripheral blood smear analysis using image processing approach for diagnostic purposes: a review. Biocybern Biomed Eng 38(3):467–480
9. Philip B, Quirk SM, Joseph P (2015) Object characterization in grey scale imagery using fractal dimension. U.S. Army Research laboratory
10. Al-Saidi NMG, Mohammed AJ, Ahmed AM (2014) Fuzzy fractal dimension based on escape time algorithm. Appl Math Sci 8(3):117–129
11. Aruna NS, Hariharan S (2014) Edge detection of sickle cells in red blood cells. Int J Comput Sci Inf Technol 5(3):4140–4144

Recognition and Counting of an Object Using Yolo and CNN



R. Vallikannu , V. Kiruthika , and M. Kaviya

Abstract Object identification and classification are vital in automation domains. It has gained conscience in the recent days owing to the growing domain of image processing and automation based of the growth of multiple sectors implementing automations. Object detection, identification and counting mechanisms are gaining momentum. Random urgencies and the trending algorithms of machine learning and deep learning strategies have made the object detection domain even more vulnerable. This paper implements a strategy to detect, classify and count objects based on image classification machine learning algorithms using Yolo. The proposed work considered training data collection using the app dataset, pre-processing using the DWT and GLCM, feature extraction using supervised ML algorithm, prune the available data and classification based on the proposed prediction models using R-CNN algorithms. The proposed methodology provided better accuracy with less computational overheads than the existing methods.

Keywords Convolution neural network (CNN) · Computer vision (CV) · You look only once (YOLOv3) · Object detection · Object counting · Python

1 Introduction

Object detection is much the need of the time. Emergence of new economic situations, emerging new markets, unearthing new opportunities in each domain such as Retail, E-Commerce, Media, Automobile and Transportation, IT Healthcare, Imaging concepts all require Imaging detection, classification and counting mechanisms to a high rate of accuracy. Figure 1 provides an illustrative view of the prevailing and the predictive market share for the object recognition markets.

Object detection is primarily used worldwide to allow clients and companies to perform advanced computational strategies in updated domains. Applications from Google Snapchat are some randomly used object detection strategical companies.

R. Vallikannu () · V. Kiruthika · M. Kaviya
Department of ECE, Hindustan Institute of Technology and Science, Chennai 603103, India
e-mail: vallikannu@hindustanuniv.ac.in

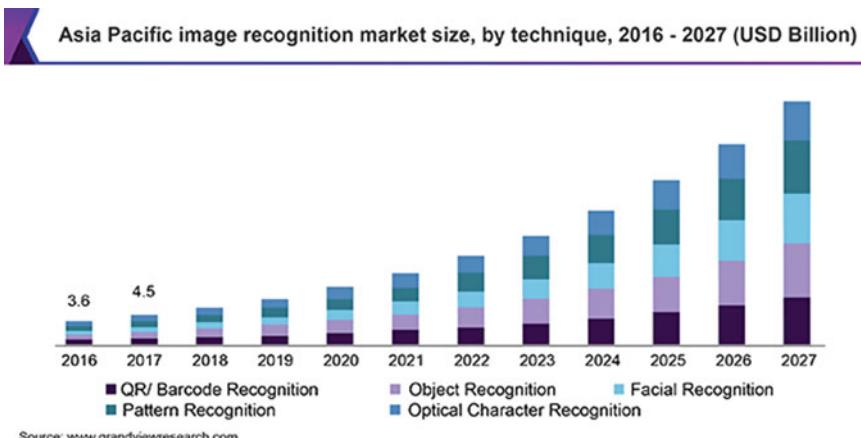


Fig. 1 Imaging market size (2016–2027)

Object detection and object counting are two trajectories of a same implementation. Datasets, termed as collection of images, are preferred and applied with specific algorithms to get the desired outputs. In addition, object detection domain have also made itself the predominant wellspring of centralized dataset for multiple applications.

While impressive examination has been deployed concerning object detection and classification investigation, restricted research has been led on foreseeing the same and applying latest technologies on the same. The results are not being fed and stored to the datasets for further use.

Thus, a strategy deploying latest machine learning or neural networks and a parallel detection algorithm in depicting the objects taking into account the fluctuations and illuminations of the environment, considering the latest datasets would be recommendable.

This article implemented Yolo, CNN Algorithms to detect, classify and count objects. The main assumption, in this paper in terms of counting objects and detection, is from an industry perception. This paper deployed convolutional neural network and YOLO for detection and supervised machine learning algorithms for feature extraction.

This article implemented the concept of applying feature extraction to the datasets and comparing them to the existing datasets for analysis and detection. Overcoming the flaws in the existing models of detection analysis and determining the highest rate of accuracy in the detection and classifications are mandatory and is the need of the hour.

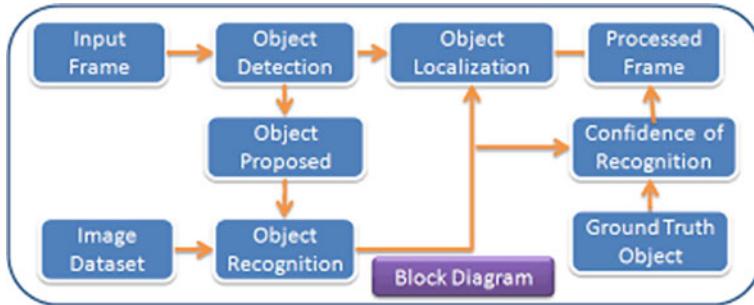


Fig. 2 Existing models object detection

2 Literature Survey

Many scholarly works have been done on object recognition specifically. But there are some restrictions on the same. Most previous studies use only the primary component analysis, i.e., only the limited sample data is considered for comparison, and the same is maintained as the reference filter value throughout detection process [1–4]. In addition, during the acquisition process, only classification techniques are used, and a paradigm is made. This may not produce the desired results, but accurate predicting of an object will not be close to it, when the same is used in real-time situations. Mostly, K-Means clustering is used for predictive purposes. There are many other algorithms available. Therefore, it is necessary to conduct a detailed study using the same and to make comparisons at the level of accuracy of the detection of stated implementation [5].

Some scientific works are implemented using MATLAB. Statically, the measures of this detection system, which is processed in terms of object detection, algorithmic processing of the frame data is to be considered during the time of pre-processing the dataset before applying to the classifier. Accuracy needs to be designed and implemented in a much more sophisticated model to achieve the desire detection accuracy levels [6–9]. Comprehensive analysis over a certain selected dataset only is possible in the existing systems (Fig. 2).

3 Methodology

Convolutional neural network approaches, specifically Discrete Wavelet Transform (DWT) and GLCM for pre-processing and feature extraction and R-CNN for classification and comparison of the frames with the existing datasets is implemented in this work. Regional convolutional networks are a perfect tool for detection. First, dataset of certain objects as training set is considered and with the help of Discrete Wavelet Transform and GLCM, feature of the datasets and the comparable data were

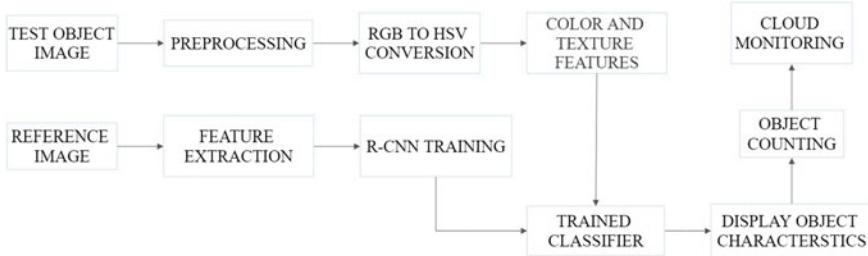


Fig. 3 Process flow diagram of the proposed method

extracted. The input data is either fetched real time with the camera or by any other source and is stored in a folder and pre-processed. A novel implementation scheme to collect the app dataset, process them using the DWT and SVM algorithm, prune the available data and classify them based on the proposed prediction models are implemented in this paper. Improvised percentage of accuracy in object detection and classification implementation of R-CNN algorithms, and graphical results are the major advantages of the proposed system (Fig. 3).

3.1 Data Collection and Training

Data streams values may be frames from the images, such as array of pixels or real images. Data is pre-processed using the desired algorithms. The test image is then converted to HSV. While the input data might be of RGB, this color conversion is mandatory to analyze the images on real time. This HSV format is as perceived by the human eye.

The steps used are pre-processing. Desired feature extraction for both the input image and dataset images. It is to increase the level of accuracy of the prediction. Figure 4 represents the flow diagram of dataset training using YOLO.

3.2 Pre-processing and Extraction

Data pre-processing the initial step toward the detection concepts. The datasets are pre-processed using the DWT Algorithm and GLCM. Feature extraction is required to decrease the number of highlights in a dataset by making new features from the current ones. This is done by using the discrete wavelet transform. This DWT algorithm helps to utilize the coefficient of the datasets to generate the same to a numerical precision (Fig. 5).

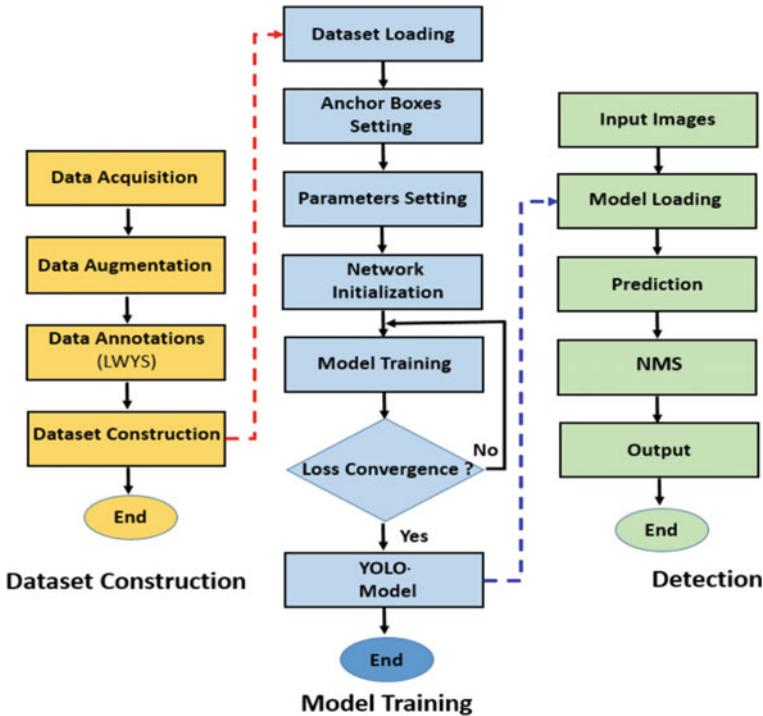


Fig. 4 Flow diagram of dataset training using YOLO

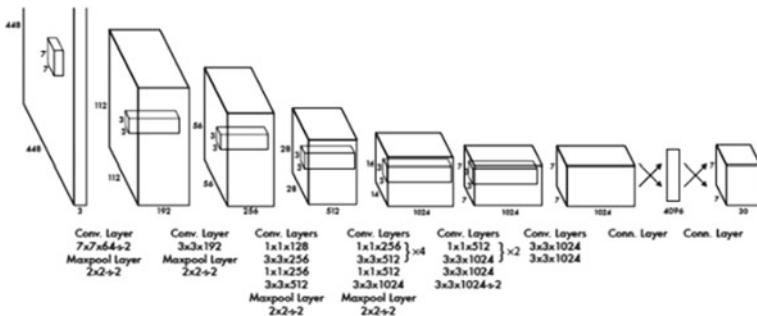


Fig. 5 Layer wise segmentation of images using R-CNN

3.3 Detection and Classification

Detection process flow is the process wherein the considered dataset is processed, suitable algorithms applied and detection AL values determined. We deploy R-CNN and Yolo Detection in this paper to determine the CV images. A R-CNN is the updated version of the Decision tree and convolutional neural networks, such as it

forms a cluster of Randomized DT's, while the data is processed for utmost accuracy and minimal loss.

3.4 Classification and Counting

The results obtained from the input data image and the pre-processed existing datasets and compared using the Neural Network Trainer. The compared results can be obtained and counted using simple python program. Multiple sets of data input image been checked with the samples of real-time data and the detections shows higher percentage of accuracy, which is very high comparatively to the existing systems available.

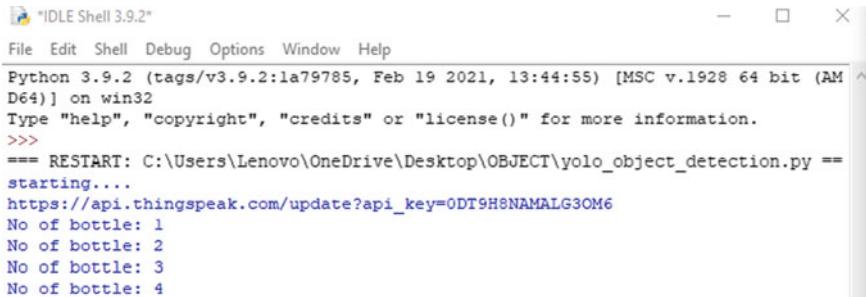
The results are stored into the database for further analysis purpose.

4 Results

The proposed system presented the prediction of object detection to a precise accuracy level. More number of datasets make prediction algorithms more accurate. With very less computational efforts the optimum results were obtained, which also shows the efficiency of proposed algorithm in recognition and object prediction. Predicted objects are counted and stored in cloud. Figure 6 presents the detection of water bottles considered as a sample object, and Fig. 7 shows the number of bottles detected.



Fig. 6 Results detecting bottle *Images*



```
*IDLE Shell 3.9.2*
File Edit Shell Debug Options Window Help
Python 3.9.2 (tags/v3.9.2:1a79785, Feb 19 2021, 13:44:55) [MSC v.1928 64 bit (AM
D64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
== RESTART: C:\Users\Lenovo\OneDrive\Desktop\OBJECT\yolo_object_detection.py ==
starting...
https://api.thingspeak.com/update?api_key=0DT9H8NAMALG3OM6
No of bottle: 1
No of bottle: 2
No of bottle: 3
No of bottle: 4
```

Fig. 7 Display of object (bottle) *counting*

5 Conclusion

The most difficult task in automated object detection and counting is accurately detecting the boundary of the object chosen. An automated algorithm was developed in this study to detect the objects and count the number of objects in the defined area considered. Pre-processing was done to enhance better segmentation. CNN and YOLO algorithm were used to identify the number of objects precisely and accurately. The proposed method has rendered promising results by giving the accurate count of the chosen object. This methodology will be of great support to the industries who deal with automation process. Such automated method will also facilitate easy and prompt production process.

6 Future Work

The accuracy level of the implemented work is very appreciable; however, real-time scenarios can be worse. It is always better to make use of the latest algorithms and check the level of accuracy for each stage of the paper implementation. Artificial Neural Network, Fuzzy Logic and hybrid algorithms can also be deployed to check the same.

The proposed work could be expanded with latest algorithms to provide optimum results with respect to existing techniques. Real-time application-based image categorization will be one of the main factors in the selection of the technique. Identifying multiple objects is a sensitive and necessary task, so reliability and accuracy will also play a major role in the selection of the method. Detection of small objects that appear in groups is difficult. The proposed algorithm finds difficult to generalize objects in new or unusual aspect ratios as the model learns to predict bounding boxes from data itself. So, any improvised method could be used for future method.

References

1. Alexe B, Deselaers T, Ferrari V (2010) What is an object? In: 2010 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, San Francisco, CA, pp 73–80. <https://doi.org/10.1109/CVPR.2010.5540226>
2. Druzhkov PN, Erukhimov VL, Zolotykh NY, Kozinov EA, Kustikova VD, Meerov IB, Polovinkin AN (2011) New object detection features in the OpenCV library. Pattern Recognit Image Anal 21(3):384
3. Alahi A, Ortiz R, Vandergheynst P (2012) 2012 IEEE conference on FREAK: fast retina keypoint, computer vision and pattern recognition (CVPR), pp 510–517
4. Han S, Pool J, Tran J, Dally W (2015) Learning both weights and connections for efficient neural network. Adv Neural Inf Process Syst 1135–1143
5. Dai J, Li Y, He K, Sun J (2016) R-fcn: object detection via region-based fully convolutional networks. In: NIPS
6. Redmon J, Divvala S, Girshick R et al (2016) You only look once: unified, real-time object detection. In: IEEE conference on computer vision and pattern recognition. IEEE Computer Society, pp 779–788
7. Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. Comput Vis Pattern Recogn (cs.CV). [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)
8. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2014) Object detection with discriminatively trained part-based models. TPAMI 3:5
9. Oquab M, Bottou L, Laptev I, Sivic J (2014) Learning and transferring mid-level image representations using convolutional neural networks. In: CVPR

A Novel Solution for Carrier Frequency Offset (CFO) Minimization for Efficient 5G Wireless Communication Using OFDM for Internet of Things (IoT)



G. Elavel Visuvanathan and T. Jaya

Abstract In the telecommunications technology, long-term evolution (LTE) will shortly be replaced by 5G wireless connectivity with improved latency, speed and power efficiency. The Internet of things (IoT) is one of the latest developments in 5G technology. IoT ensures low capacity, low cost and loose narrowband synchronization. For easy synchronization of wireless technologies, the cyclic prefix (CP) technique needs to be improved because it is not appropriate for 5G scenarios. When implemented in the OFDM waveform, a separate cyclic prefix (CP) is needed for each symbol. It is not practical for 5G application on narrowband. In this paper, a novel scheme of carrier frequency offset (CFO) estimation technique using cyclic prefix adaptive synchronization (CPAS) algorithm in non-orthogonal multiple access (NOMA) system is proposed.

Keywords Frequency · Carrier · 5G · Wireless communications · Orthogonal

1 Introduction

In this paper, we present the idea of NOMA for both spectral efficiency and down-link networks, where it allows the multiplexing of the power domain operator and exploiting the disparity between channels in mobile systems, both characteristics that were not used in current and previous cell systems [1]. We would also clarify interface

The original version of this chapter was revised: The authors affiliation has been updated. The correction to this chapter is available at
https://doi.org/10.1007/978-981-16-8721-1_74

G. Elavel Visuvanathan (✉) · T. Jaya

Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India

e-mail: elavelvg@srmist.edu.in

T. Jaya

e-mail: jaya.se@velsuniv.ac.in

G. Elavel Visuvanathan

SRM Institute of Science and Technology, Chennai, India

architecture considerations for NOMA and its combination with MIMO, including multi-user scheduling, multi-user power management and multiple-input, multiple-output (MIMO). The data rate and uplink NOMA output assessment is evaluated. For data transmission NOMA, we demonstrate the success of connecting NOMA with OFDMA's device level in conjunction with MIMO. We provide a system-level comparison of SC-FDMA and NOMA with the effect of marginal frequency reuse for uplink NOMA (FFR). In addition, we are introducing our experimental NOMA downlink test bed.

For example, for downlink NOMA, NOMA's multiplexed users' signals differ from the same transmitter, and thus, there is no problem related to synchronization, and signal overall can be reduced, since other user-multiplexed information pertaining to the demodulation or encoding of a specific user can be transmitted jointly with that user's information [2]. Downlink NOMA at 3GPP as an LTE release analysis under the title 'Downlink Super-Position Transmission' is currently addressed at 3GPP. One critical feature of cellular systems architecture is the development of multiple access schemes. It helps to ensure the spectrum-efficient and economic sharing of radio services with different users (Fig. 1).

Wireless networking at very high data rates. The users will be able to access higher data rates; making it possible to download high-definition images, which involves improved modulation with high speed and energy, is the standard application corresponding to that scenario. The Web of things (IoT). The 5G network is supposed to link up to a trillion items, allowing users to remotely access products like vehicles, washing machines, air conditioners, lights, etc. Similarly, solar, water and gas suppliers can use the linked smart meters to monitor their networks [3]. The connecting objects would have very low processing capacity and have to sporadically transfer a small volume of data, needing robust modulation to synchronize time mistakes and good performance for short communications. This scene applies

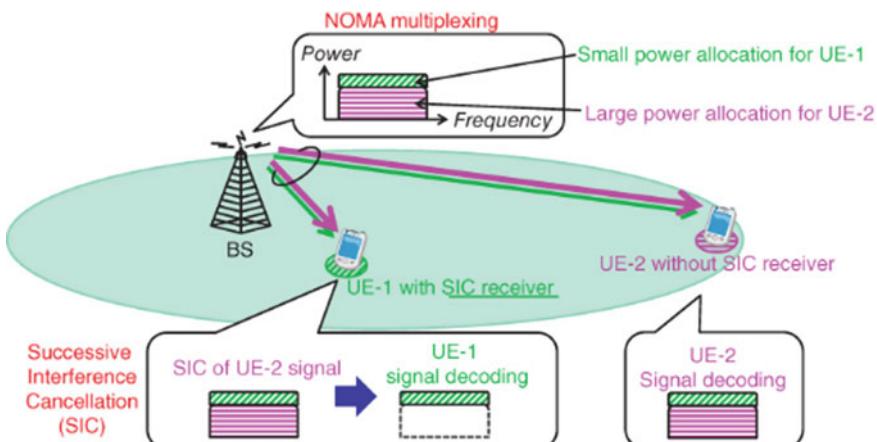


Fig. 1 NOMA multiplexing

to cyber-physical experiments in real time (for example, drones or rescue robots are remotely controlled in emergencies), which needs secure and limited latency networking services. The latency target is more than 1 m in order, and the latency of the new 4G networks is less than one order in magnitude.

2 Literature Review

Martinez (2019): The NOMA scheme has gained some significant credence as a good candidate for potential networks. The technique of frequency division (which works by separating frequencies using a NOM while in-derived subcarriers) is used in order to enable the orthogonal frequency division multiplex (OFDM) system to operate at a higher efficiency. Introducing OFDM (with respect to NOMA, offset for each carrier) errors results in the introduction of CFO errors into the overall system [4]. The topic of this paper is to study and quantify the effect of the CFO on the NOMA's operating frequency expansion on output in this study. In order to increase the efficiency of the CFO-stricken NOM-OFDM, an iterative weighted least square (WLS) algorithm is used to improve the solution process of solution of the least squares approximation method to iterate to WOLS. Using the traditional least squares (LS) as the starting point, the LS approach first is used to find initial figures of CFO costs and expenses. Now, after which, the Markov model is used to predict how long it will take to wait before a segment can be expanded to have maximum entropy, as a measurement is made. Parity game simulation findings are shown to demonstrate the influence of the NOM-OFDM error on NOMPL error operation In addition, the algorithm's output is evaluated and the algorithm's performance is presented and assessed [5].

3 Proposed Methodology

The fifth generation (5G) is driven by the rapid increases in wireless capacity requirements imposed by advanced multimedia applications with the significantly growing demand for user access needed for IoT. No orthogonal multiple access (NOMA) is proposed as a promising technology in 5G networks to solve the challenges described above. Via traditional orthogonal multiple access (OMA) techniques, substantial bandwidth efficiency enhancement can be achieved [6]. We offer a state-of-the-art NOMA in this project, focusing on the concepts of NOMA and evaluating different parameters. Cyclic prefix adaptive synchronization (CPAS) algorithm in non-orthogonal multiple access (NOMA). The presentation of a NOMA comprises variants using a cyclic prefix (CP) as a guard interval technique; windows or filters to increase the spectral decay rate of the signal transmitted; the recipients dependent both on windows and filter are included [7]. There are sufficient conditions to prevent inter-symbol and intra-symbolic intrusion on the channel impulse response and any Windows/Filtering (Fig. 2).

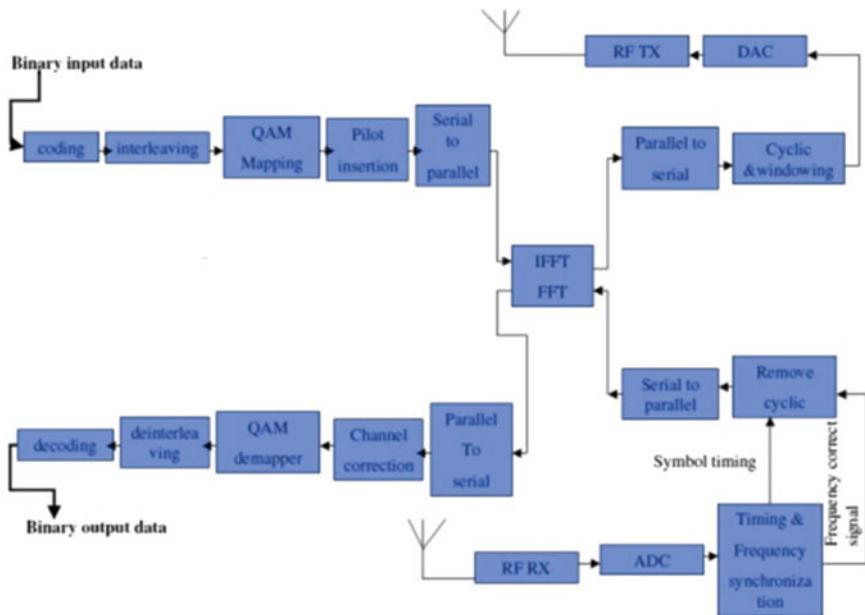


Fig. 2 QAM architecture design

The PHY latency of upcoming 5G networks could not surpass 200–300 c01-math-0001s in order to meet such an ambitious target. While not specifically linked to the tactile Internet model, other applications such as car-to-car and automobile communications will benefit from the low-latency specifications. In the past, wireless cellular networks have evolved because of the need to improve their efficiency [8]. In reality, greater data speeds became the key driver along the way from 2G systems1 to 4G networks, with data rates from tens of Kbit/s updated to the current tens of Mbit/s. The transformation from Gaussian minimum change keying, which is in use in the 2G GSM system, to quadrature amplitude modulation (QAM), which has adaptively selected cardinality schemes which are currently in 4G systems was concentrated on a physical layer (PHY). In contrast to previous wireless network iterations, 5G platforms will have to support many services and new applications and will not simply rely on increasing data performance (Fig. 3).

Uplink power management in two aspects, NOMA differs from the downlink. First of all, there is a different transmission capacity limit for optimization. The transmission capacity on the downlink is limited by just one limit: maximum transmission power of the BS, maximum transmission power of each EU is restricted in the uplink transmission power optimization. Second, there is a distinct TPC architecture approach. On the bottom link, the overlay signal received by an EU has the same channel: each EU recipient has the same channel gain for signals of different UEs. The architecture of downlink power control, therefore, seeks to artificially differentiate the signal of various UEs in the field to enable signal separation at the receiver,

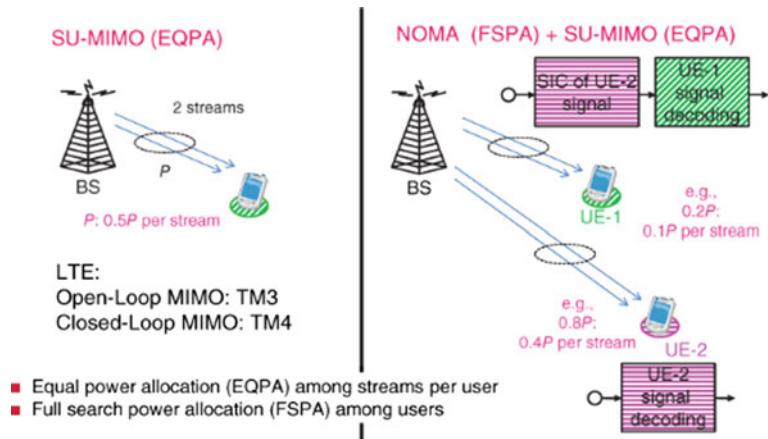


Fig. 3 NOMA and MIMO implementation

for example, SIC. With uplink, the received signal powers of different EUs already have differing power domain values, since the signals sent by various UEs that have been received at the BS are different channels. On the other hand, when NOMA is applied in uplink, ICI increases significantly because multiple UEs are simultaneously allowed for transmission, and when NOMA is applied, ICI does not increase in downlink as BS usually has constant power regardless of the amount of multiplexed UEs (Fig. 4).

As several EUs multiply and share the same non-orthographical resource, the EU transmission power limit will limit all of the EU NOMA's resource allocation, either in OFDMA downlink schedules or as SC-FDMA top link schedules without a NOMA. The transmission power limit can be limited. The group of NOMA UE {UE-1, UE-2}, to achieve the greatest transmission potential in the broader phase,

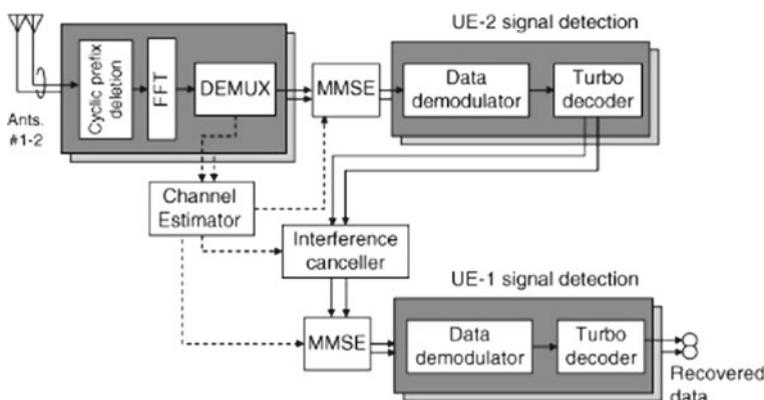


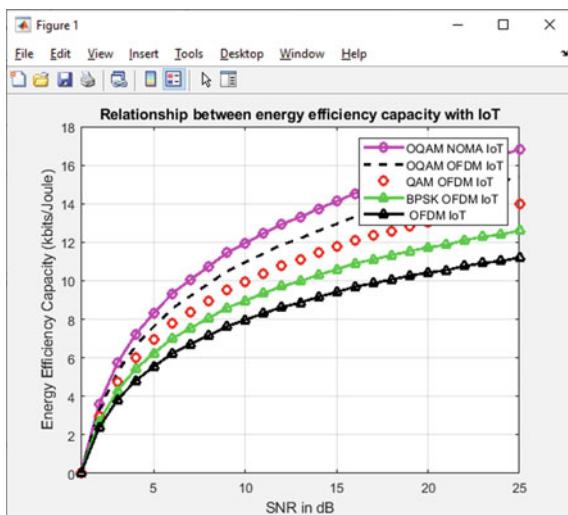
Fig. 4 Noise reduction in the signal

to ensure the subsequent allocation of resources. For SC-FDMA, if the EU reaches its maximum transmission power, BS will literally prevent sub-bands from being assigned. For NOMA, the situation is much harder, as the BS assigns sub-band groups to an EU rather than an EU group, and the transmission capabilities of UEs between the coupled UEs can be modified with the same general transmission strength. From this viewpoint, the optimum scheduling should be designed in an appropriate computation complexity with the popular architecture of the TPC, the UE selection algorithm and the algorithm for the sub-band assignment. To address this challenge, three sub-optimal solutions are possible. One is to interrupt the EU Group's expansion process, that is to say, the NOMA UE team and its subsets are no longer competing for the distribution of sub-bands. The second is to avoid the EU's expansion procedure to achieve the full capacity of transmission while further expansion of the other EUs, that is to say that the NOMA UE community is no longer competitive with the sub-band allocation but can join their subsets. In other words, both a NOMA UE party and subsists will compete for sub-band allocation, with both UEs continuing the enlargement process until the sub-band allocation has been completed. The P_{\max} is similarly distributed to all its assigned sub-bands for the union that achieves the highest transmitting capacity.

4 Experiment Result

From Fig. 5, the energy efficiency capacity in terms of Kbits/joules is calculated with respect to SNR in dB. The orthogonal quadrature amplitude modulation (OQAM) NOMA for IoT is proposed and compared with other techniques as shown in Fig. 5.

Fig. 5 SNR and capacity



The synchronization error is minimized in terms bit error rate (BER). The 5G wireless communication is established For IoT using NOMA. The OQAM NOMA IoT is proposed by minimizing synchronization error in terms of BER. It is compared with other techniques as shown in Fig. 6.

Figure 7 shows that the actual amount of data sent/received successfully through the communication connection in a throughput. The result is shown as Mbps or percentage and can vary from bandwidth because of a variety of technical problems, including latency or delay. The proposed methodology minimizes sampling time

Fig. 6 Error reduction

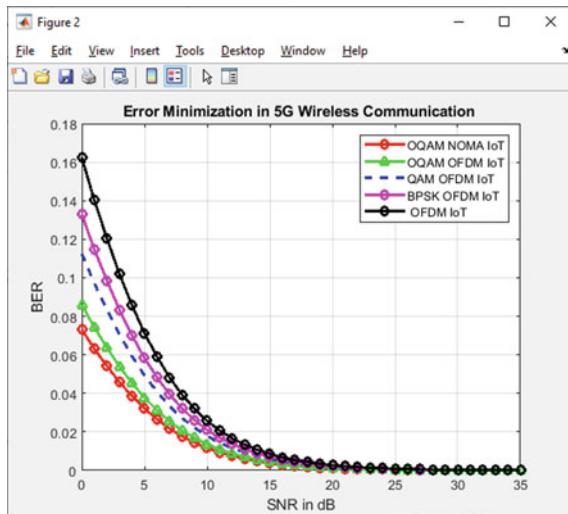
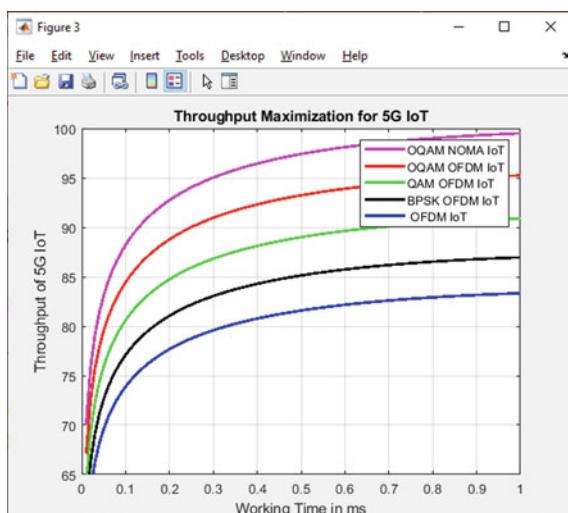


Fig. 7 Throughput maximization



offset (STO). The OQAM NOMA IoT minimizes the STO (Figs. 8 and 9). The OQAM NOMA for IoT has high throughput for 5G wireless communication. Figure 10 represents that the STO should be minimized during the reception of the transmitted data at the receiver. The STO issue of 5G wireless communication using NOMA leads to high bit error rate.

Figure 11 represents the MIMO NOMA for 5G IoT, and it is implemented on fading channel like Rayleigh channel. The Rayleigh fading channel provides unpredictable noise addition during channelization of NOMA. The NOMA IoT Rayleigh

Fig. 8 Sampling time

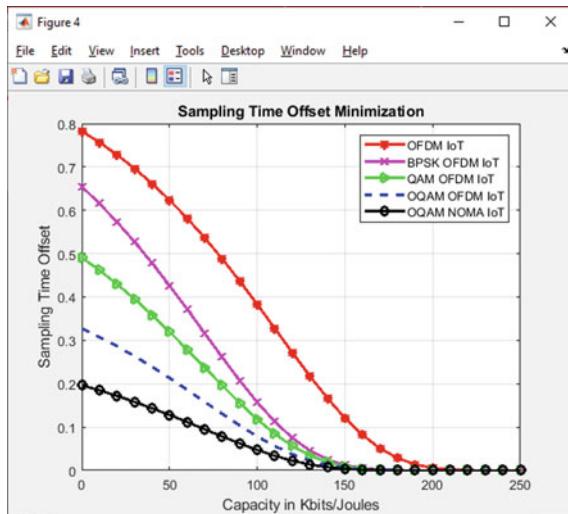


Fig. 9 5G channels

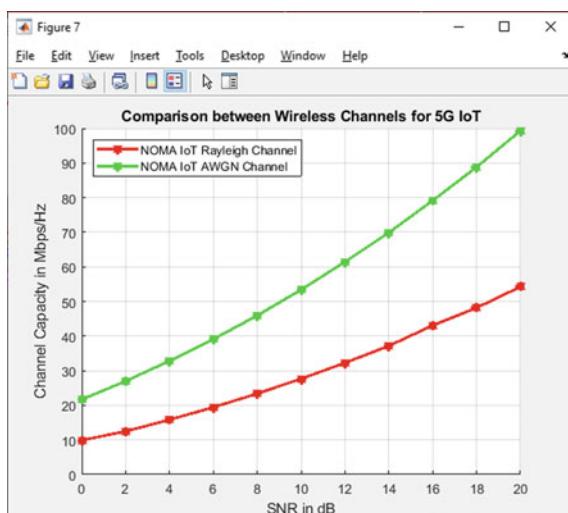


Fig. 10 Bandwidth estimation

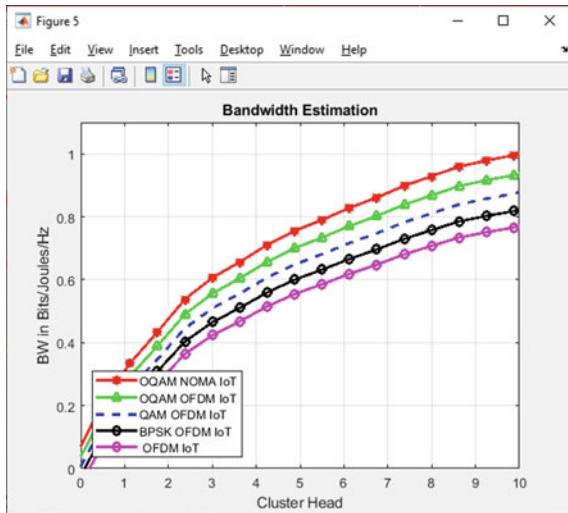
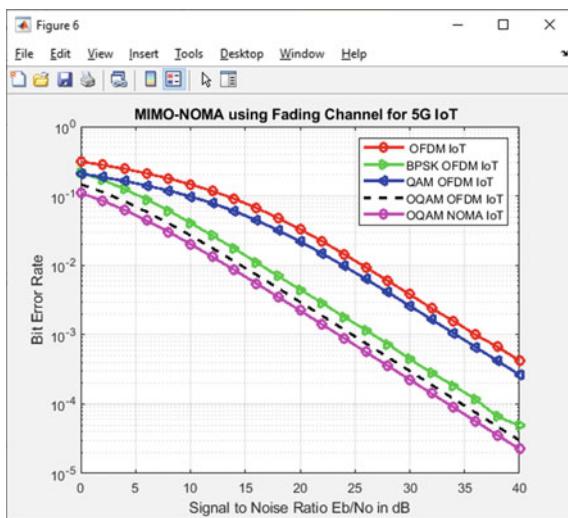
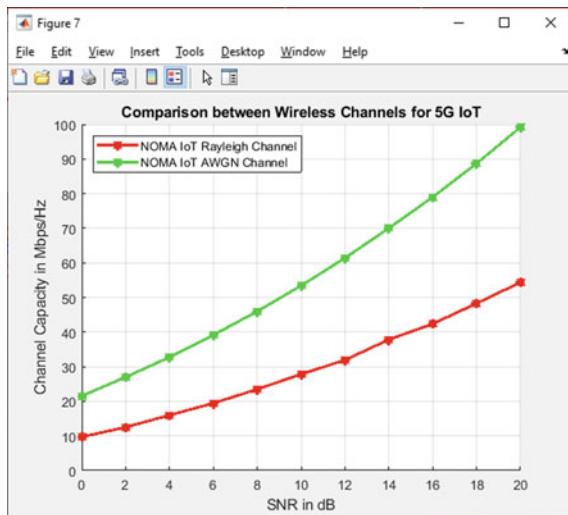


Fig. 11 Channel fading



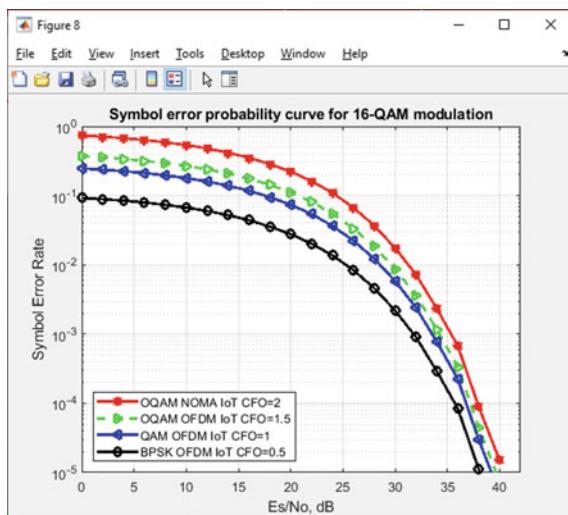
channel and AWGN channel are compared. The AWGN performance is better than Rayleigh channel. The efficiency of the bandwidth is shown in the diagram for the bits per second transmitted over the provided bandwidth (in bps/Hz), while the efficiency of the bandwidth is represented as the energy needed per bit for efficient, background-to-noise communications. In comparison with existing systems, OQAM NOMA IoT offers better results.

Figure 12 represents the MIMO NOMA 5G wireless communication for IoT, and it is implemented on fading channel such as Rayleigh channel. The Rayleigh channel is highly fading channel due to unpredictable noise during channelization. During

Fig. 12 5G IoT channels

this situation, OQAM NOMA for IoT is working better than other techniques. The Rayleigh fading channel is compared with additive white Gaussian noise (AWGN) channel. The performance of AWGN channel is better than Rayleigh channel for NOMA IoT for 5G Wireless communication.

Figure 13 represents that the carrier frequency offset (CFO) is minimized for OQAM NOMA IoT. The CFO occurrence on OQAM NOMA is tested for various CFO like 2, 1.5, 1, 0.5, 0. While CFO = 0, the performance of OQAM NOMA is

Fig. 13 Symbol error probability

better than other CFO occurrence. The CFO minimization is used to improve the signal-to-noise ratio (SNR) and reduction in BER.

5 Conclusion

The CFO and STO estimate issue for the NOMA 5G framework has been implemented in this paper. In view of the selective frequency fade and the nature of its basic receptor structure, the objective is to improve the robustness of the CFO and SFO estimates. OQAM NOMA was recognized as a promising applicant to 5th generation networks for the IoT wireless networking scheme. The orthogonal frequency divide multiplexing (OFDM) technique can be substituted by the NOMA system in order to achieve a superior bandwidth efficiency. The implementation of the NOMA method causes errors on the system of carrier frequency offset (CFO). The effect of the CFO on the success with the OQAM technique is improved in this project. The effects of CFO errors on the OQAM NOMA system will be seen in simulation results. Also presented and evaluated are the performance of the proposal estimate algorithm.

References

1. Visuvanathan Ganesan E, Thangappan J (2021) CFO and STO estimation and correction in multicarrier communications using linear filter bank multicarrier. *Trans Emerging Telecommun Technol* 32:3
2. ETSI Normalization Committee (1995) Radio broadcasting systems: digital audio broadcasting (DAB) to mobile, portable and fixed receivers; ETSI ETS 300 401; ETSI Normalization Committee, Sophia-Antipolis, France
3. ETSI Normalization Committee (2009) Digital radio Mondiale (DRM)—system specification; ETSI ES 201 980 V3.1.1; ETSI Normalization Committee, Sophia-Antipolis, France
4. ETSI Normalization Committee. (1997) Digital video broadcasting (DVB): frame structure, channel coding and modulation for digital terrestrial television (DVB-T); ETSI ETS 300 744; ETSI Normalization Committee, Sophia-Antipolis, France
5. ETSI Normalization Committee (2015) Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2); ETSI EN 302 755 V.1.4.1; ETSI Normalization Committee, Sophia-Antipolis, France
6. Institute of Electrical and Electronics Engineers (IEEE) (1999) Wireless LAN medium access control (MAC) and physical layer (PHY) specification: high-speed physical layer in the 5 GHz band; IEEE Std 802.11a; IEEE, Piscataway, NJ
7. Balogun MB, Takawira F, Oyerinde OO (2019) Uplink OFDM based NOMA plagued with carrier frequency offset errors. *IEEE AFRICON* 2019:1–5. https://doi.org/10.1109/AFRICO_N46755.2019.9133906
8. Martinez AB, Matth M, Chafii M, Fettweis G (2019) Blind carrier frequency offset estimation in generalized frequency division multiplexing. In: 2019 international conference on computing, networking and communications (ICNC), pp 869–875. <https://doi.org/10.1109/ICNC.2019.8685592>.

Middleware and Security Requirements for Internet of Things



Bharat Bhushan

Abstract The Internet of Things (IoT) confronts a future that enables the coupling of digital and physical objects where objects, users, and computing systems co-operate for economic benefits and convenience. The IoT's characteristics, like network and device-level heterogeneity, large-scale networks, numerous events generated simultaneously, will make diverse application development a challenging task. Middleware can ease such a development process by supporting interoperability within the diverse services and applications. Also, security is a rudimentary factor for various IoT applications; thus, communication stack of IoT must deliver the desired level of reliability, internet connectivity, and efficiency mechanism for protecting communications. The paper first outlined a set of middleware requirements for IoT and then presented the existing middleware solutions contrary to those requirements. Furthermore, the paper analyzed the existing protocols in the IoT. The concept of how the fundamental security requirements are ensured by the existing perspective in the IoT is also estimated.

Keywords Internet of Things · Wireless sensor networks · 6LoWPAN · Middleware requirements · IEEE 802.15.4 · End-to-end security

1 Introduction

Imposing standards is impractical in a ubiquitous computing environment. A huge number of events generated by these objects or things along with heterogeneous technologies of Internet of Things (IoT) throw light on new challenges in application development making ubiquitous computing even more difficult. A middleware can ease application development by the integration of communication devices and heterogeneous computing and supporting interoperability within various applications. To support the development of middleware solutions, many operating systems are developed. This middleware resides on different physical devices and also

B. Bhushan (✉)

School of Engineering and Technology (SET), Sharda University, Greater Noida, India

enables service deployment by providing necessary functionalities. The basic essential components of IoT are wireless sensor networks (WSNs), machine-to-machine communications (M2M), radio frequency identification (RFID), and supervisory control and data acquisition (SCADA). An IoT middleware integrates these technologies to entertain diverse application domains. No existing middleware is capable of supporting all these requirements for IoT or WSNs applications. Perera et al. studied all the existing WSN middleware and identified that these middleware solutions do not support context awareness. In the past few years, much IoT-specific middleware is reported [1].

The IoT is a collection of interconnected objects, people, devices with electronics, software, sensors, actuators, communication tools that will allow you to share data, provide communication along with better connections and accomplished the common goals of communication in a variety of environments with applications. IoT is the integration of people, equipment, and everyday items/gadgets on the World Wide Web (WWW). The IoT co-relates the human interaction domain with the machinery one [2]. The basic concept of the IoT is to be a part of the living by legalizing and gleaming the machines/devices from anywhere to run everyday tasks by way of the least of human intervention. The IoT is also being assimilated for many purposes such as television, news, shows, cartoons, films, transport, agriculture, health care, manufacturing, distribution of energy, and a lot of other areas, which require that communications over the internet are a smart way to carry out tasks without human intervention [3]. Figure 1 shows all the coverage area of IoT in which manner IoT works or connects to the physical world.

As much as there is a lot of good in the IoT, there is another hand too, that is to say, there are a lot of problems in the IoT, among them, to ensure safety is one of the most important and prior issues that need our concerns. Providing security in IoT devices is a very difficult task due to various reasons such as the lack of secure borders, and simple arithmetic, attacks by the malicious nodes in the network, lack centralized control, low power, load balance, no protection of the small channels, the dynamic topology changing behavior [2]. Table 1 describes the heterogeneity of device in IoT.

Furthermore, a summary of the involvement of this effort is enumerated as below:

- The work discusses the need and importance of IoT devices.
- The work highlights the background, features, application, challenges, security requirements of layered architecture, and security attacks of IoT devices for a better understanding of their actual need during their connection.

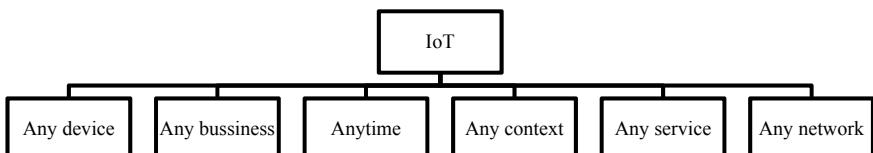


Fig. 1 Definition of IoT

Table 1 Heterogeneity of device

Device heterogeneity in IoT
Cloud or internet
The high-end computing device (SCADA)
Low-end computing devices
Middle-end computing devices (M2M communication unit)
Wireless sensors
RFID/NFC devices

- The work redefines the inspiration for securing physical, media access control (MAC), and network layer of IoT architecture that eases the connecting as well as routing facilities.
- The work explores some middleware solutions in IoT devices.

The remainder of the paper is prearranged as follows. Section 2 elaborates the background, various security attacks in IoT devices along with major consideration, where comparison of different kind of attacks is shown via table. Moreover, Sect. 3 described the security requirements for layered IoT architecture along with security parameters of physical, MAC, and network layer. Furthermore, Sect. 4 defined middleware solutions present in IoT for securing the overall network for user satisfaction. Finally, the paper concludes with Sect. 5.

2 IoT Background and Security Attacks

With the evolution of internet communications infrastructure, to secure communications among devices in the IoT applications like smart cities, smart grids, home automation, appropriate mechanisms are required. This section describes the basic information of IoT, and some kind of security attacks that are found in IoT devices are discussed.

2.1 *Background*

After many research contributions in the field of communication and low energy sensing applications, the demands of integrating these technologies with the internet are growing. The collected information by the different IoT devices may vary as it can range from temperature or humidity data, a simple heartbeat data, to the location of the user and his habits. This leads to certain privacy issues. These devices become a prime target for the adversary or attacker because of this collected information as the attackers may look to gain access to the collected confidential data. The security in IoT is a little bit neglected due to the short time to market and cost reductions

driving the development and design process. Devices that do not use any protection employ software-level solutions like firmware signing. These solutions are sometimes insufficient as various usage patterns of IoT are not considered. Moreover, this makes hardware vulnerable to attacks allowing different new attack vectors [4].

2.2 *Security Attacks*

In this section, some attacks are discussed, which are performed over the network to crack the security of the whole network as well as to harm the IoT devices.

- Replay attack: It is a fairly low-level and highly damaging attack, which only replicates the data packets and sends it over a transmission medium, again and again, that is further used to perform any malicious activity in the same device or as well as another device.
- Time attack: In this type of attack, unwanted delays are provided during the connection of two or more IoT devices and to the communication channel while data packets are sending. The main goal is to gain the unauthorized access to sensitive as well as confidential data [5].
- Data sniffing: Intercepting the communication by scanning data packets or messages, unwanted calls made by unplanned customers is also done in this attack to gain some information. This has an impact on online privacy and also known as Traffic analysis attack or Eavesdropping attack.
- Man-in-the-middle attack: All the significant data is to be inhaled through a hacker/attacker, which is located in the middle of source and destination, and also, the attacker checks the data throughout the broadcasting of the data over the channel to gain access and to harm the network as well as the user.
- Selective forwarding attack: All the signals or data packets are intercepted and then some of them are selected randomly by the attacker to send them to their destination, and the rest will have to be rejected or save them back for future purposes [6].
- Sybil attack: One node or one device got hijacked by the attacker, and then those nodes misguide the route by showing multiple identities of a single node. In short, it has an impact on accessibility as the first device takes part in communication and then left, but on the other hand, it will have a profound impact on all of the other security purposes.
- Data alteration: Data packets are seized by the hacker/attacker, then certain of the alterations have been done to them, after that simply and forwarded to their original destination.
- Sleep deprivation attack: In this attack, the device/nodes will be forced to stay awake for the minimization of the lifespan of their battery and to shut down the devices to make them out of service. In simple words, battery of devices is consumed by the attacker.

- Denial of service attack: It allows an attacker/hacker to introduce a huge number of bogus requests into the network, causing network congestion so that the communication has not been done. Sometimes, a malicious node can pretend to be very busy and deny communication with others. They have been an influence on the availability of IoT devices in the network [7].
- Queue jumps/cuts in the line: The attacker/hacker can itself or it can help to another who are standing in the waiting line, and the attacker jumps from backward to forward to make a connection and then performed malicious activities. The aim is to seize the resources and the private information. Table 2 summarized and compared all the above-discussed attacks based on some parameters.

2.3 Major Consideration

The major considerations for IoT applications are complexity, energy-efficiency, and scalability.

- Complexity: The overall complexity can be defined in two major ways: computational load involved in encoding-decoding operations and the cost involved in consumer security index (CSI) acquisition at the sensors. The method based on compressive sensing does not meet the requirements on both counts. In such compressive sensing methods, sensors before transmitting have to collect N observations, and the legitimate fusion centers are responsible for solving the minimization problem. The bit-flipping algorithm and 1-bit quantization method in type-based multiple access along with channel-aware encrypting have the lowest transmission complexity. Out of all these, computer-aided engineering (CAE) shows the lowest requirement and CSI computational complexity. This is because CAE involves non-coherent digital modulation [8].
- Energy-efficiency: Artificial noises have the worst energy-efficiency as multiple sensors waste their energy in sending jamming signals. In probabilistic ciphering, it is assumed that all the sensors in the subsystem are activated at the same time. For non-identical channels and sensors in 1-bit quantization, only the strong sensors kept activated, while others are in sleep mode. CAE and censoring show better performance in terms of energy utilization. This is mainly because of the existence of a simple threshold comparison for indicating whether the sensors should stay silent or should transmit [9].
- Scalability: In security protocols, the number of sensors increases rapidly due to complex deployments in IoT; thus, scalability is a very important aspect of IoT sensor security. The protocol parameters should involve mechanisms to minimize the recompilations if new sensors are deployed. Artificial noise methods are the best scalable as these assume sensors to make decisions autonomously regarding the amount of power allocated for jamming. This may give rise to huge co-channel interference to normal nodes. All other methods like TBMA or CAE require re-computation overhead as the number of sensors increases [10].

Table 2 Comparison of security attacks

Name	Type of attack	Action of attack	Solution	Attack performed on layer	Harm
Replay attack	Side channel	Send the message again and again	Time stamps, one time password, session identifiers	Perception layer	Availability, data freshness
Time attack	Cryptanalysis	Attacker comprises the crypto algorithm by analyzing the time which is taken to complete the crypto algorithm	Secure Advanced Encryption Standard (AES), average clock size	Perception layer	Availability, confidentiality, integrity, authorization, non-repudiation
Data sniffing	Network	Sniff the data packets to find confidential information	Use highly strong encryption methods	Network layer	Authentication
Man-in-the-middle attack	Physical/network	Sniffed important information	Must use trusted device	Network layer	Availability, confidentiality, integrity, authorization, non-repudiation
Selective forwarding attack	Cryptanalysis	Seizure of all the data signals/packets and send few of them	Acknowledge mechanism	Middleware layer	Confidentiality
Sybil attack	Network	More than one identity available for a single node	A trusted device, cross-check the certification of nodes	Middleware layer	Availability
Data alteration	Network	Data packets capture and then send after alter them	Share point central administration	Application layer	Integrity
Sleep deprivation attack	Network	Keep the node awake to minimize their life	Head-based cluster	Application layer	Availability

(continued)

Table 2 (continued)

Name	Type of attack	Action of attack	Solution	Attack performed on layer	Harm
Denial of service attack	Network	Malicious device acts as a busy device and rejecting/failing the sending of data signals/packets	Use inducement method constructed on the denial	Business layer	Availability
Queue jumping attack	Side channel	Imprisonment of the resources and all private data	Not found yet	Business layer	Availability, confidentiality

3 Security Requirements for Layered IoT Architecture

The IoT which holds a large number of smart devices that are connected to a vast network with the help of a variety of networking technologies. In short, these technologies can be wired or wireless. This will make the construction more difficult and expensive to manage. Therefore, a well-formed and secured architecture is needed, and it is a structure of the specification of the physical network components and their functional organization with configuration, rules, and procedures for the operation as well as data formats that used in its operation [11].

3.1 Security Requirements in Physical and MAC Layer

IEEE 802.15.4 standard is designed to support energy-efficiency trade-off. It also supports data rate and range of communication. The major goal of employing IEEE 802.15.4 is to enable low-energy communication at both of these layers: physical and MAC layers. It supports communications in a short range of 10 m at 250 Kbit/s. IEEE 802.15.4 which was originally introduced in 2006 was updated in 2011 to include practical deployments and market applicability of the standard [12]. As IEEE 802.15.4 supports low-energy communication, it laid a base to design many technologies at higher layers such as 6LoWPAN or CoAP. This was also adopted by other WSN standards like ZigBee-2006, ISA 100.11a, ZigBee PRO, and WirelessHART. These are industry-accepted but do not support internet communications. This manages the radio frequency transceiver and the channel selection of the sensing device. It introduces reliability by introducing ultra-wideband (UWB) and direct sequence spread spectrum [13]. The MAC layer is responsible for data service management and facilitates network beaconing, validation of frames, and security solutions. It performs two functions, namely full function device (FFD) and reduce function device (RFD). FFD coordinates and manages a network of devices. RFD communicates with other FFD and RFD devices. Carrier sense multiple access with collision avoidance (CSMA/CA) is used to manage data communication collisions. The beacon frames provide configuration information and provide device synchronization [14].

3.1.1 Security in IEEE 802.15.4

It is responsible for security at higher layers of the communication protocol stack. This is because it supports efficient symmetric cryptography in IEEE 802.15.4 sensing platforms. It uses AES to deliver the required level of security.

- Confidentiality: Applications requiring link layer communication confidentiality use AES encryption technique in the counter mode. AES security modes uses 128-bit keys for the desired security.

- Data authentication and integrity: Applications that require integrity and authentication of link layer communications use AES in cipher block chaining (CBC) mode. These results in MAC are appended to the data. The header is encrypted, and the payload is unencrypted during the data transmission [15].
- Semantic security: The key control and the frame counter fields of IEEE 802.15.4 are set by the sender to provide message replay protection and semantic security support in all the security modes of IEEE 802.15.4.
- Access control mechanisms (ACL): The access control functionalities provided by IEEE 802.15.4 enable the sensors to use the frames' destination and source addresses to provide the desired level of security. ACL entry is also employed that stores IEEE 802.15.4 address to process communication security in the address field. This uses a cryptographic key and stores a high watermark of the packet identifier which is recently received in the replay counter field [16].

3.2 *Security Requirements in the Network Layer*

Internet architecture enables packets to traverse the networks using the heterogeneous link layer mechanism required for Internet Protocol (IP) packet transport. The IETF IPv6 over 6LoWPAN was designed in 2007 to enable IPv6 packets transportation over low energy IEEE 802.15.4 and other wireless communication environments. The 6LoWPAN adaptation layer enables IPv6 communications between constrained IoT sensors and other internet entities. This layer is responsible for mapping the services of the IP layer on the services delivered by IEEE 802.15.4 [17].

3.2.1 6LoWPAN Frame Format Compression

IEEE 802.15.4 supports MAC and physical layer communications. The data payload at higher stack layers is 102 bytes in the absence of security in the link layer. The adaptation layer of the 6LoWPAN optimizes the limited payload space usage through compression of the packet header while providing mechanisms to support IPv6 operations. The various Request for Comment (RFC) documents define the adaptation layer. RFC 4944 provides mechanisms for IPv6 packets transmission over IEEE 802.15.4 networks, while RFC 6282 defines the header compression. RFC 6775 defines the neighbor discovery optimizations (NDO) [18].

3.2.2 Security in 6LoWPAN

- Security vulnerabilities: This involves forging interface addresses that may lead to compromising the singularity of 6LoWPAN. This addresses the mesh routing and neighbor discovery mechanisms in the IEEE 802.15.4 environments. The AES security protects against threats at the link layer for constrained devices. RFC 6282

facilitates the compression of 16 User Datagram Protocol (UDP) port numbers to 4 bits. This increases the risk for an application such that the application gets wrong payload leading to misinterpretation of the message content. Thus, RFC 6282 uses maximal information coefficient (MIC) codes as security mechanisms for such ports [19].

- Security requirements: RFC 4919 addresses the security aspects at various levels in the protocol stack. It also uses IPSec for enforcing security mechanisms at the network layer. RFC 6568 enforces a security mechanism for wireless sensing devices. This focuses on the serious demands for survivability and resiliency in the wireless sensing device. RFC 6606 focuses on security and addresses the use of AES at hardware sensing platforms of IEEE 802.15.4. This also outlines the importance of self-organization, security, and time synchronization for delivering desired data security. RFC 6775 employs optimization mechanisms to enable neighbor discovery in 6LoWPAN environments [20].

4 Middleware Solutions in IoT

The existing design approaches of different middleware solutions are described below. Some middleware uses a different combination of design approaches. Generally, hybrid approaches are superior to individual design categories in terms of performance.

- Event-based middleware: The applications, components, and other participants in event-based middleware interact through events. A huge number of entities or application components are present in event systems that consume and produce events. The event notifications are sent asynchronously to the subscribers. This design approach focuses on non-functional requirements like availability, reliability, scalability, real-time performance, and security. Hermes is a type of event-based middleware that can either be attribute-based or type-based and was designed for large-scale distributed services or applications. It uses a fault-tolerance mechanism and a scalable routing algorithm to withstand various failures in the middleware. This event-based middleware is useful for a system having high failures and mobility. The major advantage of an event-based middleware is that it supports the decoupling of subscribers and producers.
- Service-oriented middleware: This type of middleware design builds applications or software in the form of services. The major characteristics of service-oriented computing include neutrality, service reusability, service discoverability, service composability, and loose coupling. This middleware can be either a standalone System on a Module (SOM) or cloud computing Platform as a Service (PaaS) model. Hydra is a service-oriented middleware for ambient intelligence systems and services. Its architecture consists of many management components, event manager, service manager, device manager, and security manager. It also provides

semantic- and syntactic-level interoperability. It is also known as LinkSmart. A service-oriented middleware supports abstraction and does not deal with code management. These do not scale well in ultra-large IoT environments.

- Virtual machine (VM)-based middleware: It provides programming support for user applications for a safe execution environment. Each node present in the network holds a VM for the interpretation of the modules. Mate is an example of VM-based middleware used for resource-constrained sensors. Other examples of VM solutions are Sensorware, Melete, Magnetos, Squawk.
- Agent-based middleware: In this approach, applications are subdivided into modular programs to simplify the injection and distribution throughout the entire network with the help of a mobile agent. It leads to the design of decentralized systems for partial failure tolerating.
- Tuple space middleware: In this type of middleware, every member of the entire network infrastructure has a local tuple space. It is a data repository for concurrent access. Here, the applications write tuples in tuple space for communication. LIME, TeenyLIME, and TinyLIME are examples of tuple space middleware solutions designed for specific environments like sensor networks or mobile and ad hoc networks.
- Database-oriented middleware: Here, a sensor network is assumed to be a virtual relational database system. Applications, using a Structured Query Language (SQL), can query the database for solving complex queries. SINA, along with handling events, also determines the querying node's mobility. It also supports resource management and enables sensor applications to adapt to changes within the sensor application environment.

5 Conclusion

A middleware is required to facilitate the development of a variety of IoT applications and services. Lots of ideas are aiming at this problem. The range is quite diverse and includes the different approaches to the design of the middleware, to meet the different requirements. This paper places this work in the perspective of a holistic view of the field. Additionally, the background, various security attacks in IoT devices along with major consideration are elaborated. Moreover, the security requirements for layered IoT architecture along with security parameters of the physical layer, MAC layer, and network layer are discussed. Furthermore, middleware solutions present in IoT for securing the overall network for user satisfaction are highlighted. Although the existing middleware solutions meet many of the requirements that are associated with the middleware in the IOTs, some of the requirements and the related research questions are still relatively unexplored, such as the scale or dynamic of discovery, be a resource composition, the system-wide scalability, reliability, security, privacy, collaboration, intelligence, integration, and contextual awareness. It is a great opportunity to work in the future in these areas.

References

1. Gautam S, Malik A, Singh N, Kumar S (2019) Recent advances and countermeasures against various attacks in IoT environment. In: 2019 2nd international conference on signal processing and communication (ICSPC). <https://doi.org/10.1109/icspc46172.2019.8976527>
2. Iqbal W, Abbas H, Daneshmand M, Rauf B, Bangash YA (2020) An in-depth analysis of IoT security requirements, challenges, and their countermeasures via software-defined security. *IEEE Internet Things J* 7(10):10250–10276. <https://doi.org/10.1109/JIOT.2020.2997651>
3. Goyal S, Sharma N, Kaushik I, Bhushan B (2021) Blockchain as a solution for security attacks in named data networking of things. *Secur Privacy Issues IoT Dev Sens Netw* 211–243. <https://doi.org/10.1016/b978-0-12-821255-4.00010-9>
4. Neshenko N, Bou-Harb E, Crichigno J, Kaddoum G, Ghani N (2019) Demystifying IoT security: an exhaustive survey on IoT vulnerabilities and a first empirical look on internet-scale iot exploitations. *IEEE Commun Surv Tutor* 21(3):2702–2733, thirdquarter 2019. <https://doi.org/10.1109/COMST.2019.2910750>
5. Jaitly S, Malhotra H, Bhushan B (2017) Security vulnerabilities and countermeasures against jamming attacks in wireless sensor networks: a survey. In: 2017 international conference on computer, communications and electronics (Comptelix). <https://doi.org/10.1109/comptelix.2017.8004033>
6. Shin D, Yun K, Kim J, Astillo PV, Kim J, You I (2019) A security protocol for route optimization in DMM-based smart home IoT networks. *IEEE Access* 7:142531–142550. <https://doi.org/10.1109/ACCESS.2019.2943929>
7. Bhushan B, Sahoo C, Sinha P, Khamparia A (2020) Unification of Blockchain and Internet of Things (BIoT): requirements, working model, challenges and future directions. *Wireless Netw.* <https://doi.org/10.1007/s11276-020-02445-6>
8. Bhatt G, Bhushan B (2020) A comprehensive survey on various security authentication schemes for mobile touch screen. In: 2020 IEEE 9th international conference on communication systems and network technologies (CSNT). <https://doi.org/10.1109/csnt48778.2020.9115731>
9. Hassija V, Chamola V, Saxena V, Jain D, Goyal P, Sikdar B (2019) A survey on IoT security: application areas, security threats, and solution architectures. *IEEE Access* 7:82721–82743. <https://doi.org/10.1109/ACCESS.2019.2924045>
10. Saxena S, Bhushan B, Ahad MA (2021) Blockchain based solutions to secure IoT: background, integration trends and a way forward. *J Netw Comput Appl* 103050. <https://doi.org/10.1016/j.jnca.2021.103050>
11. Liao B, Ali Y, Nazir S, He L, Khan HU (2020) Security analysis of IoT devices by using mobile computing: a systematic literature review. *IEEE Access* 8:120331–120350. <https://doi.org/10.1109/ACCESS.2020.3006358>
12. Arora A, Kaur A, Bhushan B, Saini H (2019) Security concerns and future trends of internet of things. In: 2019 2nd international conference on intelligent computing, instrumentation and control technologies (ICICICT). <https://doi.org/10.1109/icicict46008.2019.8993222>
13. Goel AK, Rose A, Gaur J, Bhushan B (2019) Attacks, countermeasures and security paradigms in IoT. In: 2019 2nd international conference on intelligent computing, instrumentation and control technologies (ICICICT). <https://doi.org/10.1109/icicict46008.2019.8993338>
14. Haque AK, Bhushan B, Dhiman G (2021) Conceptualizing smart city applications: requirements, architecture, security issues, and emerging trends. *Expert Syst.* <https://doi.org/10.1111/exsy.12753>
15. Lounis K, Zulkernine M (2020) Attacks and defenses in short-range wireless technologies for IoT. *IEEE Access* 8:88892–88932. <https://doi.org/10.1109/ACCESS.2020.2993553>
16. Varshney T, Sharma N, Kaushik I, Bhushan B (2019) Architectural model of security threats & their countermeasures in IoT. In: 2019 international conference on computing, communication, and intelligent systems (ICCCIS). <https://doi.org/10.1109/icccis48478.2019.8974544>
17. Sharma V, You I, Andersson K, Palmieri F, Rehmani MH, Lim J (2020) Security, privacy and trust for smart mobile- internet of things (M-IoT): a survey. In: IEEE access, vol 8, pp 167123–167163. <https://doi.org/10.1109/ACCESS.2020.3022661>

18. Malik A (2020) Steganography: step towards security and privacy of confidential data in insecure medium by using LSB and cover media (December 12, 2020). SSRN Electron J. <https://doi.org/10.2139/ssrn.3747579>
19. Mukherjee A (2015) Physical-layer security in the internet of things: sensing and communication confidentiality under resource constraints. Proc IEEE 103(10):1747–1761. <https://doi.org/10.1109/JPROC.2015.2466548>
20. Khanam S, Ahmedy IB, Idna MY, Idris, Jaward MH, Bin AQ, Sabri M (2020) A survey of security challenges, attacks taxonomy and advanced countermeasures in the internet of things. In: IEEE access, vol 8, pp 219709–219743. <https://doi.org/10.1109/ACCESS.2020.3037359>

Secure Multipath Key Establishment Solution in WSN



Charu Sharma and Rohit Vaid

Abstract Security is the common concern in wireless sensor networks. In many research, Diffie-Hellman key exchange algorithm is used to provide secure data communication between nodes. But, this algorithm is a non-authenticated protocol due to which in the absence of authentication, it is easily susceptible to man-in-the-middle (MITM) attacks. The malicious node can easily interrupt the communication, by appearing as a valid node in the network. To overcome this problem, this paper presents a mechanism secure multipath key establishment with modified Diffie-Hellman (SMKE-DH) which detects and mitigates Sybil nodes and makes the network trustful with the purpose of resolving MITM attack problem, after then secure keys are established between two communicating nodes for data transmission. Performance evolution of SMKE-DH is evaluated on NS-2 which shows that proposed algorithm performs better than existing elliptic curve cryptography (ECC) algorithm by considering two QoS parameters average E2E delay and throughput.

Keywords Diffie-Hellman · Elliptic curve cryptography · Wireless sensor networks

1 Introduction

In modern era, wireless sensor networks (WSNs) are deeply investigated and thus have attracted major interest from the research area due to their potential in a broad collection of applications such as battlefield sensing, environmental monitoring, forest fire detection, hazard leakage detection. Unlike wired network with high processing power and large bandwidth, WSNs have some unique features. In WSNs, the sensor nodes (SNs) have limited resources in terms of energy, bandwidth,

C. Sharma (✉) · R. Vaid
CSE Department, M. M. Engineering College, M.M (Deemed to be University), Mullana,
Ambala, Haryana 133207, India
e-mail: er.charusharma@mmumullana.org

R. Vaid
e-mail: rohitvaid@mmumullana.org

and computational capabilities [1–3]. These nodes are randomly deployed in an unattended fashion which makes it susceptible to physical attacks. The topology of these networks changes very frequently due to link failure or mobility of sensor nodes. Unattended nature and wireless medium increases the possibility of various attacks in the network.

1.1 Multipath Routing

Single-path routing between the source and the destination node is very general, but it suffers with the two major problems, i.e., path failure which results in loss of data packets in the network, and if that single path is compromised, there is no alternative to send the data to the destination securely. To overcome this problem, ad hoc on demand multipath distance vector routing protocol (AOMDV) is used which provides alternative paths to reach the data packets to the destination.

1.2 Security Attacks in WSNs

Various types of security attacks in WSNs are:

- **Denial of service attack:** The aim of this attack is to stop the proper functioning of the network. The attacker blocks the resources for authorized nodes.
- **Wormhole attack:** This attack may have more than one malicious node and a tunnel between them. The malicious node captures the packets from one place and tunnels those packets to another far located malicious node which in turn retransmit those packets locally.
- **Sybil attack:** This attack disturbs the normal functioning of the network. A malicious node with the same identity can present in several locations in a single network [4].
- **Selective forwarding attack:** In this type of attack, nodes act as a black hole and deny sending each packet the node receives.
- **Replay attack:** In this, mugger repeats the broadcast of precedent information to take benefit of the communication at the stage of sending.
- **Message suppression attack:** In this, malicious node can selectively drop data packets from the network. These data packets possibly might have essential information for the receiver.

1.3 Security Requirements in WSNs

Security is the important and challenging task in WSNs. The goal of security is to protect information and resources from attacks. So, communication between SNs

and with the base station (BS) should be protected against security threats. Various security requirements in WSN's are:

- **Confidentiality:** It ensures that private data is protected, and only legitimate users can read a communication; eavesdroppers cannot.
- **Authentication:** This process assures that the source of the data is correctly recognized before using it in the network.
- **Integrity:** It ensures that the original data is transmitted from source to destination accurately without any alteration.
- **Availability:** It ensures that all required resources are accessible by the SNs in the network whenever needed.
- **Non-repudiation:** It provides the assurance that the node cannot deny that it had not send the message in the network.
- **Forward secrecy:** It gives assurance that when current secret key is compromised, it will not be allowed to compromise any secret keys further.
- **Backward secrecy:** It gives assurance that when current secret key is compromised, it will not be allowed to compromise any prior secret keys.

1.4 Existing Techniques

- **Diffie-Hellman key exchange (DHKE):** For security applications, DHKE procedure is widely used. DHKE provides robust safety and has very less key pair generation time. But, DHKE is vulnerable to MITM attack as it does not provide authentication between nodes. To overcome the mentioned drawback, a way to provide validation for DHKE procedure is required [5].
- **Elliptic curve cryptography:** ECC is a node authentication protocol. It covers system initialization, SN registering, and validation process. It carries on to use the main scheme's high powerful, and efficient multiplication operation process and the SN keep the other ID's hash value as the proof of the authentication. ECC with key size of 162 bits provides the same safety as RSA with key size of 1024 bits. ECC is widely used for IoT nodes for verification purposes [6].

2 Related Work

Araujo et al. [7] highlighted different challenges in WSNs and also describe various types of attacks along with countermeasures to control these attacks.

Authors in [8] presented a survey on classification on different attacks in WSNs and their respective defensive measures. In this research work, author classifies various secured routing protocols with comparison of each with different parameters to build a highly secured network.

In [9], authors discussed various schemes and protocols on multipath strategy which defines the areas for future development in wireless multimedia networks.

Authors in [10] established taxonomy of different forms of Sybil attack and an enhanced scheme countermeasure compared to each type and examine their efficiency.

Rosheen Qazi et al. presented a novel algorithm in [11], which is based on ECC for WSNs. The algorithm not only provides security but also saves memory space on SNs using elliptic curve digital signature (ECDSA) cryptographic scheme and can be used for many real-world networks.

Authors in [12] use ECC with beta and gamma functions for safe transmission in WSNs. There are many schemes that use ECC for WSNs, and this is the reason why there is a need for a new secured algorithm for WSNs. Authors also highlighted the drawback of using DH algorithm for WSNs. DH is easily susceptible to the MITM attack if applied without node authentication.

To overcome this problem, proposed algorithm constructs a network of trusted nodes to spot and remove Sybil nodes. To offer safe communication between nodes, hashing is used. DH with hash function is used for key generation and key exchange as validation. Only the trusted nodes can break the hashing because correct hash value can only be determined if the correct process is applied.

3 Proposed Work

The proposed algorithm works in two steps:

Step 1: First create a network of trusted nodes which finds and removes Sybil nodes from the network.

The SNs are randomly deployed. After network setup, topological verification of all the SNs is done by the BS. The nodes with minimum data packet drop in the network are selected as trustworthy nodes. Cluster nodes are selected from trustworthy nodes. Each cluster nodes is assigned some member nodes. Each member node sends its ID and power value to its own cluster node. Cluster node compares the power value of each member nodes with the threshold value. If the power value is below threshold, then the node is spotted as Sybil node and is removed from the network.

Step 2: Apply Diffie-Hellman along with hash function (modified-DH).

DH along with hash function is used as validation during key generation and key exchange process. Two communicating nodes select both public keys and private keys. Both the communicating nodes generate the keys using Diffie-Hellman. Hash values of generated keys are calculated. After then, between two communicating parties, exchange the generated keys as well as the calculated hash value. Both the nodes again calculate hash value of received keys. After validating both the keys by matching hash value, now, use the interchange keys to generate the official keys. During this process, only trusted nodes are capable to break the hashing and validate the data.

4 Results and Discussion

The network is simulated using NS-2 simulator with a network size of 1000×1000 . Different QoS parameters are used to evaluate the efficiency of WSNs. Four different cases are considered by varying number of nodes 25, 50, 75, and 100 for evaluation.

Case 1: WSN without IoT

The SNs are randomly deployed, and AOMDV protocol is used for multipath routing for transmitting data from nodes to BS. Table 1 shows the QoS parameters of WSN without IoT.

Case 2: WSN with IoT

The SNs are randomly deployed along with IoT devices. As a result, SNs are able to transmit data to BS through IoT devices with minimum delay. This also enhances the PDR and throughput of the network. Table 2 shows the QoS parameters of WSN with IoT.

Figure 1 shows that the PDR of the network without IoT is less due to funneling effect in the network.

As shown in Fig. 2, WSN with IoT achieves better performance in terms of throughput as compared to the network without IoT, and the average E2E delay of the data transfer in the network reduces when IoT is used in the network as shown in Fig. 3.

Table 1 WSN without IoT

No. of nodes	PDR (%)	Throughput (Kbps)	Average end-to-end delay (ms)
25	93.58	20.02	0.042835
50	84.94	13.49	0.038882
75	80.63	6.98	0.051096
100	73.76	4.15	0.051397

Table 2 WSN with IoT

No. of nodes	PDR (%)	Throughput (Kbps)	Average end-to-end delay (ms)
25	100	21.39	0.001537
50	100	15.89	0.001656
75	99.4	8.60	0.001471
100	99.2	5.32	0.001517

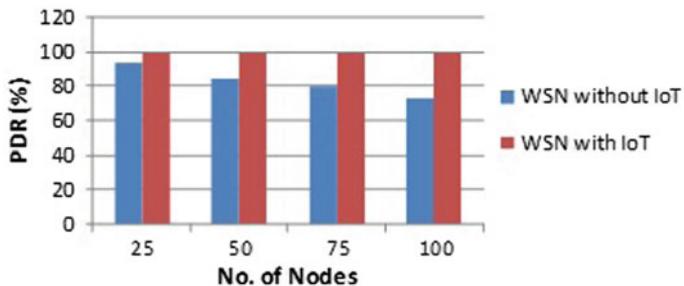


Fig. 1 PDR (%) comparison of WSN without IoT and WSN with IoT

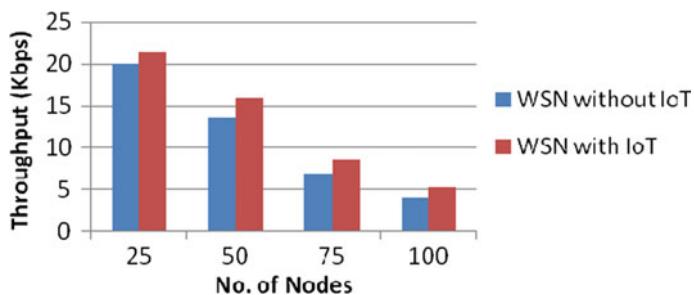


Fig. 2 Throughput comparison of WSN without IoT and WSN with IoT

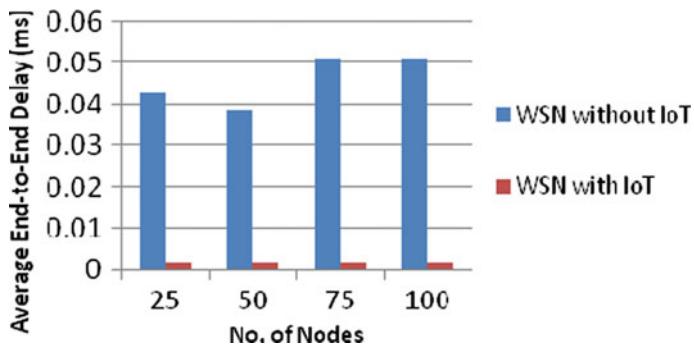


Fig. 3 Average end-to-end delay comparison of WSN without IoT and WSN with IoT

Case 3: SMKE-DH

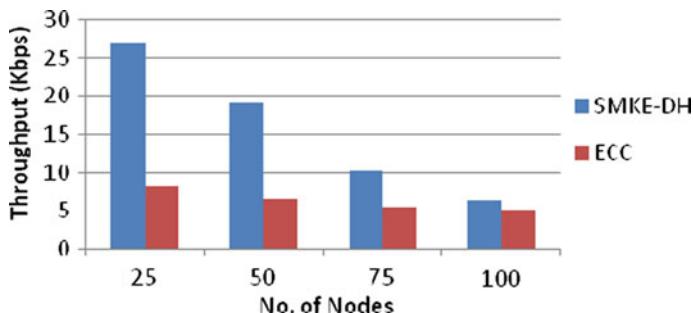
In this scenario, we have evaluated the QoS parameters of proposed SMKE-DH as shown in Table 3.

Table 3 SMKE-DH

No. of nodes	Throughput (Kbps)	Average end-to-end delay (ms)
25	26.99	0.001529
50	19.12	0.001562
75	10.11	0.001637
100	6.36	0.002013

Table 4 ECC

No. of nodes	Throughput (Kbps)	Average end-to-end delay (ms)
25	8.26	0.031500
50	6.46	0.031590
75	5.51	0.031639
100	5.07	0.031657

**Fig. 4** Throughput (Kbps) comparison of SMKE-DH with ECC

Case 4: QoS Parameters Comparison with Existing Method

To evaluate the efficiency of proposed algorithm, it is contrast with the existing ECC. Two performance metrics are considered for evaluation as shown in Table 4.

As shown in Figs. 4 and 5, proposed SMKE-DH achieves better performance in throughput and minimizes average delay when compared ECC.

5 Conclusion and Future Scope

Data safety is a key concern in WSNs. DH is susceptible to MITM attack as it does not provide authentication between nodes during communication. In this paper, we presented an algorithm SMKE-DH which is used to detect and mitigate Sybil nodes from WSN and also establish multipath using AOMDV. Simulation results show that

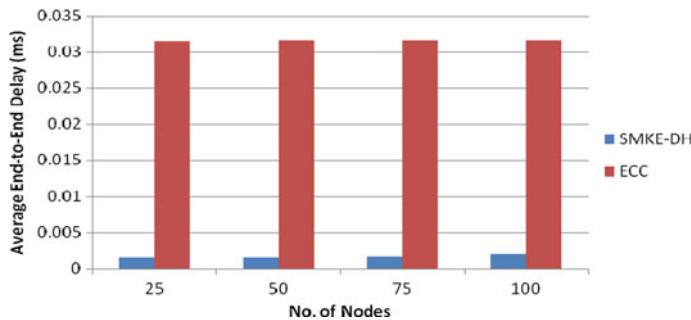


Fig. 5 Average end-to-end delay comparison of SMKE-DH with ECC

SMKE-DH gives higher throughput, lower average E2E delay as compared to ECC. There exist some possible additions to our work such as to calculate computational complexity of SMKE-DH.

References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. *Comput Netw* 38(4):393–422
2. Vaid R, Katiyar V (2013) Security issues and remedies in wireless sensor networks—a survey. *Int J Comput Appl* 79(4):31–39
3. Bansal S, Juneja D, Mukherjee S (2011) An analysis of real time routing protocols for wireless sensor networks. *Int J Eng Sci Technol (IJEST)* 3(3):1797–1801
4. Newsome J, Shi E, Song D, Perrig A (2004) The sybil attack in sensor networks: analysis & defenses. In: Third international symposium on information processing in sensor networks. IPSN 2004. IEEE, pp 259–268
5. Pal O, Alam B (2017) Diffie-Hellman key exchange protocol with entities authentication. *Int J Eng Comput Sci* 6(4):20831–20839
6. Chang Q, Zhang YP, Qin LL (2010) A node authentication protocol based on ECC in WSN. In: International conference on computer design and applications IEEE, vol 2, pp V2-606
7. Araujo A, Blesa J, Romero E, Villanueva D (2012) Security in cognitive wireless sensor networks. Challenges and open problems. *EURASIP J Wireless Commun Netw* 2012(1):1–8
8. Azeem MA, Pramod A (2011) V: security architecture framework and secure routing protocols in wireless sensor networks-survey. *Int J Comput Sci Eng Surv* 2(4):189
9. Anasane AA, Satao RA (2016) A survey on various multipath routing protocols in wireless sensor networks. *Proc Comput Sci* 79:610–615
10. Newsome J, Shi E, Song D, Perrig A (2004) The sybil attack in sensor networks: analysis & defenses. In: Third international symposium on information processing in sensor networks, 2004. IPSN 2004. IEEE, pp 259–268
11. Qazi R, Qureshi KN, Bashir F, Islam NU, Iqbal S, Arshad A (2021) Security protocol using elliptic curve cryptography algorithm for wireless sensor networks. *J Ambient Intell Humaniz Comput* 12(1):547–566
12. Ali S, Humaria A, Ramzan MS, Khan I, Saqlain SM, Ghani A, Zakia J, Alzahrani BA (2020) An efficient cryptographic technique using modified Diffie–Hellman in wireless sensor networks. *Int J Distrib Sensor Netw* 16(6):1–24

Controlling Node Failure Localization in Data Networks Using Probing Mechanisms



K. Saritha, Bingi Manorama Devi, Muralidhar Kurni,
Debabrata Samanta , and Niju P. Joseph

Abstract In this paper, we prospect the potency of node failure localization in network communication from dual states (normal/fizzle) of the source to destination paths. To localize the failure nodes individually in the scheduled nodes, dissimilar path states must connect with various events of failure nodes. But, this situation is inapplicable or not easier to investigate or apply on enormous networks due to the obligation of any viable failure nodes. This objective is to deploy the set of adequate conditions for recognizing a set of failures in a set of arbitrary nodes which can be verified in a stipulated time. To avoid the above situation, probing mechanisms are assimilated additionally as a combination for network topology and locations of scrutinizes. Three probing mechanisms are considering which vary depending on measurement paths. Both the procedures can be transformed into single-node possessions by which they can be calculated effectively based on the given conditions. The exceeding measures are proposed for measuring the potency of failure localization which can be utilized for assessing the effect of different factors, which comprises topology, total monitors, and probing mechanisms.

Keywords CAP · Internal network · MANET · Message authenticate code · Probing mechanisms

K. Saritha

S. V. Degree & P. G. College, Anantapur, Andhra Pradesh, India

B. M. Devi

Department of CSE, K.S. R. M College of Engineering, Kadapa, Andhra Pradesh, India

M. Kurni

Department of Computer Science, SoS, GITAM (Deemed to be University), Hyderabad, Telangana, India

D. Samanta · N. P. Joseph

Department of Computer Science, CHRIST Deemed to be University, Bengaluru, India

N. P. Joseph

e-mail: nijup.joseph@christuniversity.in

1 Introduction

Nowadays, overlay routing has been suggested in recent years as a compelling method to accomplish certain routing properties for avoiding the lengthy and tedious procedure of stabilization and global positioning of another routing protocol. It is crucial for network operatives to have efficient monitoring of network performance while constructing unwavering network communication that are hard to deal with disruptions. So as to attain this objective, the monitoring arrangements ought to be capable of recognizing the network mischiefs like abnormally great damage/latency, un-availability, and establishing the localized sources anomalies like failure of appropriate routers in a proper and accurate manner [1].

Having information about risky elements of network exist in the network is practically beneficial for quick recuperation of service facility; for example, the network service provider can relocate services that are concealed. In any case, restricting the elements of networks that effect the disruption of service is challenging process. The direct methodology of examining the strength of unique elements (e.g., by assembling the topology descriptions) is impractical because of deficiency of protocol interoperability like ad hoc networks or providing partial access to internal network nodes [2].

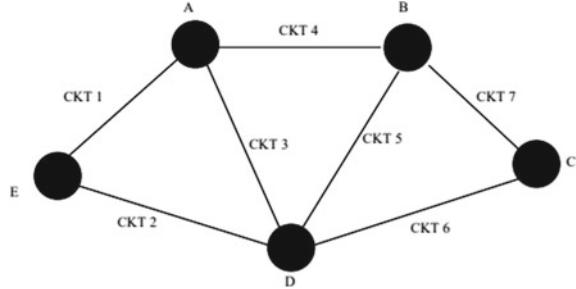
The dynamic system has the subsequent improvement. The current system intends a measure called maximum identifiability to illustrate the potency of network in localized failure as the maximum amount of node failures [3]. We impose a necessary/sufficient condition for restricting the number of localized failures that are relevant to all probing mechanisms. Using the three types of probing mechanisms (CAP, CSP, and UP), we transform the above situations into most tangible conditions.

We reveal that the distinct relationship between the supposed conditions that proposes the maximum/minimum bounds on the index of maximum identifiability as a strategy for providing the boundaries [4]. By using the above method, the bounds are created for polynomial time which are used and computed using CAP and CSP probing mechanisms. To compute in UP probing mechanism, Np-hard method will be considered.

We examine the suggested mechanisms on all three probing mechanisms on sample and actual topologies. The results displays that particularly controllable probing CAP drastically progresses the functionality of localization node failure over uncontrollable probing mechanism [5]. The evaluation demonstrates new perceptions into the distribution of single-node index of maximum identifiability.

2 Literature Survey

Secure routing techniques protect MANET users against assaults and fraudulent data. Along these protocols, information transfer from foundation to objective transpires in numerous paths [6], as well as public keys for data encryption, symmetric keys

Fig. 1 Logical topology

are utilized [7]. The authors of [8] presented a technique in which each sensor node sends to the particulate directions. When a node takes delivery of a message, it confirms to see if it is valid. Even if the data is encrypted, a data replacement assault is unavoidable. The authors of [9] proposed a method for determining numerous paths. Hash functions are used to encrypt the route request messages. Top-k queries are commonly used in disseminated and record systems to extract only the data items that are required from large amounts of data. Authors suggested approaches in [10] that adapt to movement, provide high accuracy, and reduce congestion. The authors of [11] suggested a secure query processing solution for a network with malicious nodes. A method was proposed in [12] in which a rogue node generates fresh and bogus data. Techniques for several standing coordination's are planned in [13]. Every mobile node handles the standing ideals of its neighbors in [14]. Each node calculates its reputation value by studying the communications of its neighbors. Authors presented a mechanism in a reputation system that is resistant to false notification attacks in [15] and [16]. Figure 1 shows logical topology.

This method involves the sender and receiver exchanging a cryptographic key in advance and also delivers their ID in encrypted form, along with their previous and current reputation scores. The received reputation scores can be decoded and confirmed by the receiving node. The major goal is to categorize those nodes that are convolutional nodes that supply to the aggregation computation so that erroneous standing achieves can be discarded in n-networks where aggregation is utilized to acquire the outcome on sensor network. To identify misbehaving nodes and avoid them from contribute in the collective estimate, [17–19] use a secure hierarchical in network aggregation.

3 Proposed System

In the proposed system, efficient algorithms for monitor placement are proposed, allowing for the detection and localization of any single failure. Range tomography not only localizes the failure but also evaluates its severity, which improves the resolution in defining failures. Multiple failure localization is fraught with risk. Boolean network tomography's primary purpose is to decipher this coordination of Boolean

equations. The current research can be divided into two categories: solitary disappointment localization and multiple breakdown localization. Multiple concurrent failures occur with little frequency, according to single failure localization. For the above situation, [20–23] propose effective algorithms for placement of monitors, so that we can detect and localize any single failure. To enhance that characterizing failure resolution, range tomography along with the localizing the failure additionally evaluates its rigorousness (e.g., jamming stage). However, the mechanisms will not consider the point that numerous failures happen regularly than assumptions [24]. Here, the common situation of localizing numerous failures is considered. Intrinsic ambiguity is confronted by the multiple failure localization. Utmost current work reveals this ambiguity through finding the minimal set of network elements which describes the identified node failures. By using this methodology [25], tree topology is suggested for acquiring the network solution that is then enhanced to general topologies. A new Bayesian approach, a dual stage solution, is proposed; the different links failure probabilities are analyzed in the first stage; the maximum likely failure set for subsequent measurements is gathered in second stage.

Algorithm: cal_Met(root_Ca R, fail_Sign_F)

```

1: for (root_Ca r ∈ R) do
2:   r.hit_Rat = |R ∩ F|/|R|;
3:   r.cov_Rat = |R ∩ F|/|F|;
4: end for

```

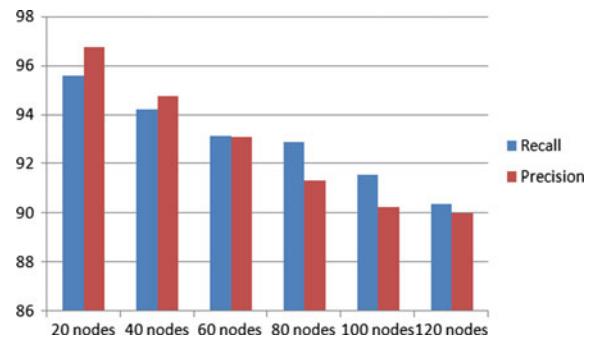
The demerit of the current system, furthermore, the mechanism of integrated monitoring on elements of network cannot identify the complications that are triggered by unexpected communication among network layer, even path may exist between individual network elements, but source to destination connection is interrupted. Network tomography is a simple method for identifying the strength of network elements based on the strength of source to destination communication observed among the factors. It focuses on gathering the characteristics of internal networks based on source to destination measurements from a subset of nodes with monitoring capability, known as monitor. In contrast to traditional measurement, network tomography relies solely on data packets' source to destination performance (e.g., path connectivity) to address issues such as overhead, protocol support, and silent failures [original ref]. This technique is known as Boolean network tomography when the network feature of interest is dual (normal or fizzled). Table 1 expresses the summary of performance.

In this paper, to calculate proficiency of localization node failure, we have used two measures: (1) maximum identifiability bound that describes the greatest number of concurrent failures in a given set of nodes, (2) maximum identifiable set that provides the maximum limit on the concurrent failures, and this denotes the biggest node set, where we can identify the failures that are uniquely localized when the happening of failures is within the limit. Figure 2 shows comparative tabulation with different nodes.

Table 1 Summary of performance

Topology	Recall	Precision
20 nodes	95.63	96.78
40 nodes	94.25	94.787
60 nodes	93.17	93.12
80 nodes	92.89	91.35
100 nodes	91.56	90.25
120 nodes	90.37	90.004

Fig. 2 Comparative tabulation with different nodes



These measures are evaluated in different probing mechanisms on random and real topologies. We have used the naïve Bayesian technique to improve the performance of crucial features of the existing and emerging routing techniques where the naïve Bayesian technique is simple, clear, and active to predict class of the given dataset. Our framework is to compute the given of probing mechanism that identifies and detects the localized network node failure. Figures 3, 4, and 5 represent snapshot of the simulation environment at the original state, intermediate state, and final state, respectively.

Fig. 3 Snapshot of the simulation environment at the original state

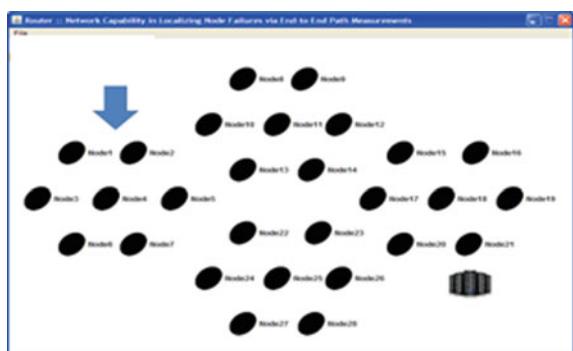


Fig. 4 Snapshot of the simulation environment at the intermediate state

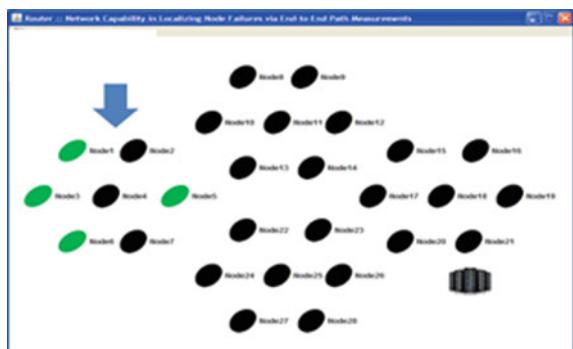
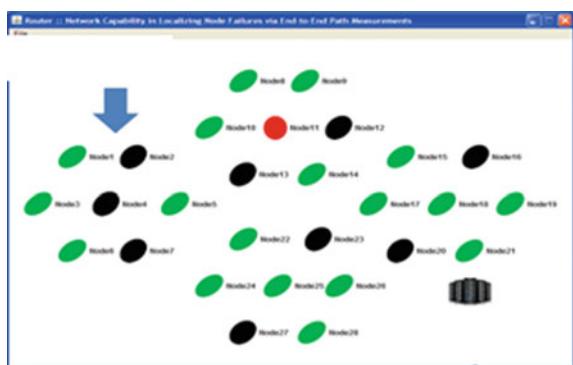


Fig. 5 Snapshot of the simulation environment at the final state



4 Conclusion

The monitors' flexibility is visibly reduced as a result of these probing procedures, and their ability to localize problems is also reduced. However, they also offer increased simplicity of consumption. The stretchiest probing tool, CAP, provides an upper bound on the capacity to locate failures. In classical networks, CAP is possible at the IP layer if all nodes have stringent source routing enabled, 3 or at the application layer if the application supports similar "source routing." Furthermore, CAP is possible under software-defined networking (SDN), an emerging networking paradigm in which monitors can direct the SDN controller to set up subjective channels for probing traffic. UP, on the other hand, is the simplest fundamental probing method that can be used in any communication network and offers a lower bound on the capacity to locate failures, as illustrated. CSP is an intermediate situation that allows routing control while adhering to the basic condition that routes be cycle-free. The "explicit routing" mode of multiprotocol label switching (MPLS) allows one to set up a customizable, non-shortest path using labels as long as the path is cycle-free. CSP can also be deployed over IP networks by installing virtual private networks (VPNs), where the

cycle-free property is also compulsory for choosing pathways between VPN endpoints. The main elements of numerous present and developing routing strategies are captured by these three probing mechanisms. To increase performance, we employed the naive Bayesian technique, which is simple and quick to forecast the test dataset's class assuming the assumption of independence holds. Our goal is to determine how a probing mechanism's flexibility affects the network's ability to locate faults.

References

1. Hagihara R, Shinohara M, Hara T, Nishio S (2009) A message processing method for top-k query for traffic reduction in ad hoc networks. In: Proceedings MDM, pp 11–20
2. Amagata D, Sasaki Y, Hara T, Nishio S (2013) A robust routing method for top-k queries processing in mobile ad hoc networks. In: Proceedings MDM, pp 251–256
3. Liu K, Deng J, Varshney PK, Balakrishnan K (2007) An acknowledgement-based approach for the detection of routing misbehavior in MANETs. IEEE Trans Mobile Comput 6(5):536–550
4. Sasaki Y, Hara T, Nishio S (2011) Two-phase top-k query processing in mobile ad hoc networks. In: Proceedings NBiS, pp 42–49
5. Yu C-M, Ni G-K, Chen I-Y, Gelenbe E, Kuo S-Y (2014) Top-k query result completeness verification in tiered sensor networks. IEEE Trans Inf Forensics Secur 9(1):109–124
6. Zhang R, Shi J, Liu Y, Zhang Y (2010) Verifiable fine-grained top-k queries in tiered sensor networks. In: Proceedings INFOCOM, pp 1–9
7. Chen B, Liang W, Zhou R, Yu JX (2010) Energy-efficient top-k query processing in wireless sensor networks. In: Proceedings CIKM, pp 329–338
8. Lee SJ, Gerla M (2001) Spilt multipath routing with maximally disjoint paths in ad hoc networks. In: Proceedings ICC, vol 10, pp 3201–3205
9. Tsuda T, Komai Y, Sasaki Y, Hara T, Nishio S (2014) Top-k query processing and malicious node identification against data replacement attack in MANETS. In: Proceedings MDM, pp 279–288
10. Samanta D et al Cipher block chaining support vector machine for secured decentralized cloud enabled intelligent IoT architecture. IEEE Access. <https://doi.org/10.1109/ACCESS.2021.3095297> @SCI-Q1
11. Guha A, Samanta D, Banerjee A, Agarwal D (2021) A deep learning model for information loss prevention from multi-page digital documents. IEEE Access 9:80451–80465. <https://doi.org/10.1109/ACCESS.2021.3084841> @SCI-Q1
12. Mekala MS, Patan R, Islam SKH, Samanta D, Mallah GA, Chaudhry SA (2021) DAWM: cost-aware asset claim analysis approach on big data analytic computation model for cloud data centre. Secur Commun Netw. Article ID 6688162, 16 pages. <https://doi.org/10.1155/2021/6688162> @SCI-Q2
13. Samanta D, Karthikeyan MP, Banerjee A, Inokawa H (2021) Tunable graphene nano-patch antenna design for on-chip integrated terahertz detector arrays with potential application in cancer imagining. Nanomedicine, nnm-2020-0386. ISSN:1743-5889. <https://doi.org/10.2217/nmm-2020-0386> @ SCI
14. Biswal AK, Singh D, Pattanayak BK, Samanta D, Yang M-H (2021) IoT-based smart alert system for drowsy driver detection. Wireless Commun Mob Comput. Article ID 6627217, 13 pages. <https://doi.org/10.1155/2021/6627217> @ SCI
15. Maheswari M, Geetha S, Selva Kumar S, Karuppiah M, Samanta D, Park Y PEVRM: probabilistic evolution based version recommendation model for mobile applications. IEEE Access. <https://doi.org/10.1109/ACCESS.2021.3053583> @ SCI
16. Sasaki Y, Hagihara R, Hara T, Shinohara M, Nishio S (2010) A top-k query method by estimating score distribution in mobile ad hoc networks. In: Proceedings DMWPC, pp 944–949

17. Malhotra B, Nascimento MA, Nikolaidis I (2011) Exact top-k queries in wireless sensor networks. *IEEE Trans Knowl Data Eng* 23(10):1513–1525
18. Wu M, Xu J, Tang X, Lee WC (2007) Top-k monitoring in wire-less sensor networks. *IEEE Trans Knowl Data Eng* 19(7):962–976
19. Hu Y-C, Johnson DB, Perrig A (2003) SEAD: secure efficient distance vector routing for mobile wireless ad hoc networks. *Ad Hoc Netw* 1(1):175–192
20. Liu X, Xu J, Lee WC (2010) A cross pruning framework for top-kdata collection in wireless sensor networks. In: Proceedings MDM, pp 157–166
21. Kurosawa S, Nakayama H, Kato N, Jamalipour A, Nemoto Y (2007) Detecting blackhole attack on AODV-based mobile ad hoc networks bydynamic learning method. *Int J Netw Secur* 5(3):338–346
22. Chan H, Perrig A, Song D (2006) Secure hierarchical in-network aggregation in sensor networks. In: Proceedings CCS, pp 278–287
23. Chen S, Zhang Y, Liu Q, Feng J (2012) Dealing with dishonest recommendation: the trials in reputation management court. *Ad Hoc Netw* 10(8):1603–1618
24. Dewan P, Dasgupta P (2010) P2P reputation management using distributed identities and decentralized recommendation chains. *IEEE Trans Knowl Data Eng* 22(7):1000–1013
25. Buchegger S, Le Boudec J-Y (2002) Performance analysis of the CONFIDANT protocol. In: Proceedings MobiHoc, pp 226–236

Rendering View of Kitchen Design Using Autodesk 3Ds Max



Ritwika Das Gupta, Debabrata Samanta^{ID}, and Niju P. Joseph

Abstract The method of creating a 3D kitchen design model is clarified, including setting up the sources, working with editable poly, information in the inside of the kitchen design, and applying turbo-smooth and symmetry modifier. The way materials are introduced to the model which is defined in addition to lighting the environment and setting up the renderer. Rendering methods and procedures are also defined. Multiple images were drawn to create the final rendering. The goal of our research is to produce a kitchen design that uses materials to enhance models. Cylinder, sphere, box, plane, and splines were the shapes employed. Editable poly, editable spline, and UVW map are the modifiers. Finally, we enhanced the model using a material editor and target lighting.

Keywords Autodesk · Render · Turbo smooth · Shell modifier · Bend modifier

1 Introduction

Kitchen design is a part of the food storage model. This refrigerator was designed using a box which is a standard primitive. The box was divided into two segments using connect option of edit poly modifier by connecting edges. The edge of the front face that divided the box's front face was then extruded inward using extrude option of edit poly modifier. All the corner edges and the segmented edge were then chamfered to make the corners smooth. The handles were made by boxes by removing the left and right faces of the box and scaling it appropriately and using turbo smooth to make it smooth and circular corners [1]. The back, front, and the bottom face of the box was removed, and shell modifier was used to give thickness to

R. Das Gupta · D. Samanta (✉) · N. P. Joseph

Department of Computer Science, CHRIST Deemed to be University, Bengaluru, India

R. Das Gupta

e-mail: ritwika.gupta@science.christuniversity.in

N. P. Joseph

e-mail: nijup.joseph@christuniversity.in

the entire rack. A dish was made using a cylinder which is extruded and rescaled the former dish-like structure. This dish was then cloned into several dishes and placed appropriately within the rack. The cooking pan was made using a cylinder, whose top face was extruded and rescaled to make it larger. This process of scaling and extruding continued until the base of the pan was formed [2]. The cap of the pan was again formed by a cylinder, whose top face was extruded and rescaled to make it smaller until a cap-like structure was formed. The drinking glasses and the water bottles are a part of utensils models. The drinking glass was made by a cylinder. The top face of the cylinder was extruded and scaled to make it larger [3, 4].

2 List of Modifier

Modifiers are used to modify any models into its perfect shape and sizes. The major list of modifiers used for designing this model is:

Turbo Smooth—The turbo smooth modifier smoothens any object to give it perfect smooth shape. It has an iteration operation, upon increasing the value of this iteration the amount of smoothness can be increased. In this model, turbo smooth modifier was used for smoothening the glasses, mixer grinder, dishes, spoon, cooking utensils, bottles, and so on [5].

Smooth—The smooth modifier also helps in smoothening any object. Unlike turbo smooth modifier, smooth modifier makes sure that the geometry of the object is kept intact while smoothening it. The auto-smooth option smoothes the object automatically to the threshold amount required and the other numbers options smooth the objects up to a certain level given by the number. In this model, the smooth modifier was used for smoothening the shelves keeping its proper geometry intact.

Edit Poly—The edit poly modifier helps in editing the objects and forming the model. It has various options for editing edges, vertices, faces, and polygon of an object. This modifier has options like chamfer, extrude, bevel cut which was used in designing this model. The chamfer option was used to create smoothness in corners of refrigerator and sharp corner areas of any objects. The extrude option was used in making mixer grinder, glasses, dishes, bottles, utensils, cooking pan, gas stove, microwave oven, shelves, and so on. The cut option was used to create edges and cut faces, vertices or edges and extrude them as per requirement [6].

Shell Modifier—Shell modifier is used to give thickness to any object. It has two major options which are—the inner and outer thickness options. These options help in increasing the inner and outer thicknesses of any object, respectively. In this model, shell modifier was used to increase the thickness of glasses, mixer utensils of mixer grinder, racks, and handles of various objects and wherever required.

Bend Modifier—This modifier is used to bend objects in appropriate directions. There are direction, angle, axis, and limit options used for setting an object's bend

limit and bending it in a particular direction, angle, and axis. To create the cooking stove bend modifier was used [7].

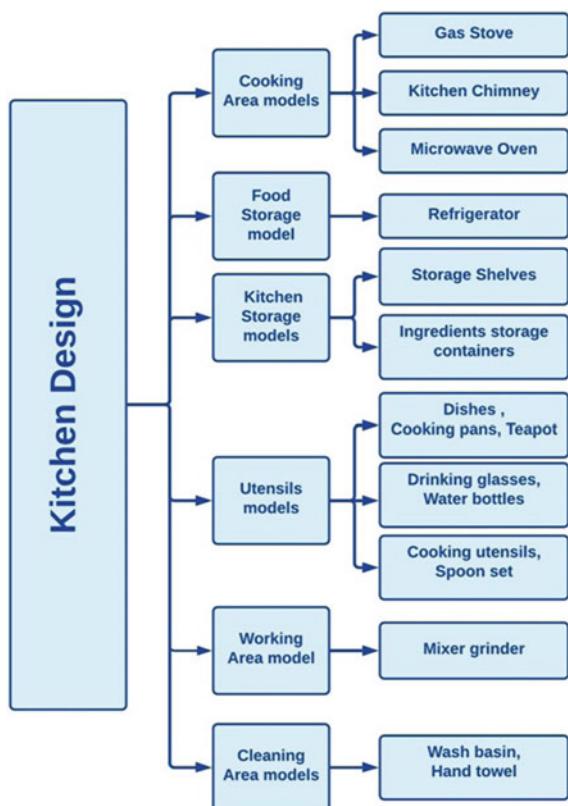
UVW Map—This modifier is used for modifying the textures mapped to an object. It has box, cylinder, sphere, and other options which helps in mapping the object in the form of certain shape. And increasing or decreasing the size of that shape-mapped textures is also possible. It also allows enabling real-world mapping. In this model, most of the textures were used in shelves and as marble on floor, top of shelves and walls are mapped using this modifier [8].

Lathe Modifier—This modifier converts 2D shapes to 3D model. It was used to make the wire of the mixer grinder.

3 Flowchart

Figure 1 represents all the basic model of virtual kitchen design system structure.

Fig. 1 Virtual kitchen design system structure



4 Advantages

There were many tools used to design this model. Mentioning all the major tools, starting with the clone tool helps in copying objects as instance, copy or references. The utensils in kitchen, glasses, and the bottles on racks were created by cloning one single object. The mirror tool is used to create a mirror image of the object [9,10]. For creating cooking utensils and area over gas stove, mirror was used in the design. The snap tool helps to snap any object to other by vertices, edges, midpoints, etc. In kitchen design, snap was used to cut along vertices or edges or midpoints for proper extrude. It was also used to snap the racks inside the top shelf. The major advantage of scale, rotate, and move tool is that it can easily help in transforming a single object into a model. The material editor used helps in editing the material and textures given to each object, with the help of UVW Map [11].

Many modifiers were used to create and modify the objects to form an appropriate model [12]. The major advantages of such modifiers starting with edit poly modifier are that, edit poly modifier was used in kitchen design for extrude, chamfer, bevel, and selection of edges, vertices, faces, and polygon. It was also used for cutting vertices and edges. The turbo smooth and smooth modifiers are smoothing modifiers to make objects smooth. Bend modifier was used to create the area over gas stove by bending a box. Shell modifier was used to give thickness, the mixer; jars and glasses are given thickness by shell. The lathe modifier was used to convert line to wire of the mixer grinder. The UVW map was used to map the textures and edit them in proper manner [13, 14].

5 Result with Discussion

The final result (i.e., the kitchen design) was achieved by using various modifiers and tools. As mentioned in the flowchart, the final result has many models related to a modular kitchen. kitchen design structure express in Fig. 2. These models majorly are:

Refrigerator—For reflection under maps option as reflection material fall off was used and its mix curve pointer was decreased to a certain level from the center, to create a proper reflection. The black shiny material for handles and legs was made using the material editor. For this, ray caster material was used, the glossiness and the specular level was increased to give it appropriate shine. The design on the material is given by an image used as bitmap on diffuse option.

Storage Shelves and Ingredients Storage Containers—This smaller face was then extruded outwards and rescaled down again. This process of extruding and rescaling was continued to create the designs on the shelves. At the bottom center of the shelf, the bottom face was moved upwards using move tool to create the design. At the bottom front face of the shelf, the edges were connected in such a way that it gets



Fig. 2 Render view kitchen design structure

divided into two segments, one smaller segment at the top and one larger segment below. The edges of these segments were further connected using connect option of edit poly modifier to create a square face. The top face of the box was extruded upwards, and each division on the box was connected by edges to create different square faces according to the design. These square-faced edges were then extruded inwards an outwards and scaled appropriately to create shelf and drawer designs. A sphere was used to create the knobs of the shelves; this fear was rescaled according to the knob size and placed in correct position for creating knobs of the shelves. The wood material and marble material were created as a standard material using the material editor tool. These materials were images of a wood texture and marble texture which were added to the diffuse option present in the material editor tool. These materials were then assigned to each face of the shelves and properly aligned using UVW map modifier.

Mixer Grinder—The mixer grinder (Fig. 3) is a part of working area model. The cap of the utensil was again made by a cylinder which was extruded and resized to form

Fig. 3 Mixer grinder design structure



of gap-like structure. The speed regulator off the mixer grinder was made using a cylinder, which was extruded and rescaled to form the base of the regulator and then cut option was used from edit Poly to cut from one vertex to another to create a holder of the regulators. This face which was cut was then extruded to form the holder of the regulator. Different text shapes such as inch, off, Roman numbers were used to create the text around the regulator. Then loft option was used which is the option present in compound objects, to give thickness to the text around the regulator. The switchboard was made using a box, which was divided into different segments using connect option of edit poly modifier an extruded outwards and rescaled to form a board-like structure. The top face of the board was then divided into five segments using connect option, and square edges of each division were extruded invert to create columns for switches [15, 16]. Each division was then further divided into two small square faces using connect option, and those square edges were extruded inwards. The shiny gray and shiny white materials were also made by ray caster material in material editor by increasing the glossiness and specular level. For reflection under maps option as reflection material fall off was used and its mix curve pointer was decreased to a certain level from the center, to create a proper reflection. The black shiny material for base of the mixer was made using the material editor. For this ray caster material was used, the glossiness and the specular level was increased to give it appropriate shine. All these materials were mapped to their respective objects.

Cooking Utensils, Spoon Set—These spoons were mirrored and cloned and align into correct position. The jar of the spoon set (Fig. 4) was made by extruding a cylinder and rescaling them accordingly so that it takes a jar shape. The top face of the jar was removed, and shell modifier was used to give thickness to the jar. The shiny gray (steel material) was also made by ray caster material in material editor by increasing the glossiness and specular level. For reflection under maps option as reflection material fall off was used, and its mix curve pointer was decreased to a certain level from the center, to create a proper reflection. The black shiny material for handles of the utensils was made using the material editor. For this, ray caster material was used, the glossiness and the specular level was increased to give it appropriate shine. All these materials were mapped to their respective objects.

Washbasin and Hand Towel—All these objects were placed in a proper position and grouped together as a tap. The hand towel was made using a plane which was extruded as required. This extruded plane was then segmented into many segments.

Fig. 4 Spoon stand design structure





Fig. 5 Drinking glasses design structure

Each edge of the segments was moved outwards and inwards alternatively. Then each edge was extruded inwards and outwards alternatively. All these materials were mapped to their respective objects.

Dishes, Cooking Pans, and Teapot—The side handles after cooking pan were made by a box, which's back top and bottom faces were removed and thickness was provided by shell modifier. The top lid cap was made by an oil tank. All these objects were grouped together and cloned to form two cooking pans. The teapot is a standard primitive which was placed in the design. Different colors were given to the object using the material editor tool and providing different color materials with slight shine onto the colors. The shine was given by using ray caster material and by increasing the glossiness and specular level. All these materials were mapped to their respective objects. Figure 5 shows drinking glasses design structure.

Drinking Glasses, Water Bottles—This glass material was then assigned to the glasses. The water bottles were made by cylinder. By extruding the cylinder and rescaling to make it larger for the base of the water bottle. This process of extruding and rescaling continued until a bottle structure was formed. The shine was given by using ray caster material and by increasing the glossiness and specular level. All these materials were mapped to their respective objects.

Microwave Oven—The shiny gray (steel material) and shiny white materials were also made by ray caster material in material editor by increasing the glossiness and specular level. For reflection under maps option as reflection material fall off was used and its mix curve pointer was decreased to a certain level from the center, to create a proper reflection. The shiny gray and shiny white material was then used for handle and button, respectively. The glass material used at the central box was made by ray caster material in material editor by increasing the glossiness, specular level and transparency to give a glass look. For mapping different textures to different

faces of the microwave oven, multi-sub/object was used so that different faces can be provided with different ids and each id can constitute a different material in the material editor.

Gas Stove—The materials used in the entire cooking stove were: the black shiny material for the base of the gas stove which was made using the material editor. For this ray caster material was used, the glossiness and the specular level was increased to give it appropriate shine. The shiny gray and shiny white materials were also made by ray caster material in material editor by increasing the glossiness and specular level.

Kitchen Chimney—The kitchen chimney is a part of the cooking area models. This chimney was made by a cylinder and a cone. In the lower part, the lowest face of the cone was removed and shell modifier was used to increase the thickness of the cone. The shiny gray material was made by ray caster material in material editor by increasing the glossiness and specular level. For reflection under maps option as reflection material fall off was used and its mix curve pointer was decreased to a certain level from the center, to create a proper reflection. This material was then used for the chimney.

6 Conclusion

In the UVW map modifier box option was used to align the wood texture. The glass material used at the shelves was made by ray caster material in material editor by increasing the glossiness, specular level, and transparency to give a glass look. For reflection under maps option as reflection material fall off was used and its mix curve pointer was decreased to a certain level from the center, to create a proper reflection. For mapping different textures to different faces of the shelves, multi-sub/object was used so that different faces can be provided with different ids and each id can constitute a different material in the material editor. To make this container smooth, turbo smooth modifier was used after grouping the entire container. This grouped container was cloned several times to make many such containers. Then the entire rack of container was cloned to make many racks and placed in its perfect position. These racks were mapped with a wood texture, each container was mapped with the silver texture, and the caps of the containers were well-mapped with shiny black texture.

References

1. Arayici Y, Hamilton A (2005) Modeling 3D scanned data to visualize the built environment. In: Ninth international conference on information visualisation (IV'05), p 509–514. ISSN: 2375-0138

2. Castellani U, Fusillo A, Murino V, Papaleo L, Puppo E, Pittore M (2005) A complete system for on-line 3D modelling from acoustic images. *Signal Process Image Commun* 20(9–10):832–852
3. De Luca L, Veron P, Florenzano M (2006) Reverse engineering of architectural buildings based on a hybrid modeling approach. *Comput Graph* 30(2):160–176
4. Samanta GSD, Paul M (2011) Segmentation technique of sar imagery using entropy. *Int J Comput Technol Appl* 2:1548–1551
5. Esteban CH, Schmitt F (2003) Silhouette and stereo fusion for 3D object modeling. In: Fourth international conference on 3-D digital imaging and modeling. 3DIM 2003. Proceedings, pp 46–53
6. Guha A, Samanta D, Banerjee A, Agarwal D (2021) A deep learning model for information loss prevention from multi-page digital documents. *IEEE Access* 9:80451–80465
7. Lian Q, Li D-C, Tang Y-P, Zhang Y-R (2006) Computer modeling approach for a novel internal architecture of artificial bone. *Comput Aided Des* 38(5):507–514
8. Maheswari M, Geetha S, Selva Kumar S, Karuppiah M, Samanta D, Park Y (2021) PEVRM: probabilistic evolution based version recommendation model for mobile applications. *IEEE Access* 9:20819–20827
9. Park S-Y, Subbarao M (2005) A multiview 3D modeling system based on stereo vision techniques. *Mach Vis Appl* 16(3):148–156
10. Sainz M, Pajarola R, Mercade A, Susin A (2004) A simple approach for point-based object capturing and rendering. *IEEE Comput Graph Appl* 24(4):24–33
11. Samanta D, Sanyal G (2012) Segmentation technique of sar imagery based on fuzzy c-means clustering, pp 610–612
12. Sun W, Starly B, Nam J, Darling A (2005) Bio-CAD modeling and its applications in computer-aided tissue engineering. *Comput Aided Des* 37(11):1097–1114
13. Samanta D, Sanyal G (2012) A novel approach of sar image classification using color space clustering and watersheds, pp 237–240
14. Kureethara V, Biswas J, Samanta D, Eapen NG Balanced constrained partitioning of distinct objects. *Int J Innov Technol Exploring Eng.* ISSN: 2278-3075(Online). <https://doi.org/10.35940/ijitee.K1023.09811S19>
15. Biswal AK, Singh D, Pattanayak BK, Samanta D, Yang M-H (2021) IoT-based smart alert system for drowsy driver detection. *Wireless Commun Mobile Comput.* Article ID 6627217, p 13. <https://doi.org/10.1155/2021/6627217>
16. Manu MK, Roy S, Samanta D (2018) Effects of liver cancer drugs on cellular energy metabolism in hepatocellular carcinoma cells. *Int J Pharm Res* 10(3). ISSN-0975-2366. <https://doi.org/10.31838/ijpr/2018.10.03.079>

Preserving Security and Privacy in IoT Using Machine Learning and Trust Management



Avinash Kumar, Trisha Bhowmik, Rohit Sharma, and Abhishek Bhardwaj

Abstract The digital era has given rise to various services, which are provided to the user in automation. The Internet of things (IoT) has now been implemented in all walks of life, starting from smart devices to intelligent offices and from smart homes to smart industrial systems. These industries drive huge data and hence attract hackers to exploit for personal gain. This paper presents various threats and adequate solutions that can make IoT more reliable for users. The paper has also covered the most recent work done in this field related to security and privacy. Also, it covers the vital future research directions that could make IoT safer systems to be used by users.

Keywords IoT · SIoT · Cyber-attacks · RFID · ARP · Deep brief network

1 Introduction

The Internet of things (IoT) deals with the communication of interconnected devices via the internet and work together for dedicated tasks. The IoT has laid the foundation for various smart applications and smart systems. These have been covered in below subsections.

1.1 Internet of Things (IoT)

Internet of things (IoT) is a collection of interrelated objects, applications, services, and people. By communicating and sharing data and information across areas and

A. Kumar (✉) · T. Bhowmik · A. Bhardwaj
School of Engineering and Technology (SET), Sharda University, Noida, India

R. Sharma
Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India

applications, these objects can achieve common goals. Many different areas of IoT applications are being implemented today, such as healthcare, transportation, energy production, agriculture, and distribution. Identical and heterogeneous devices within an Internet of things are identified with identity management. IP addresses can also denote IoT regions, yet each entity in each region has its IP address. Intelligent devices around us perform tasks and chores for us by making them do tasks for us automatically. That is what the Internet of things is all about. IoT is a term that is used to describe smart homes, cities, transportation, and infrastructure. IoT is widely applied at both the personal and the enterprise level [1].

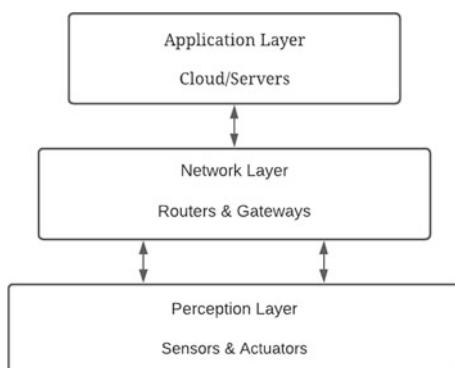
1.2 IoT Architecture

IoT manifests itself as layers defined by their functions and their devices. IoT is considered to have different layers based on different opinions. As described by several researchers [2–4], the IoT is primarily composed of three layers called perception, network, and application. Security issues are inherent to every layer of the IoT. Figure 1 illustrates the three-layered architecture of IoT.

1.2.1 Perception Layer

In the IoT, this layer is sometimes referred to as the “sensors” layer. Sensors and actuators are used in this layer to obtain environmental data. Information is collected, processed, and transmitted from this layer to the network layer. The IoT nodes in this layer can also collaborate over local networks and short-range networks [3].

Fig. 1 Three-layered architecture of IoT



1.2.2 Network Layer

The Internet of things network layer acts as the data flow mechanism that traverses the Internet from various devices and hubs to the final destination. Network gateways act as the intermediary between the different IoT devices by aggregating, filtering, and transmitting data to and from different sensors [4].

1.2.3 Application Layer

In this layer, data are authenticated, integrity is maintained, and confidentiality is guaranteed. This layer determines the formation of smart environment.

There are various researches that have been carried out for securing IoT devices and IoT systems. Moreover, to the best of our knowledge, there has been no past work done to accumulate the vital aspects that could guard the IoT system from cyber-attacks. This paper fills the gap of using different vital technology to secure IoT and produces state-of-the-art solutions. In summary, the major contributions of this paper are as follows.

- This work presents the various threat vectors that are rising in IoT, which were previously existing on a traditional client–server system.
- This work presents the most vital IoT attack model that could help to analyze the IoT system and its environment.
- This work scrutinizes various metrics that could enhance customer reliability for IoT-based system.
- This work presents various approaches that could be used in IoT to make it more defensive against cyber-attacks.

The remainder of the paper is organized as follows: Sect. 2 represents the effects arising in IoT related to security and privacy. Section 3 describes the vital IoT attack model. Section 4 presents the most probable solution for securing the IoT system. Finally, identification of future research directions has been covered in Sect. 5 followed by a conclusion in Sect. 5.

2 IoT Characteristic Leading to Security and Privacy Issues

The IoT being heterogeneous affects the security and privacy of the system to a greater extent. The devices have to deal with various challenges and threats, which have been discussed in the below subsections.

2.1 *Interdependence*

The interdependency is one of the significant issues in IoT security, and its impact is covered in the proceeding sections.

2.1.1 Description

It is becoming easier for IoT devices to interact with each other as the technology evolves, and guiding humans is unnecessary. Devices that connect to the IoT are not necessarily communicating explicitly like computers and smartphones. It also allows connectivity between many devices and implicit control of their behavior or environmental circumstances using smart guidelines in the cloud, implemented via IFTTT [5], widely adopted in IoT platforms, e.g., SmartThings of Samsung [6], HomeKit of Apple [7], AWS IoT of Amazon [8].

2.1.2 Threats

It may not be possible to compromise the target device itself, but attackers can manipulate surrounding devices and the environment interdependent with the target device. This may mean that attackers may use this feature maliciously to target the target devices with a low difficulty level and bypass the original defense mechanism.

2.1.3 Challenges

IoT security is posed by interdependent behaviors, which most researchers do not appreciate. A single device is generally protected by researchers. A defensive boundary can be established for IoT devices, but it is not easy to apply static access control and privilege management methods due to their interdependent features. It is also difficult to specify a particular set of fine-grained permissions as other devices or environmental conditions could change their behaviors. Thus, overprivileged applications on existing IoT platforms have become a common problem [9].

2.1.4 Solutions and Opportunities

It was realized very early that cross-device interdependencies existed, and a set of new security plans were proposed for detecting anomalous behavior [10]. The increasing number of devices will complicate and make these policies impractical.

2.2 *Diversity*

The diversity is a very important part in IoT. It is a combination of different technologies and devices. The below subsections deeply explain various parameters.

2.2.1 **Description**

Unlike traditional IoT devices, heterogeneous IoT products are designed for particular tasks and interact closely with their physical environments to meet the needs of various applications. Therefore, they require specific hardware, software, and processes. Furthermore, there is also a difference in communication protocols needed for different application scenarios.

2.2.2 **Threats**

In spite of new IoT devices being built with inadequate security checks beforehand, the Ali mobile security team at [11] found that above 90% of IoT firmware themselves have safety vulnerabilities, such as key hardcoded and common vulnerabilities linked to web browsers, which hackers could exploit. New protocols usually suffer from many security problems because practical experience is lacking for new IoT functions, such as bootstrapping IoT devices [12].

2.2.3 **Challenges**

The challenge of securing heterogeneous IoT devices is a problem due to the diversity of the devices [13], especially for industries. The IoT devices present so many security vulnerabilities that a new understanding of how to detect and remediate them is imperative. Researchers should investigate the general security concerns of every protocol, as every protocol differs from others in a few ways. This is also important from the network security point.

2.2.4 **Solutions and Opportunities**

Researchers investigated additional IoT devices' firmware and source code to uncover and fix potential vulnerabilities [14]. An embedded system's firmware can be analyzed dynamically using a framework created by Dovgalyuk et al. [15]. However, the emulator cannot simulate every action taken by the real devices and needs to communicate those actions through a physical connection with the real devices. Therefore, it cannot be used for analyzing large-scale firmware.

2.3 *Constrained*

The IoT has various constrained because of its complexity. These arise due to heterogeneity and these are discussed in below subsections.

2.3.1 Description

Many IoT devices have been designed to be light and small due to cost and physical limitations. This applies especially to industrial sensors and implantable medical devices. They, therefore, have the relatively low computing power and storage capacity compared to conventional computers or mobile phones. The military, industrial, and agricultural markets use devices that must operate in environments without charging, so they consume large amounts of power.

2.3.2 Threats

IoT devices are unable to deploy critical system and network defenses due to their constrained feature sets. Lightweight IoT devices, on the other hand, do not have a memory management unit (MMU). These devices cannot be isolated from memory and have a randomized address space layout (ASLR) and other memory safety features. These devices cannot also be used to implement complex encryption and authentication algorithms.

2.3.3 Challenges

Researchers face a significant challenge when it comes to protecting lightweight IoT devices with less system software and hardware resources. As a result, time and power constraints must also be met by such system protections in practical applications. Additionally, researchers have difficulty deploying complex encryption and authentication algorithms on tiny IoT devices due to their low latency and computing power.

2.3.4 Solutions and Opportunities

Prior studies have focused on designing lightweight security mechanisms for constrained IoT devices to enhance system security. However, most of them are unable to satisfy both the operational and security requirements. ARMor [16] is a lightweight software fault isolation solution that can protect small embedded processors from critical application code errors. Although this caused some programs to

run slower, it had a high-performance overhead since they had to check addresses often.

2.4 *Myriad*

This component decides various important elements that relate with the challenges in IoT. The below subsections cover these aspects.

2.4.1 Description

The IoT devices that are rapidly proliferating will generate enormous amounts of data that will be transmitted, used, and transmitted again. IoT devices and data consist of an enormous amount of data, which we describe as an IoT feature.

2.4.2 Threats

The Mirai botnet assassinated more than 1 million IoT devices in 2016, resulting in the most significant cyberattack ever with attack traffic exceeding 1 Tbps [17]. Also, malware distribution botnets like IoTroop [17] increasingly exploit unsecured devices such as IoT rather than computers or smartphones. Several of these can be used to attempt large-scale denial of service (DDoS) attacks due to their speed of spread.

2.4.3 Challenges

IoT devices are generally not protected against intrusion and do not have anti-virus software to detect malware. Furthermore, IoT devices have a limited power supply and computing power, as previously discussed. Detection and resistance to IoT botnet viruses are therefore challenged for researchers. While botnet proliferation cannot be prevented, it can be prevented to a certain degree.

Solutions and Opportunities: Mirai's characteristics have been analyzed by many researchers in order to detect IoT botnets. Jenkins et al. [18] designed a tool that can detect potential vulnerabilities in IoT systems with several attack vectors gleaned from the Mirai botnet.

3 IoT Attack Model

The attack model helps to analyze various attacks that could take place in IoT. These have been covered in below subsections.

3.1 *DoS Attackers*

IoT devices cannot obtain services because of the intrusions that flood the target server with too many requests [19]. In a DoS attack, one of the most dangerous techniques involves using thousands of Internet protocol addresses to request IoT services, making it difficult for the server to identify real IoT devices from attackers. Attacks on distributed IoT systems that utilize lightweight security protocols are particularly common. There is more than one unique IP address involved in DDoS attacks. Many hosts are infected with malware, so there are potentially thousands of infected computers. An attack on a network with more than 3–5 nodes is typically referred to as a distributed denial of service, while one with fewer nodes might be referred to as a DoS attack. There can be multiple attack machines that generate more attack traffic than one. Numerous attack machines can be harder to disable and eliminate than one attack machine. Each attack machine will exhibit stealthier behavior, making it difficult to track and eliminate.

3.2 *Jamming*

Incoming attacks disrupt the ongoing radio transmissions of IoT devices and exacerbate their failed communication efforts that further drain the resources of IoT devices or sensors (bandwidth, energy, CPU, memory) [20]. A jammer attacks wireless networks as part of a denial-of-service (DoS) campaign. A jammed wireless network is caused by RF frequencies that interfere with the network's operation. Normally jamming does not result from malicious intent but can be caused by other wireless devices operating on the same frequency of the wireless network. Using spectrum analysis, hackers can use denial-of-service (DoS) jamming to disrupt wireless networks by transmitting massive signals to interfere with the networks' communication. It is also possible for attackers to interfere with radar operation by intentionally emitting radio frequency signals during jamming attacks.

3.3 Spoofing

Spoofing nodes impersonate legal IoT devices by imitating their identity along with their medium access control (MAC) addresses and RFID identifiers and then enables them to launch DoS attacks and man-in-the-middle attacks [21]. In particular, man-in-the-middle attacks against computer networks can be carried out by using Internet protocol (IP) spoofing and address resolution protocol (ARP) spoofing. In the TCP/IP suite of protocols, there is often no method to authenticate the source or destination of a message, leaving the applications vulnerable to spoofing attacks when applications take no additional precautions. In addition to firewalls that perform deep packet inspection, spoofing attacks using TCP/IP suite protocols may be moderated by measures that verify the identity of a sender and receiver. Basically, spoofing refers to the practice of someone or something pretending to be different in order to gain our confidence, gain access to our systems, steal money, or propagate malware.

3.4 Malware

Malware is malicious code that is crafted intelligently to attack the system. The malware's main objective is to encrypt the target in order to victimize it. The malware could either encrypt the system, or it can encrypt the documents and other resources. The malware has increased in terms of both the number of attacks and severity. The malware attacks from various means and variants such as botnets, crypto-jacking, malvertising. The last few years have seen the evolved form of malware that is ransomware, which not only encrypts the system but also demands ransom or money in return for decrypting the entire system. WannaCry, Petya are some of the most dangerous malware that can attack any system.

3.5 Eavesdropping

Eavesdropping is an action in which a third party intercepts the communication of two parties. In the world of cybersecurity, the term has also come to hold much significance since the beginning of the digital age. Data drives the Internet. Thousands of transactions are completed each day using a digital medium that involves inputting personal information into a Web site, whether it is for a purchase, a loan application, or any other purpose. Data from Web sites can be spied on through digital network eavesdropping. Insecure networks are compromised through the use of special programs built by hackers to record sensitive data communications. Once the info-packets are gathered, they are analyzed using advanced cryptographic tools, or they can simply be listened to or read in hopes of finding valuable information. Table 1 gives a summary of the mechanism for detecting and protecting.

Table 1 Mechanism for detecting and protecting

Heading level	Example	Concepts used
DoS attackers	Securing IoT	Neural networks [4] Access management [5]
Jamming	Securing IoT	Q-learning [16]
Spoofing	Authentication IoT	Dyna-Q [8]
Malware	Detecting malware	K-neural networks [19]
Eavesdropping	Authentication of user	Bayesian [21]

4 Securing IoT Through Multi-technological Concepts

The IoT system's main backbone is the Internet, which helps them to communicate with interconnected devices. Therefore, the communicating channel must perform various trust management for user authentication mechanisms as well as secure the routing of the packets.

4.1 *Intrusion Detection Using ML*

The power of machine learning (ML) is to learn the sample and then produce fruitful decisions based upon the learned data set. The IoT, being heterogeneous, becomes vulnerable to attacks. The IoT attack model can provide useful solutions through all three learning approaches: supervised, unsupervised, and reinforcement learning [22]. He et al. [23] suggest the use of ML, as well as deep learning (DL) in IoT, which increases the level of security. These two approaches are efficient as the system learns the threats and accordingly defends the system. IoT needs to be defended at the network level also, and hence, it focuses on various security devices, among which intrusion detection system (IDS) and tool for learning data set [24]. The use of ML is used for detecting intruders inside the network [25].

4.2 *Trust Management*

Trust management is yet another powerful mechanism that should be incorporated in IoT for securing the channel. The mobile ad hoc networks (MANETS) and vehicular ad hoc networks (VANET) are sometimes used in the IoT for communication. These communicating technologies, such as MANET and VANET, are secured using a trust management scheme and are very efficient. Therefore, implementing trust management (TM) would be an edge for securing IoT systems and IoT devices. Wireless sensor network (WSN) security is managed using TM as one of the concepts.

The TM has been analyzed as one of the practical uses for securing the WSN. Hence, the IoT system that is using WSN could be guarded against WSN network-based cyber-attacks. The TM-based has been successfully implemented for small IoT systems.

4.3 Securing IoT by Securing the Routing Protocol

The routing protocol is the driving unit used for the transition of a packet from one to another. In an IoT, the network also works on the routing protocol. Cakir et al. [26] performed an attack on IoT with nodes ranging from 10 to 1000, and the proposed method was able to find the deep attack inside the network. So, the routing protocol, once secured in an IoT, can reduce cyber-attacks. The protocol security has high accuracy for threat detection when this security model is used [26]. The other way is to protect routing packets by either detecting infected or malicious nodes. The use of a deep brief network (DBN) is very effective in finding both infected nodes or malicious nodes by analyzing anomalies in the IoT environment [27]. The anomaly-based detection for the IoT environment is crucial as it helps identify the attack, which is crucial for removing the attack vector. The attacks common in routing path are sinkhole, wormhole, and selective forwarding, which could be easily detected using optimum path forest (OPF), which work upon the MapReduce approach [28]. In dealing with removing malicious code and malicious nodes for a network, the most important part is finding that infected part. Hence, these approaches will enhance the security of the IoT system.

4.4 Trust Metric to Secure the IoT Environment

Trust metrics (TM) helps IoT to provide quality of service (QoS). The TM uses various standards and patterns with adequate monitoring of various services executing on the IoT environment. Trust metrics is very useful in achieving accurate calculation and precise evaluation [28]. The measurement highly depends upon the type of devices that are being used in the system. The QoS is used for analyzing the IoT environment efficiency. When the IoT system becomes more specialized like social IoT (IoT), elements such as honesty, collaborations, and community interest become vital assets. These trust metrics decide the security and privacy in the IoT system. If more data transaction is required, the requirement for honesty metrics increases. The nodes need to cooperate with other nodes via their social relationship. Router behavior is another vital trust metric that enhances IoT security where routing of the data and detecting routing path plays a vital role. The behavior of routing is monitored to identify any anomaly in the system.

5 Conclusion and Future Research Direction

The IoT is burgeoning implementation-wise in smart cities, smart healthcare, smart offices, and many others from either a single unit or a system. The paper tried its best to enumerate all the vital security and privacy issues emerging in IoT. It then focuses on various IoT security models and proceeds with the probable solution using modern technologies.

The IoT is a heterogeneous system where the devices are from different manufacturing companies. These devices may differ in their security policies, and some even do not have patch update features required to increase existing devices' security. Hence, it is very challenging to provide one solution that can suit every device used in the IoT. Moreover, the field sensors (FSs) that are an essential part of any IoT device or IoT system do not have any security features. It is hard to detect if the fake sensor is replaced by the real one by the adversary. These security failures need special attention in order to make IoT more reliable, secure, and efficient for the users. The paper outlines itself with its contributions and future research directions.

References

1. Bhushan B, Sahoo C, Sinha P, Khamparia A (2020) Unification of blockchain and internet of things (BIoT): requirements, working model, challenges and future directions. *Wireless Netw.* <https://doi.org/10.1007/s11276-020-02445-6>
2. Haque AK, Bhushan B, Dhiman G (2021) Conceptualizing smart city applications: requirements, architecture, security issues, and emerging trends. *Expert Syst.* <https://doi.org/10.1111/exsy.12753>
3. Miori, V., & Russo, D. (2017). Improving life quality for the elderly through the social Internet of Things (SIoT). 2017 Global Internet of Things Summit (GIoTS). <https://doi.org/10.1109/giots.2017.8016215>
4. Li S (2017) Security architecture in the Internet of things. *Secur Internet of Things* 27–48. <https://doi.org/10.1016/b978-0-12-804458-2.00002-0>
5. Xu R, Zeng Q, Zhu L, Chi H, Du X, Guizani M (2019) Privacy leakage in smart homes and its mitigation: IFTTT as a case study. *IEEE Access* 7:63457–63471. <https://doi.org/10.1109/acc.ess.2019.2911202>
6. Karimi K, Krit S (2019) Internet of thing for smart home system using web services and android application. *Smart Netw Inspired Paradigm Approaches IoT Appl* 191–202. https://doi.org/10.1007/978-981-13-8614-5_12
7. Feiler J (2016) Exploring the HomeKit world as a developer, designer, or device manufacturer. *Learn Apple HomeKit on IOS* 73–87. https://doi.org/10.1007/978-1-4842-1527-2_6
8. Rizvi JR, Killough S, Cherry B, Gowda S (2018) Lessons learned and cost analysis of hosting a full stack open data cube (ODC) application on the Amazon web Services (AWS). In: IGARSS 2018—2018 IEEE international geoscience and remote sensing symposium. <https://doi.org/10.1109/igarss.2018.8518084>
9. Katsinoulas L, Papoutsidakis M, Tseles D (2017) Smart home applications for energy saving and increased security. *Int J Comput Appl* 175(8):38–44. <https://doi.org/10.5120/ijca2017915650>
10. Payne BR, Abegaz TT (2017) Securing the Internet of things: Best practices for deploying iot devices. *Comput Netw Secur Essentials*, 493–506. https://doi.org/10.1007/978-3-319-58424-9_28

11. Showail J (2021) Internet of things security and privacy. Internet of Things [Working Title]. <https://doi.org/10.5772/intechopen.96669>
12. Sanchez-Gomez J, Garcia-Carrillo D, Marin-Perez R, Sanchez-Iborra R, Gomez AF (2020) Secure bootstrapping and header compression for iot constrained networks. In: 2020 global internet of things summit (GloTS). <https://doi.org/10.1109/giots49054.2020.9119644>
13. Lin C-W, Sangiovanni-Vincentelli A (2017) Security threats in cyber-physical systems. Secur-Aware Design Cyber-Phys Syst 5–8. https://doi.org/10.1007/978-3-319-51328-7_2
14. Busch M, Dirsch K (2020) Finding 1-day vulnerabilities in trusted applications using selective symbolic execution. In: Proceedings 2020 workshop on binary analysis research. <https://doi.org/10.14722/bar.2020.23014>
15. Dovgalyuk P, Fursova N, Vasiliev I, Makarov V (2018) Work-in-progress: introspection of the linux-based embedded firmwares. In: 2018 international conference on embedded software (EMSOFT). <https://doi.org/10.1109/emsoft.2018.8537186>
16. Besson F, Blazy S, Dang A, Jensen T, Wilke P (2019) Compiling sandboxes: formally verified software fault isolation. Program Lang Syst 499–524. https://doi.org/10.1007/978-3-030-17184-1_18
17. Dhayal H, Kumar J (2018) Botnet and p2p botnet detection strategies: a review. In: 2018 international conference on communication and signal processing (ICCSP). <https://doi.org/10.1109/iccsp.2018.8524529>
18. Kolias C, Kambourakis G, Stavrou A, Voas J (2017) DDoS in the IOT: Mirai and OTHER BOTNETS. Computer 50(7):80–84. <https://doi.org/10.1109/mc.2017.201>
19. Bhushan B, Sahoo G (2017) Recent advances in attacks, technical challenges, vulnerabilities and their countermeasures in wireless sensor networks. Wireless Pers Commun 98(2):2037–2077. <https://doi.org/10.1007/s11277-017-4962-0>
20. Jaitly S, Malhotra H, Bhushan B (2017) Security vulnerabilities and countermeasures against jamming attacks in wireless sensor networks: a survey. In: 2017 international conference on computer, communications and electronics (Comptelix). <https://doi.org/10.1109/comptelix.2017.8004033>
21. Bhushan B, Sahoo G (2019) \$\$E^2\{2\} SR^2\{2\} E 2 S R 2: an acknowledgement-based mobile sink routing protocol with rechargeable sensors for wireless sensor networks. Wireless Netw 25(5):2697–2721. <https://doi.org/10.1007/s11276-019-01988-7>
22. Xiao L, Wan X, Lu X, Zhang Y, Wu D (2018) Iot security techniques based on machine learning: how do iot devices use ai to enhance security? IEEE Signal Process Mag 35(5):41–49. <https://doi.org/10.1109/msp.2018.2825478>
23. Bhushan B, Sahoo G (2019) ISFC-BLS (intelligent and secured fuzzy clustering algorithm using balanced load sub-cluster formation) in WSN environment. Wireless Pers Commun. <https://doi.org/10.1007/s11277-019-06948-0>
24. Chaabouni N, Mosbah M, Zemmari A, Sauvignac C, Faruki P (2019) Network intrusion detection for iot security based on learning techniques. IEEE Commun Surv Tutor 21(3):2671–2701. <https://doi.org/10.1109/comst.2019.2896380>
25. Jha K, Anumoti S, Pronika, Soni K (2021) Security issues and architecture of iot. In: 2021 International conference on artificial intelligence and smart systems (ICAIS). <https://doi.org/10.1109/icais50930.2021.9395962>
26. Cakir S, Toklu S, Yalcin N (2020) Rpl attack detection and prevention in the Internet of things networks using a gru based deep learning. IEEE Access 8:183678–183689. <https://doi.org/10.1109/access.2020.3029191>
27. Ge M, Syed NF, Fu X, Baig Z, Robles-Kelly A (2021) Towards a deep learning-driven intrusion detection approach for Internet of things. Comput Netw 186:107784. <https://doi.org/10.1016/j.comnet.2020.107784>
28. Diro AA, Chilamkurti N (2018) Distributed attack detection scheme using deep learning approach for Internet of things. Futur Gener Comput Syst 82:761–768. <https://doi.org/10.1016/j.future.2017.08.043>

Design and Implementation Wireless Sensor Node with Security Algorithm Based on Microcontroller ESP8266



Zahraa A. Msekh and Alyaa A. Msekh

Abstract Due to the vast development in wireless sensor networks, it has become necessary to preserve the data sent over the Internet and shared networks by using security algorithms. There are many encryption algorithms, and one of the most secure algorithms is the advanced encryption standard (AES algorithm). AES is a robust security algorithm recommended by National Institute for Standardization and Technology (NIST). AES has key space equal to (10^{128}) enough to resist brute force attacks. In this work, we design and implement wireless sensor networks containing four sensor nodes for object tracking with (AES) algorithm to encrypt object locations. Each sensor node will use a microcontroller ESP8266 board and GPS. When the client calls with the sensor node, the GPS will specify the longitude and latitude, and the ESP8266 microcontroller will make the AES encryption process for the specified site. The encryption data send in the wireless channel between a sensor node and client by the protocol TCP/IP in real time.

Keywords Wireless sensor networks · AES algorithm · Encryption · Security · ESP8266 microcontroller

1 Introduction

In wireless networks, the Internet of things (IoT) became significant where the Internet of things consists of connected things that communicating with each other to make form worldwide networks and introduced concept of the IoT to enable full access to the node [1].

The data in the networks are faced with risks like eavesdrop by the attacker [2]. Some sensor nodes (SN) use in essential things like military actions or tracking

Z. A. Msekh (✉)

Energy Department, College of Engineering Al Musaib, Babylon University, Hillah, Iraq

Engineering Technologies for Medical Devices Department, Hilla University College, Hillah, Iraq

A. A. Msekh

Computer Department, College of Science for Woman, Babylon University, Hillah, Iraq

important persons. So, the sensor nodes (SN) use in unsecure places that lead to attackers' easy to attacks. Therefore, sensor nodes should be safe by making secure algorithm to encrypt in WSNs [3]. One of the safest algorithms is the advanced encryption standard algorithm (AES). Advanced encryption standard (AES) is a type of symmetric key encryption used to secure data [4]. AES algorithm has security, great speed, and faster implementation in hardware and software. Also, it is new encryption standard recommended by "National Institute of Standards and Technology" (NIST) [5]. Many researchers focus on AES, secure algorithms, and wireless sensor networks in recent years. In 2012, Chandra and et al. suggested a protocol for authentication of the agent and session key foundation and depends on public key and executed the SNs in (RSA), algorithm by connecting the base station and the external agent by public key [6]. In 2014, Madhumita Panda proposed to use two algorithms of the public key, RSA and elliptic curve cryptography (ECC) with wireless sensor networks [7]. In 2020, Khairul Muttaqin, Jefril Rahmadoni proposed the process of encryption and decryption on the file, where AES is used cryptographic algorithm that also uses the Rijndael algorithm that will encrypt and decrypt the data over (128 bits) [8]. This paper designed and implemented sensor nodes based on AES security algorithm. AES considers a strong algorithm because it has high key space equal to (10^{128}) enough to resist brute force attacks [9] where designed a sensor node that will send its encrypted locations to client by a wireless channel. The client collects the data of encrypted locations by makeing a TCP/IP connection with the sensor nodes designed.

2 The Sensor Nodes

The sensor nodes (SNs) are a device that possesses sensing and communication abilities, as the central part of the WSNs. It is used in wide fields, like a monitoring (temperature, light, motion, pressure, humidity, etc.) [10]. The SN's processing does counting on the sensed and received data from other devices. Also, the SNs use to transmit and receive the data from and to other node [11] will process all information collected by the connection between Internet protocol (IP), with servers of storage in web and transmitted at the right site and time to be used in diverse applications [12].

3 The Security for, Wireless Sensor Networks

Encryption is the most important factor in the design of information and network systems that may be exposed to security attacks on WSNs, which are similar to wired networks [11]. WSNs are more vulnerable to attacks due to the prevalence of SNs in unsafe cities [13]. The symmetric and asymmetric keys are types of keys in encryption algorithms [4]. Symmetric means that the sender and receiver use the

same key (single key or secret key), while asymmetric means the sender uses a key that differs from the receiver key. In a symmetric key of encryption, the original data (plaintext) will be converted and encrypted to random nonsense data known as ciphertext by an encryption process that consists of an algorithm and a key [5, 8]. The value of the encryption key is independent of the plaintext, and the encryption algorithm produces an output depending on the used key. The output of the algorithm changes when changing the key. Advanced Encryption Standard (AES) is a type of symmetric key encryption used to secure data [5].

3.1 Advanced Encryption Standard (AES)

AES algorithm is one of the block cipher/encryption algorithms that were published by “the National Institute of Standards and Technology (NIST)” [5]. One of the most important features that NIST was considered to choose algorithm is security. The main reason behind this was evident because the main aim of AES was to develop the security issue of the data encryption standard (DES) algorithm [5, 14]. AES can protect sensitive data from attackers and not break the encrypted data compared to other proposed algorithms [15]. This was achieved by making a lot of testing on AES against theoretical and practical attacks. AES has key sizes 128, 192, and 256 bits and a linear substitution transformation network with 10, 12, and 14 rounds depending on the size of the key [16, 17]. The data encrypted by AES is divided into blocks as an array of bytes. Each round in AES’s function has four layers. Figure 1 illustrates AES encryption and decryption [5, 9].

1. Add round key transformation in this step, the case a plain text is added with the key corresponding to the current round and which had been previously created by the key schedule. This addition is an exclusive OR (XOR) process performed bitwise from the key and the state.
2. Sub-bytes transformation in this process, all the bytes in the state are replaced with different bytes according to an 8-bit look-up table called the substitution box (S-Box). Figure 2 illustrates S-Box used in substitute byte operation.
3. ShiftRows transformation this process performs byte shifting on the current state is illustrated in this operation in Fig. 3. “The first row of the state is left untouched, and the remaining three rows are shifted to the left each by different amounts where the second row is a 1-byte offset, the third row is a 2-byte offset, and the fourth is a 3-byte offset”.
4. MixColumns transformation in this process, each column which consists of four bytes is multiplied by a known 4×4 matrix described as follows:

$$\begin{bmatrix} 2 & 3 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 3 & 1 & 1 & 2 \end{bmatrix}$$

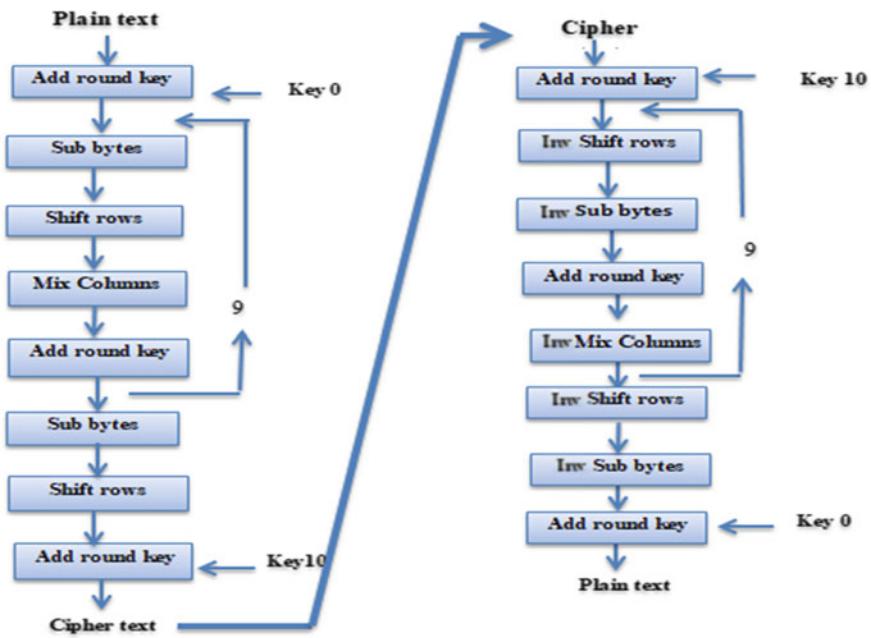


Fig. 1 AES encryption and decryption process

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	63	7C	77	7B	F2	6B	6F	C5	30	01	67	2B	FE	D7	AB	76
1	CA	82	C9	7D	FA	59	47	F0	AD	D4	A2	AF	9C	A4	72	C0
2	B7	FD	93	26	36	3F	F7	CC	34	A5	E5	F1	71	D8	31	15
3	04	C7	23	C3	18	96	05	9A	07	12	80	E2	EB	27	B2	75
4	09	83	2C	1A	1B	6E	5A	A0	52	3B	D6	B3	29	E3	2F	84
5	53	D1	00	ED	20	FC	B1	5B	6A	CB	BE	39	4A	4C	58	CF
6	D0	EF	AA	FB	43	4D	33	85	45	F9	02	7F	50	3C	9F	A8
7	51	A3	40	8F	92	9D	38	F5	BC	B6	DA	21	10	FF	F3	D2
8	CD	0C	13	EC	5F	97	44	17	C4	A7	7E	3D	64	5D	19	73
9	60	81	4F	DC	22	2A	90	88	46	EE	B8	14	DE	5E	0B	DB
A	E0	32	3A	0A	49	06	24	5C	C2	D3	AC	62	91	95	E4	79
B	E7	C8	37	6D	8D	D5	4E	A9	6C	56	F4	EA	65	7A	AE	08
C	BA	78	25	2E	1C	A6	B4	C6	E8	DD	74	1F	4B	BD	8B	8A
D	70	3E	B5	66	48	03	F6	0E	61	35	57	B9	86	C1	1D	9E
E	E1	F8	98	11	69	D9	8E	94	9B	1E	87	E9	CE	55	28	DF
F	8C	A1	89	0D	BF	E6	42	68	41	99	2D	0F	B0	54	BB	16

Fig. 2 Standard S-box

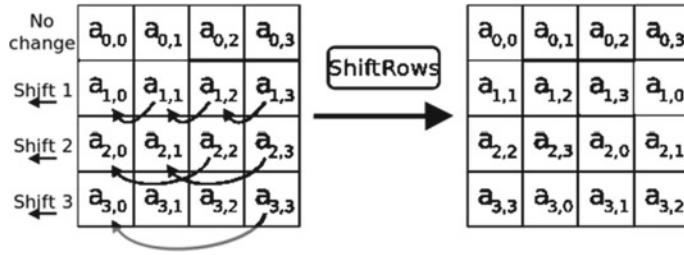


Fig. 3 Shift rows transformation

4 The Model of the Proposed System

The proposed system work in two stages, the first stage is the transmitter and the second stage is the receiver. In this work, we used four sensor nodes (A, B, C, and D) in transmitter. Each sensor node contains a microcontroller ESP8266 board, GPS, and battery. The longitude and latitude will specify by GPS. Each node (A, B, C, and D) will have a special IP address that will allow a client to connect with sensor nodes. The sensor nodes will receive the request from the client to locate them by GPS. The ESP8266 microcontroller will make an encryptions process on the specified location data by applying the AES algorithm. These data will be considered input data for the AES algorithm encryption process.

In this work, we used the key size 128-bit and 10 rounds. In the beginning step in the encryptions process, the data will convert into a block state (4×4 matrixes of bytes) to be the first input for the XOR gate, and the key in AES will be the second input for the XOR gate. Then, the XOR gate's output will be a state of the matrix (4×4 matrix of bytes) and will process by AES steps. The one round of AES has sub-processes. The first step in sub-processes is sub-bytes transformation. This step uses S-box to change a byte in the state to another. The second step is shift row; in this step, the bytes of the state will shift to left where the first row will keep without move and the second row is a one-byte offset, the third row is a 2-byte offset, and the fourth is a three-byte offset. The third step is mix columns transform; it is multiplying each row of matrix transformation with all column of the state. The fourth step is to add a round key by multiplying 28-bits of the round key with the mix column's final state by using the XOR gate. The steps mentioned above will be repeated until all rounds (10 rounds) are completed. Final will get ciphertext (encrypt data) sent to the client at the receiving side by wireless connection. Figure 4 explains a general block diagram for the proposed system and how sensor nodes work. Figure 5 shows the final construction of sensor nodes in the proposed system.

To receive the encrypted data, it has been used two main protocols, TCP/IP connections, transmission_control_Protocol (TCP) and Internet_Protocol (IP). TCP provides a connection between client and transmitter before data is sent. Meanwhile, IP will provide addressing and routing. Figure 6 demonstrates the client interface. When the client chooses the node to locate, they call by TCP/IP connections with

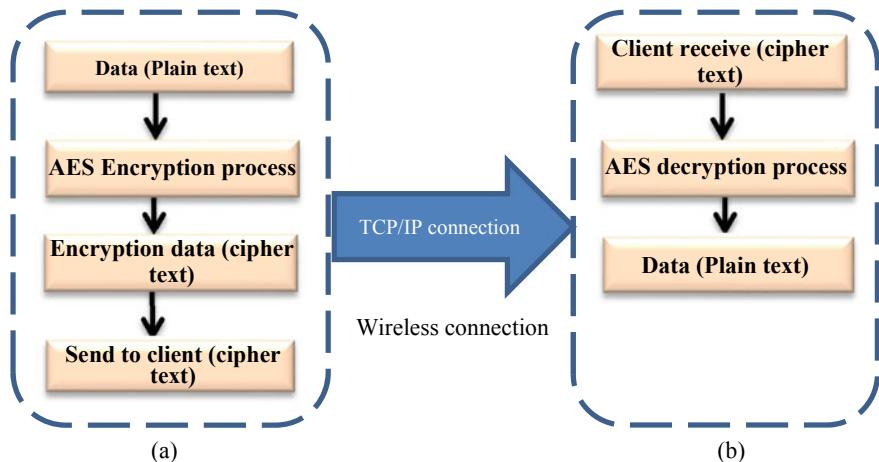


Fig. 4 General block diagram for the proposed system **a** transmitter process, **b** receiver process



Fig. 5 Sensor nodes of the proposed system



Fig. 6 Client interface in the proposed system

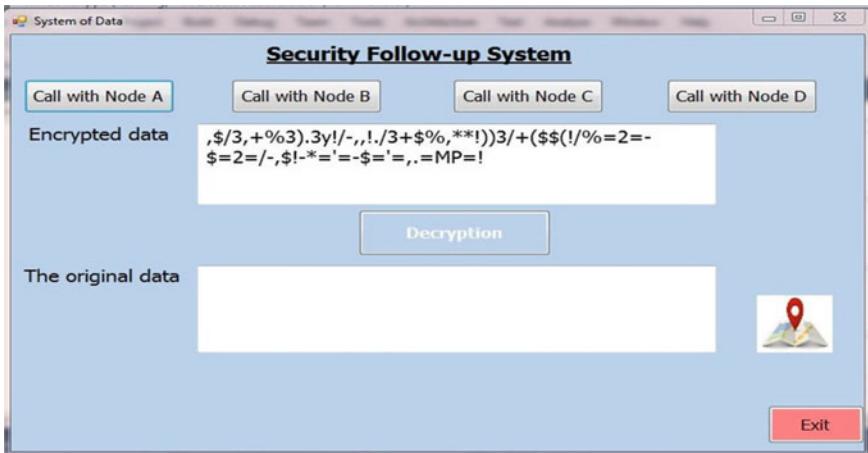


Fig. 7 Receive encrypted data from sensor node

sensor node to collect encrypt data (ciphertext) and make decryption by reverse encryption process to get the original data.

5 Results

This section will explain how the proposed system work. When making a call connection with the node, the sensor node will encrypt the location's data and send it to the client, shown in Fig. 7. We get the original data when pressing the (decryption) icon.

Finally, when pressing on the map icon, we get the specified location on the google map. Figure 8a, b shows the implementation of the proposed system by making a call connection between the client and the node with location on the Google Map.

6 Conclusions

This paper used the popular encryption algorithms AES. It provides high security to the network and data. Also, algorithms AES is useful for real-time encryption and efficient in terms of speed implementation. The original data (plaintext) was obtained without any error during the decryption process. It was easy to achieve the connection calls between client and sensor nodes. Used the microcontroller ESP8266 in sensor nodes where the performance of the implementation was perfect and accurate. Also, the microcontroller ESP8266 has a small size than other microcontrollers like Arduino and Raspberry Pi. This system can be used in military tasking, and tracking important person and can use to track blind peoples.

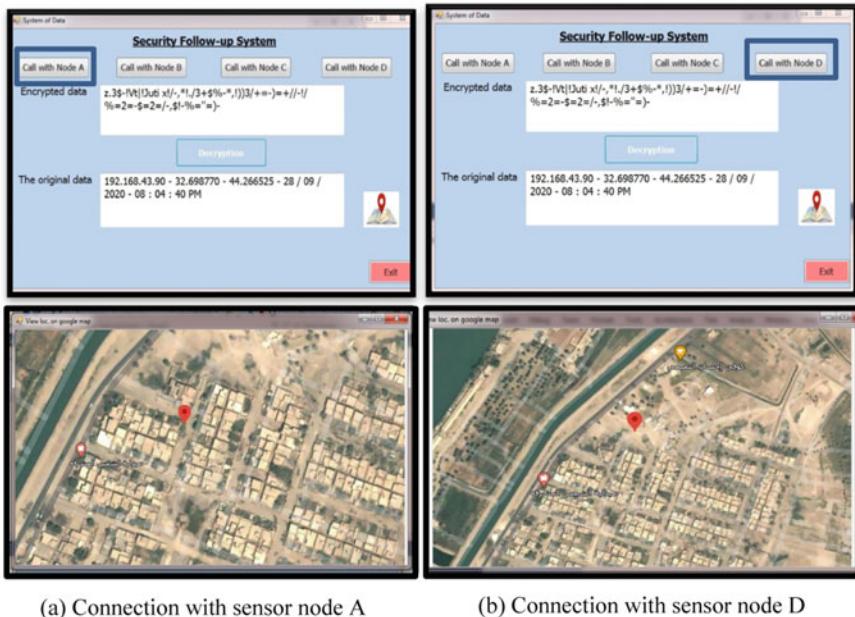


Fig. 8 Implementation of the proposed system

References

1. Dogo EM et al (2019) Blockchain and internet of things-based technologies for intelligent water management system. Artificial intelligence in IoT. Springer, pp 129–150
2. Azzabi T, Farhat H, Sahli N (2017) A survey on wireless sensor networks security issues and military specificities. In: 2017 international conference on advanced systems and electric technologies (IC_ASET). IEEE
3. Kun Y, Han Z, Zhaozui L (2009) An improved AES algorithm based on chaos. In: 2009 international conference on multimedia information networking and security. IEEE
4. Jain R et al (2014) AES algorithm using 512 bit key implementation for secure communication. Int J Innov Res Comput Commun Eng 2(3):3516–3522
5. Abdullah A (2017) Advanced encryption standard (aes) algorithm to encrypt and decrypt data. Cryptogr Netw Secur 16
6. Sekhar VC, Sarvabhatla M (2012) Security in wireless sensor networks with public key techniques. In: 2012 international conference on computer communication and informatics. IEEE
7. Panda M (2014) Security in wireless sensor networks using cryptographic techniques. Am J Eng Res (AJER) 3(01):50–56
8. Muttaqin K, Rahmadoni J (2020) Analysis and design of file security system AES (advanced encryption standard) cryptography based. J Appl Eng Technol Sci (JAETS) 1(2):113–123
9. Zeghid M et al (2007) A modified AES based algorithm for image encryption. Int J Comput Sci Eng 1(1):70–75
10. Al Shehri W (2017) A survey on security in wireless sensor networks. Int J Netw Secur Appl (IJNSA) 9(1):25–32

11. Bashaa MH, Al-Alak SM, Idrees AK (2019) Secret key generation in wireless sensor network using public key encryption. In: Proceedings of the international conference on information and communication technology
12. Kadhim JM (2020) Security of wireless sensor nodes. *Iraqi J Sci* 1773–1780
13. D'Souza M et al (2007) A wireless sensor node architecture using remote power charging, for interaction applications. In 10th Euromicro conference on digital system design architectures, methods and tools (DSD 2007). IEEE
14. Mahajan P, Sachdeva A (2013) A study of encryption algorithms AES, DES and RSA for security. *Global J Comput Sci Technol*
15. Singh G (2013) A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. *Int J Comput Appl* 67(19)
16. Kumar P, Rana SB (2016) Development of modified AES algorithm for data security. *Optik* 127(4):2341–2345
17. Obaid AJ (2021) Critical research on the novel progressive, JOKER an opportunistic routing protocol technology for enhancing the network performance for multimedia communications. In: Kumar R, Quang NH, Kumar Solanki V, Cardona M, Pattnaik PK (eds) Research in intelligent and computing in engineering. Advances in intelligent systems and computing, vol 1254. Springer. https://doi.org/10.1007/978-981-15-7527-3_36

Anomaly Detection of Ceramic Images Using Bag of Features



Zaid T. Omer and Amel H. Abbas

Abstract Ceramics may be a fabric embedded in many areas of our existence. What is required is to discover the high quality that we can benefit from in everything. To solve this problem, algorithms have been used that are involved in employing many computer programs, including machine learning. One of these algorithms, bag of feature, was used in this way. A strategy based on the classification and collection of similar images is proposed in this paper. A bag of moderated words is used for ready-made ceramic pictures. A dataset of 4000 images of 227 * 227 size, collected by a smartphone, was used. The feature bag method is a method of representing images as unordered collections of local features. And the results obtained are as good as 99.95%.

1 Introduction

Because once made, ceramic is nearly indestructible, and it is been found in giant quantities within the majority of archaeologic sites chemical analysis from the neolithic amount onward. Therefore, their rhetorical study has invariably been central to the archaeologic interpretation of the positioning, area, and period, and as a result, ceramics are a serious focus of studies of anthropology or archaeology since its beginnings within the 1950.

Ceramics area unit all around USA. This class of materials includes things like tiles, bricks, panels, and glass. Ceramics will be found in merchandise like watches, cars, and phone lines. They will even be found on area shuttles and airplanes. Counting on however its formed, the ceramics will be dense or lightweight. Usually, it will show wonderful strength and toughness properties; but, it is typically fragile

Z. T. Omer (✉)
Mustansiriyah University, Baghdad, Iraq
e-mail: zaidalradi@uomustansiriyah.edu.iq

A. H. Abbas
Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq
e-mail: dr.amelhussein2017@uomustansiriyah.edu.iq

in nature. Ceramics can even be formed as semiconducting materials, objects that permit electricity to submit to their mass, or insulators, that area unit materials that block the flow of electricity.

With all the advantages higher than, ceramics area unit in the midst of some defects like loud sound, coldness, slipping, and permanent modernization. Since precedent days, folks are discovering ceramics whether or not its natural or abnormal; however, with the event of science and technology, the invention of faults in ceramics has become easier which by mistreatment several techniques like neural network (NN) or convolutional neural network (CNN) and different technologies wherever these technologies method ceramics to supply terribly various business merchandise in terms of size, form, and detail. The aim of treating ceramics to associate degree engineering science is that the natural results of increasing the flexibility to refine, develop, and characterize ceramic materials.

Trung associate degreed Kim [1] propose an automatic thresholding technique, that is associate degree improvement of Otsu's technique, mistreatment associate degree entropy coefficient theme.

Lin et al. [2] he planned a way for sleuthing anomalies of advanced unsmooth Earth mistreatment the Fourier rework. The tactic removes Earth texture data by suggesting that of zero high-frequency elements and detects defects within the reconstructed image with a smart operator.

Roy et al. [3] this study measured the changes in roughness and mass lost from cobalt-chromium (CoCr) leg bone elements tested during a knee machine and compared them to those determined in ceramic elements.

Ammar et al. [4] chemistry electrical resistance (EIS) is taken into account to be a robust method that employments a little annoyance AC flag to review a chemistry cell [5]. EIS method was wont to characterize the interface of the metal inundated in solution employing a Potentiostat. Knowledge gotten from EIS produced two plots: the primary, a bespeak plot during which magnitude of electrical resistance and part move was arranged against recurrence; the moment, Nyquist plot wherever unreal electrical resistance was planned against real electrical resistance. Erosion resistance of the coatings was analyzed abuse each plots.

Yumerci et al. [6] during this study, associate degree investigation was conducted by mistreatment separate ripple parcel rework (DWPT) and support vector machine (SVM) strategies to work out intact and split plates. The setup was wont to arrive rise to impacts on plates amid this exploratory consider.

Min et al. [7] during this paper, we tend to propose to use the convolutional neural network (CNN) to find defects, like crack, bubble, scratch, and burr in ceramic-made merchandise. The initial experiment results indicate that the planned model achieves high defect detection accuracy.

2 Propose Method

2.1 *Bag of Visual Words*

Sack of Visual Words?

In pc Vision, an identical thought is utilized inside the sack of visual words. Here instead of taking the word from the content, picture patches and their include vectors area unit extricated from the picture into a bag. Options vector is nothing in any case a particular design that we are going to take note in a picture. To put it simply, sack of visual word is nothing in any case speaking to a picture as a set of unordered picture patches, as appeared inside the underneath outline.

2.2 *What is the Advantage?*

The image feature mainly consists of main points and recipes. The cardinal points are the interesting focuses within the picture, and indeed on-the-off chance that the picture is turned, contracted, or broadened, its essential focuses will continuously be the same. The descriptor is nothing but a depiction of the most point. The most assignment of portraying a key point is to depict a curiously rectification (key point) in a picture.

2.3 *Image Classification with Visual Word Bag*

The image classification is divided into three main steps:

- Features extraction—Define image features for a specific label.
- Code building—Building visual lexicon by gathering, taken after by recurrence examination.
- Classification—Sort pictures based on lexicon made with SVM.

2.4 *Extracting Features*

The primary step is to extricate the highlights by removing the descriptors from each picture in our dataset. They include representation strategies that bargain with how spots are spoken to as numerical vectors. These vectors are called highlight descriptors.

A great descriptor ought to have the capacity to handle thickness, revolution, scale, and reactivity varieties to a few degrees. One of the foremost common descriptors

is fastened scale parameter conversion (SIFT) and also the difference is ORB. SIFT changes over every fix into a 128-dimensional vector. When this step, every picture may be a set of vectors of an equivalent dimension (128 for SIFT), wherever the arrangement of the various vectors isn't vital [8].

2.5 Encoding Words and Building the Code

The directives created within the feature extraction step higher than are currently regenerated into coded words that agree on the words found in text documents. Passwords are solely a vector illustration of comparable spots. This code additionally leads to a codebook just like a word lexicon. This move is usually Superior by means of the k-means clustering calculation. The K-means cluster graph is appeared underneath—given as:

1. Choose the first midpoint randomly.
2. Assign each object to the block with the closest midpoint.
3. Calculate each midpoint as the average of the objects allocated to it.
4. Rehash steps 2 and 3 until it does not alter.

In summary, each fix in an image is distributed to a specific password through the method of grouping and can be identified to the image by coded word graph.

3 Classification

The other step is to represent each image in a coded word map. This is done first by applying a key point finder or including extractor and descriptor to each preparing picture, at that point coordinating each base point with those within the codebook.

The result is a graph wherever the containers match the cipher words, and also, the range of each basket compares to the quantity of times the code word matches a key purpose within the such that picture. During this method, the picture is delineated by a graph of coded words. Training picture graphs can at that point be utilized to figure out the classification demonstration. Here I am victimization SVM as a classification show.

4 Goal

Determine which images are normal and which are abnormal abstract algorithm.

A group of training images classified by viewer is used to train a model to create a classification model. Once the pattern is obtained, it is tested on the test kit to verify the model.

Step 1

Images are spoken to utilizing pixel values, and an important representation of these pictures is required to classify these pictures. We perform small image representations and sieve bags small image representation to implement a baseline, and we obtain a 16×16 px representation of the image and normalize it. The resized image is then oriented and used as a feature that represents the image.

Step 2

To implement a baseline for the classifier, we use the K-nearest neighbor algorithm K-closest neighbor algorithm. Due to the feature of the image and a set of features of the pre-classified images, we use the nomenclature of the K-nearest neighbors in the feature space to determine the label of the selected image.

Step 3

For better implementation of the workbook, we use support vector machine method. Transport truck supports on a large scale, due to its binary classification of data, the support vector machine finds an excessive level of data splitting.

For the classification of images, we implement the One-Vs-All method; i.e., for each category, we find the hyperlink that divides the features of the image into a category and not a category. Hence, we have as many super levels as the number of classes we classify. The raster product of the super plane with the feature of the image represents the distance of the data from the section and hence gives a rough estimate of the confidence. We use this information to identify the actual class by selecting the category whose hyperbolic level is furthest from representing an image feature.

Step 4

Bag of words.

A fixed volume was generated from the vocabulary of the image's features (Surf in our case) using the existing training imagery. For simplicity, points are sampled at a fixed distance (step size) and a SURF representation of each is calculated. All SIFT features are then passed through all images via the K-means algorithm to find a fixed number ($K = \text{vocab_size}$) from the feature set. This K vector set will be used as a vocabulary here next. For each training image, SURF vectors are computed with a fixed step size (they do not need to be the same as used in the vocabulary).

A histogram is calculated for each image, and this serves as a distinct representation of the image. The histogram indicates the number of image features closest to each vocabulary feature. The algorithm works best for lower step sizes but due to hardware limitations, step size 4 with SURF and 10 with SURF + GIST was the least possible.

5 Datasets

A data set of 40,000 ceramic images at a scale of 227 * 227 was collected on the Temple College campus using the smartphone as an information sensor. Each image is explained through numerous annotations. In this study, to achieve a great compromise between computing and location resolution, each test can be a fix for a 99 * 99 pixel image.

3. Channels (RGB) produced by the inspection technique are shown in the following steps:

1. A fix whose center is within $f = 5$ pixels from the midpoint of the fracture is considered a positive fix; Another thing is negative patch.
2. To reduce the convergence of the priming tests, the cover of two positive spots P1 and P2, which are connected as $O = \text{area}(P1 \cup P2) = \text{area}(P1) + \text{area}(P2)$, should be kept at a low level. In this reasoning, we choose the separation of the centers of adjacent spots to be $d = 0.75w$, where w is the width of the fix. For negative spots, the two adjacent spots should not have a covering.
3. Given the repair center c , each filter repairs around c is triggered by a random point $2[0; 360]$. This plays a necessary role to expand the number of cleavage tests since the break corrections because they consisted of a small range of images collected.

Among the tests created from the previous steps, 1500 were used as the training set, and 1220 were used as the approval kit for mutual verification. When setting up the bag of feature, the number of fracture and non-cracking corrections was set in all the information sets.

6 Result

Two pictures of the used dataset are shown (Fig. 1).

The bag of feature was used to make training images 60%, validation 20%, and testing 20%. The location of the feature points was determined using the grid method. The SIFT features have been extracted because they reduce noise and shadows at specific feature point locations.

- The bag of feature was used to make training images 60%, validation 20%, and testing 20%.
- The location of the feature points was determined using the grid method.
- The SIFT features have been extracted because they reduce noise and shadows at specific feature point locations.

It appeared when using K-means groups to create a visual vocabulary of 500 words (Table 1).

Clustering was completed on 78/100 iterations at 24.58 s/it.



Fig. 1 Type of datasets

Table 1 K-means clustering

Number of features	3,672,884
Number of clusters (K)	500

The feature bag has been completed

- Image category classifier evaluation for two classes of training set.
- Category 1: Negative
- Category 2: Positive
- 1632 images are evaluated ... and upon completion of the evaluation of all test groups. The confusion matrix for this test suite appears (Fig. 2 and Table 2).
- Evaluation of the image class classifier for two classes of the validation set.
- Category 1: Negative
- Category 2: Positive

Fig. 2 Plot the histogram of visual word occurrences

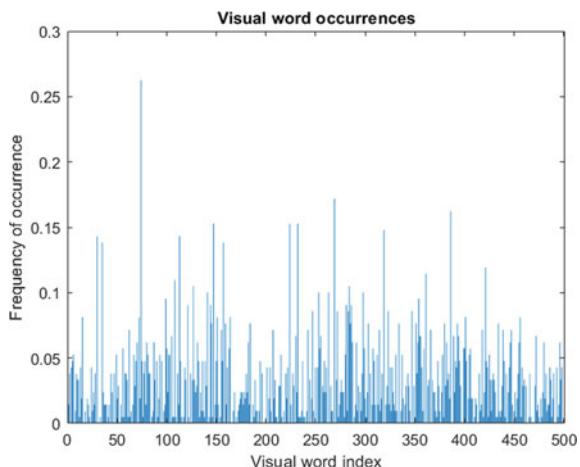


Table 2 Predicted training

KNOWN	Negative	Positive
Negative	1.00	0.00
Positive	0.00	1.00
Average accuracy: 1.00		

Table 3 Predicted validation

KNOWN	Negative	Positive
Negative	1.00	0.00
Positive	0.01	0.99
Average accuracy: 0.99		

Table 4 Predicted test

KNOWN	Negative	Positive
Negative	1.00	0.00
Positive	0.01	0.99
Average accuracy: 0.99		

- 544 images are evaluated ... and upon completion of the evaluation of all validation groups. The confusion matrix for this test suite appears (Table 3).
- Image category classifier evaluation for two classes of test set.
- Category 1: Negative
- Category 2: Positive
- 544 images are evaluated ... and upon completion of the evaluation of all test groups. The confusion matrix for this test suite appears (Table 4).

To find out the level of accuracy between the three groups, we use the following formula:

$$X.\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

The answer is 0.9959.

7 Conclusion

The representation of the distinctive bag is noteworthy due to its relative simplicity and robust performance in a number of vision tasks. Experimental results showed that with the proposed application to ceramic images, the accuracy reached 99.95%.

However, challenges remain. Since it is the success of bag of feature representation that creates the need for massive image datasets, their deployment is constrained by time and cost. Examination of the procedure is still an open issue.

The premium bag approach may be less appropriate than other measures. One can imagine the ‘Where’s Waldo’ challenge, in which bag of feature incorrectly classifies the image as containing Waldo because local highlights have many red and white stripes. Since bag of feature offerings do not recognize and display objects in photos, the admission could be wrong. The photograph appears to contain a question because her bag contains the same type and posting of salient elements as other photos containing the protest. With no data on the course of action of the highlights, many false discoveries can be expected in real-world applications.

References

1. Truong MTN, Kim S (2018) Automatic image thresholding using Otsu’s method and entropy weighting scheme for surface defect detection. *Soft Comput* 22(13):4197–4203
2. Lin B, Chen S, Han X, Liang X (2013) Automatic damage detection of engineering ceramics ground surface based on texture analysis. *Trans Tianjin Univ* 19(4):267–271
3. Roy ME, Whiteside LA, Ly KK, Gauvain MJ (2020) Cobalt-chromium total knee femoral components developed scratches and released metal debris in simulated wear whereas magnesia-stabilized Zirconia ceramic femoral components did not scratch or roughen. In: Orthopaedic proceedings, vol 102, No SUPP_9, pp 49–49. The British Editorial Society of Bone & Joint Surgery
4. Ammar AU, Shahid M, Ahmed MK, Khan M, Khalid A, Khan ZA (2018) Electrochemical study of polymer and ceramic-based nanocomposite coatings for corrosion protection of cast iron pipeline. *Materials* 11(3):332
5. Loveday D, Peterson P (2004) Evaluation of organic coatings with electrochemical impedance spectroscopy; part 1; Gamry instruments: Warminster. PA, USA
6. Yumurtaci M, Gokmen G, Akinci TC (2020) Determining damages in ceramic plates by using discrete wavelet packet transform and support vector machine. *Int J Electrical Comput Eng* 10(5):4759
7. Min B, Tin H, Nasridinov A, Yoo KH (2020) Abnormal detection and classification in i-ceramic images. In: 2020 IEEE international conference on big data and smart computing (BigComp), pp 17–18. IEEE
8. Al-asadi TA, Obaid AJ (2016) Object-based image retrieval using enhanced SURF. *Asian J Inf Technol* 15:2756–2762. <https://doi.org/10.36478/ajit.2016.2756.2762>

An Analysis of the Effect of Wireless Network Channel on Radio Fingerprint Authentication



Mohammed Mahdi Salih Altufaili, Hussein Ali Mousa,
and Yasir Abdulzahra Flaiyh Alaabedi

Abstract Radio fingerprint aims to provide an authentication and authorization for wireless devices by identification. It improves the security of wireless network at the physical layer which also enhance the time complexity by eliminating the need of authentication at higher layers. Artificial intelligence (AI) has proven its significant improvement in the radio fingerprinting, especially a multidimensional mapping feature of AI enhances the accuracy of radio fingerprint. However, the wireless network channel reduces the accuracy of radio fingerprinting algorithm. Hence, it is important to study the data classification of radio fingerprinting with lager amount of dataset collected from various wireless channels and quantitative analysis of effect of wireless network on radio fingerprint. In this research, 8 terabytes of data are collected from various wireless devices having similar radio frequency circuit. This dataset is used to analyze the effect of wireless network on radio authentication. Experimental results show the effect of wireless network channel on the classification accuracy of artificial intelligence algorithm which varies from 98 to 8% for the collected dataset. It is also observed that balancing I/Q data increases fingerprinting accuracy.

Keywords Radio fingerprint · Artificial intelligence · Radio fingerprint authentication · Wireless network channel

M. M. S. Altufaili (✉) · H. A. Mousa

Department of Computer Engineering Techniques, College of Engineering Techniques, University of Alkafeel, Al Najaf 54001, Iraq

e-mail: mohammed.altufaili@alkafeel.edu.iq

H. A. Mousa

e-mail: hussein.almealy@alkafeel.edu.iq

Y. A. F. Alaabedi

Department of Medical Laboratory Techniques, College of Medical and Health Techniques, University of Alkafeel, Al Najaf 54001, Iraq

e-mail: yasir.alaabedi@alkafeel.edu.iq

1 Introduction

With the enhancement of Internet of things (IoT), the importance of wireless network has grown drastically in last one decade. With the reliability and flexibility which embedded wireless modules provide, it is being used in multiple commercial and residential tasks. IoT has simplified the human life [1]. However, as other tools and services required authentication, wireless devices are no exception [2]. Hence, it is required to build and enhance the authentication and authorization requirements of wireless devices. There exist many authentication algorithms available for security purpose, but this either provide a limited security or consumes a lot of computational power for complex security checks.

Many research organizations have shown interest in the development and enhancement of wireless fingerprinting [3, 4]. Radio frequency circuit imposes the imperfection in the form of waveform, but devices cannot imitate this. This could be caused by phase noise, offset value of frequencies or imbalanced I/Q signals [5].

2 Related Work

In the early age of radio fingerprinting, various researches proposed radiometric signatures for wireless authentication and device identification [6]. Xu and Zheng published “Device Fingerprinting in Wireless Networks: Challenges and Opportunities.” This study also focusses and compares various supervised and un-supervised algorithm for wireless fingerprinting. Existing issues and problems occur in feature extraction step are also discussed in this research [7]. Few researches propose a fingerprinting method to distinguish devices over wireless LAN using 802.11 11 probe request frames. This enhances the security of wireless devices [8].

With the researches, it was also observed that low signal-to-noise ratio affects the accuracy of wireless fingerprinting [9]. Bratus and Cornelius proposed a wireless fingerprint-based method in their research “Active Behavioral Fingerprinting of Wireless Devices” to analyze and detect using the devices which might be spoofing MAC addresses [10].

Belgiovine in the research studied the effect of wireless network channel on the wireless fingerprinting. Finite impulse response filter is used to restore the signal at transmitter’s end reducing the I/Q signal imbalance [11]. But the detailed effect of network channel on the machine learning algorithm was not studied.

3 Data Collection and Architectures

3.1 Data Collection

Data collection process from various Wi-Fi setups is carried out for 10 days. In this, each transmission is of 30 s, this transmission is sliced into the I/Q files. These files are stored as a device burst. Each device burst contains I/Q files generated from the 10 different transmission which is transmitted in the gap of 60 s of time. For the consistency purpose, file size of each burst is maintained at around 25 GB. Every burst file in this process generates 3 I/Q data sample files.

We decorate the samples using the metadata for analyzing the data in the later stage. For this purpose, signal metadata format is used. Advantage of signal metadata format over other format is its greater readability as it is present in JSON format. So, each sample consist of I/Q data and the metadata which specifies the core information like rate of sampling, recording timing details of the I/Q data.

3.2 ResNet-50-1D CNN Architecture

Along with baseline architecture, ResNet 50 1D model is also used to analyze the effect of channel. Figure 1 shows the architecture of ResNet 50 1D model. Like baseline CNN, this network also consists of convolutional network consisting of 128 neurons, rectified linear units action function and max pool of 2×2 to minimize the data overfitting issue. This network block contains one convolutional block and identity block, which consist of filter. Which is then passed to Avgpool and softmax layer for minimizing the data overfitting and normalization of output.

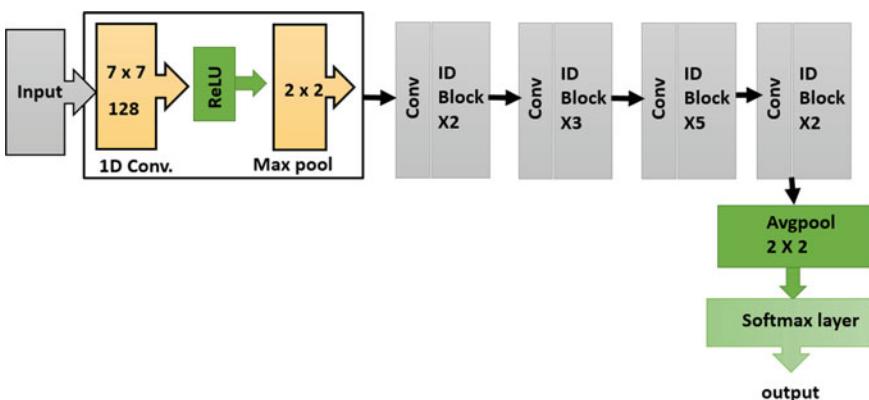


Fig. 1 Architecture of ResNet 50 1D

4 Experimental Setup

For the experiment purpose, testbed is used. Testbed is a platform for conducting rigorous, transparent and replicable testing of research theories and new technologies [12]. 20 SDR is used for conducting this experiment, it consists of N200 receiver and 7X310 Universal Software Radio Peripheral. N200 receiver also consist of CBX-120 1.2–6 GHz Rx/Tx daughterboard. This in the wild testbed is consisting of VERT2450 vertical antenna of 2.4–2.5 and 4.9–5.9 GHz. It has a total of 24 antenna which is connected to the universal software radio peripheral which controlled by servers. For another setup, anechoic chamber is used. An anechoic chamber is a room designed to completely absorb reflections of either sound or electromagnetic waves. This chamber absorbs the unwanted radio frequency waves [13]. Anechoic chamber can help us to understand the impact of channel. As this chamber is isolated, other radio frequency cannot produce noise in the transmission.

For experimental analysis four different setups were prepared.

4.1 *Multiple Antenna—In the Wild*

In this setup, each device is connected to the one antenna, i.e., one antenna for each software defined radio. Using this scenario, each device has different antenna, eventually it changes multiple configuration setting for each device like distance from the receiver, variation in the hardware impairment. This data collection in this setup was carried out for ten days.

4.2 *Single Antenna—In the Wild*

In this setup, all antennas are connected to the single antenna. This synchronizes the same distance from the receiver, same hardware impairment was experienced, and hence multipath conditions was same for all the devices [14]. This data collection setups were carried out for two days. Data collected from day one is used as training dataset, while other dataset is used in testing.

4.3 *Wired Connection*

In this setup, each transmitter is connected to receivers using subminiature version—a coaxial radio frequency cable. Similar to setup 2, in this setup, all the transmitters experience the same environment, and hence, multipath condition was same for all

the devices. This data collection setups were carried out for two days. Data collected from day one is used as training dataset, while other dataset is used in testing.

4.4 Single Antenna—Anechoic Chamber

In this setup, all transmitters are present in the anechoic chamber, and they all connect to the same antenna. As this chamber is isolated, other radio frequency cannot produce noise in the transmission. This data collection in this setup was carried out only for one day.

5 Experiment Results

In the experimental results, different channel conditions were tested with the dataset on the same day or on the next day. For the analysis of wireless network channel impact, the test records were compared. For analysis purpose, slices of 100 K size were generated from each device. Out of the collected sample, 70% data is used for training purpose, 20% is used in validation, and 10% is used for testing and evaluation of dataset.

5.1 Multiple Antenna—In the Wild

Chart 1 shows the analysis of in the wild data testing with multiple antenna which insures that the receiver distance and hardware impairment could be different for each transmitter. Hence, have multipath dataset.

Chart 1 Same and next day comparison in multiple antenna—in the wild

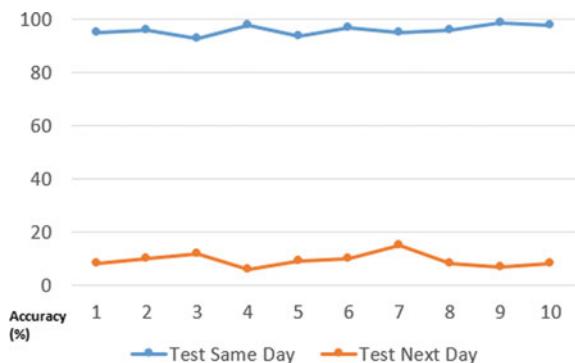
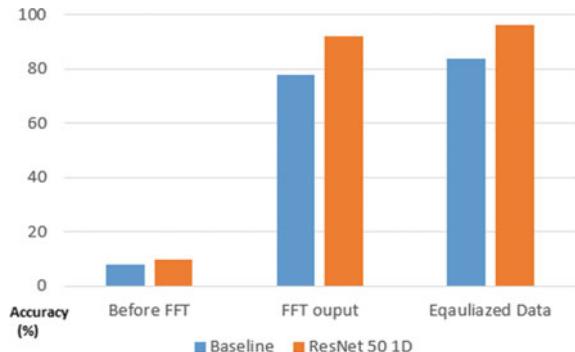


Chart 2 Comparison w.r.t. different architecture and data stage



As shown in the Chart 1, when the data accuracy is tested on the same day of data collection, the accuracy result is around 95–98%, while when we test it on the next day keeping the same training data, accuracy drops to 9–11%. This is a major drop in the accuracy which is definitely not caused by the hardware impairment, as on the next day, there were very slight changes. It confirms that the wireless network channel affects the accuracy of the algorithm.

For testing the data accuracy before the Fourier transformation filter, after transformation and I/Q balanced data is considered. Chart 2 illustrates the comparison of the data before Fourier transformation filter, after transformation and I/Q balance. It is observed that the performance of ResNet 50 1D architecture is better than baseline CNN architecture with the collected dataset. Data comparison at collected at different stages shows that equalizing the I/Q data increases the algorithm also the accuracy of data collected before FFT is to less to consider as this data does not contain any important data to distinguish the difference in two transmission slices. As the equalized data after FFT contains the binary phase-shift keying, samples are distinguishable from others.

5.2 Single Antenna—In the Wild and Wired Connection

In case of analysis with respect to single antenna, the result is very much similar in case of data before FFT. Hence, it is not considered in the analysis. Only output of FFT and output of equalized data are considered for the comparison. Tested data on the next day is not satisfying; it shows the variation in the input on both day to CNN. Accuracy of same day result was around 86.5% while it drops to 7% on the next day data; it indicates that the neural network still considers the wireless network information. Chart 3 illustrates the analysis comparison of single antenna—in the wild dataset for 10 days.

In case of analysis with respect to wired connection, the result is worse than in the wild comparison of data after FFT. Accuracy of same day result was around 87.5% while it drops to 29.5% on the next day data. However, when compared to

Chart 3 Same and next day comparison in single antenna—in the wild

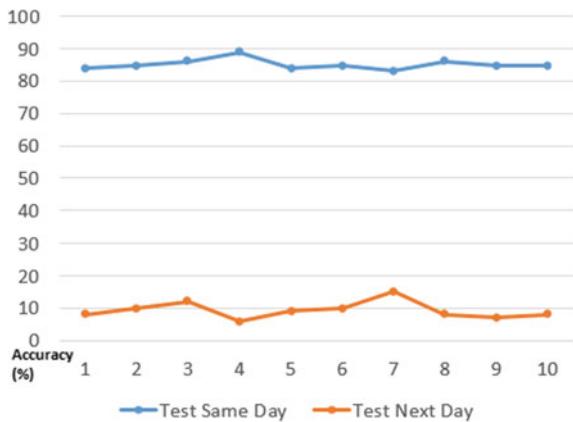
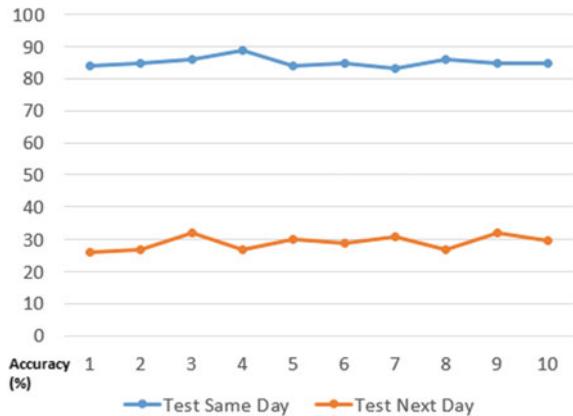


Chart 4 Same and next day comparison in wired connection

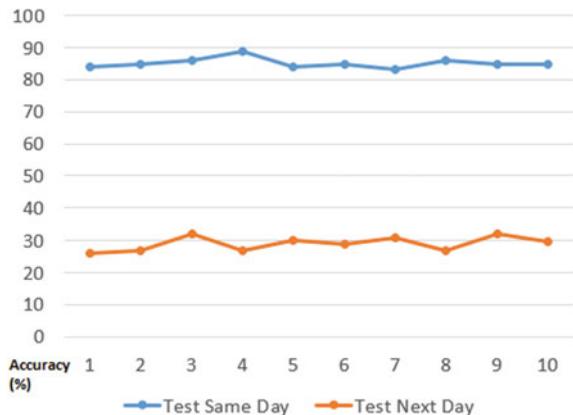


previous two cases, the difference is far better. This indicates that the algorithm is modeled better to identify the device based on the hardware impairment in case of wired environment. This is because the wired channel is same for all the inputs and neural network does not take it as a feature for identification. Chart 4 illustrates the analysis comparison of test same day and next day in wired environment.

5.3 Single Antenna—Anechoic Chamber

Anechoic chamber consists of one receiver and 10 Universal Software Radio Peripheral devices. In this setup, all transmitter is present in the anechoic chamber, and they connect to the same antenna. As this chamber is isolated, other radio frequency cannot produce noise in the transmission. The aim of testing the setup in anechoic chamber is to create the same environmental setup as wired, as clarified in (Chart 5).

Chart 5 Same and next day comparison in anechoic chamber



6 Conclusion

In this research, large amount of dataset was collected from various sources and various hardware configuration [15]. Four different experimental setups were used for the performance analysis. Two types of neural network architecture models were used. The comparison of data before FFT, after FFT, and equalized data is compared. To determine the channel effect, testing of dataset was conducted on the same day of training as well as on the next day for ten days. For the experimental study, it is observed that the accuracy of ResNet 50 1D architecture is better than baseline architecture. Data comparison collected at different stages shows that equalizing the I/Q data increases the algorithm also the accuracy of data collected before FFT is to less to consider as this data does not contain any important data to distinguish the difference in two transmission slices. As the equalized data after FFT contains the binary phase-shift keying, samples are distinguishable from others. Accuracy of neural network is around 90–98% for in the wild situation with single or multiple antennas; however, this accuracy reduces on the next day to 8–10% as the channel introduce different noise, interference, and fading. The result comparison of anechoic chamber, wired connection with in the wild analysis concludes the impact of wireless network channel on the CNN output.

References

1. Mobo FD (2018) The IoT evolution and its impacts on human life. Orient J Comput Sci Technol 11(4):188–189
2. Dabbagh YS, Saad W (2019) Authentication of wireless devices in the internet of things: learning and environmental effects. IEEE Internet Things J 6(4):6692–6705. <https://doi.org/10.1109/JIOT.2019.2910233>
3. Lin K, Chen M, Deng J, Hassan MM (2016) Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings. IEEE Trans Autom Sci Eng 13(3):1–14. <https://doi.org/>

- [10.1109/TASE.2016.2543242](https://doi.org/10.1109/TASE.2016.2543242)
- 4. Bansal R et al (2021) J Phys Conf Ser 1963:012170
 - 5. Zhuo F, Huang Y, Chen J (2017) Radio frequency fingerprint extraction of radio emitter based on I/Q imbalance. Procedia Comput Sci 107:472–477. <https://doi.org/10.1016/j.procs.2017.03.092>
 - 6. Lucia O, Isong B, Gasela N, Abu-Mahfouz AM (2019) Device authentication schemes in IoT: a review. In: Conference: 2019 international multidisciplinary information technology and engineering conference (IMITEC). <https://doi.org/10.1109/IMITEC45504.2019.9015902>
 - 7. Xu Q, Zheng R, Saad W, Han Z (2016) Device fingerprinting in wireless networks: challenges and opportunities. IEEE Commun Surv Tutor 18(1):94–104. <https://doi.org/10.1109/COMST.2015.2476338>
 - 8. Desmond LC, Yuan CC, Pheng TC, Lee RS (2008) Identifying unique devices through wireless fingerprinting. In: WiSec'08, 31 Mar–2 Apr 2008, Alexandria, Virginia, USA
 - 9. Xing Y, Hu A, Zhang J, Peng L, Li G (2018) On radio frequency fingerprint identification for DSSS systems in low SNR scenarios. IEEE Commun Lett 22(11):2326–2329
 - 10. Bratus S, Cornelius C, Kotz D, Peebles D (2008) Active behavioral fingerprinting of wireless devices. In: Proceedings of the ACM conference on wireless network security (WiSec). ACM, pp 56–61. <https://doi.org/10.1145/1352533.1352543>
 - 11. Restuccia F, D'Oro S, Al-Shawabka A, Belgiovine M, Angioloni L, Ioannidis S, Chowdhury K, Melodia T (2019) DeepRadioID: real-time channel-resilient optimization of deep learning-based radio fingerprinting algorithms. In: Proceedings of the ACM international symposium on mobile ad hoc networking and computing (ACM MobiHoc). ACM, pp 51–60
 - 12. Soltanieh N, Norouzi Y, Yang Y, Karmakar NC (2020) A review of radio frequency fingerprinting techniques. IEEE J Radio Freq Identif PP(99):1–1. <https://doi.org/10.1109/JRFID.2020.2968369>
 - 13. Vo-Huu TD, Vo-Huu TD, Noubir G (2016) Fingerprinting wi-fi devices using software defined radios. In: Proceedings of the 9th ACM conference on security & privacy in wireless and mobile networks. ACM, pp 3–14
 - 14. Xie F, Wen H, Li Y, Chen S, Hu L, Chen Y, Song H (2018) Optimized coherent integration-based radio frequency fingerprinting in internet of things. IEEE Internet Things J 5(5):3967–3977
 - 15. Chen F, Yan Q, Clancy T (2012) On passive wireless device fingerprinting using infinite Hidden Markov Random Field

Analysis of IoT Device Network Traffic: Thinking Toward Machine Learning



Vian Adnan Ferman and Mohammed Ali Tawfeeq

Abstract The proliferation and diversity of Internet of things (IoT) devices increase IoT system risks. This open many challenging area for research like identifying authorized and unauthorized IoT devices, classifying events of devices, detecting network traffic anomalies generated by such devices, etc. To do so, machine and deep learning algorithms should be applied after extracted a set of features from the associated device's traffic. The identification accuracy as well as the computational time are both very important factors, especially with the limited resources system. These considerations are firstly dependent on which features list is taken. So, the IoT network traffic analysis is the more important and difficult step in IoT device identification. In this paper, the traffic of six IoT home devices is analyzed using wireshark network protocol analyzer. A set of features and protocols are studied to gain insight into the important ones that can be utilized for creating a device fingerprint for identification purposes in the IoT environment. The result of the analytical study was that some features are unsuitable to be certified in the case of the testbed devices' traffic, while others are expressive features that can be used to identify devices according to their manufacturers.

Keywords Features extraction · Internet of things · Machine learning · Network analysis · Wireshark network protocol analyzer

V. A. Ferman (✉) · M. A. Tawfeeq

Computer Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

e-mail: egma018@uomustansiriyah.edu.iq

M. A. Tawfeeq

e-mail: drmatawfeeq@uomustansiriyah.edu.iq

1 Introduction

Internet of Things (IoT) is the fourth industrial revolution major technology which defines as a network of interconnected various types of devices and sensors through a wireless or wired medium that connects each other with no need for human intervention. These devices are characterized as low computational capabilities that communicate with remote servers and may use specific software instructions to transfer their data to cloud servers [1]. The rapid development of IoT technology led to introduce several opportunities for IoT markets to industrialize a variety of devices. The IoT combined markets may reach 520 billion by the end of 2021 [2] while the number of worldwide connected devices was over 23 billion in 2018 and may exceed 75 billion by 2025 [3].

Since the customers looking for economical cost devices, IoT systems are inherently insecure because adding security to these systems consider as an additional cost for IoT endpoint manufacturer [4]. A significant threat is that compromised devices can be utilized to launch a massive scale of distributed denial-of-service attacks [5, 6].

The first step of securing IoT networks is to identify the connected devices through their resulted traffic then enforce rules upon the unknown traffic [7]. Many researchers have focused on machine learning (ML) or deep learning (DL) to fulfill traffic identification depending on distinct network features. Moreover, the selected features might do not base on rules or tended to be important from their views (or up to the nature of devices' traffic) without shedding further light on the analysis step which is regarded as the backbone of the identification to achieve good accuracy with a minimum computational cost. Indeed, the performance of the model depends on how important features are selected. Therefore, network features must be chosen carefully so that they are not affected by the environment or by conceivable obfuscation techniques [8].

There are two major categories of features: (i) inherent features which represent the device characteristics itself like equipment-specific parameters, device manufacturer, and other information. (ii) Protocol features that represent the protocol features of IoT devices like logical addresses and source and destination port IDs [9], while device fingerprint may include either behavioral or flow-based features [10].

In [10–16], ML algorithms have been applied. In [10], 67 features were generated from both flow-based features and behavioral features include eight features related to interarrival time (IAT). In [11], 19 traffic measurement metrics were extracted from TCP flows then leveraged principal component analysis to identify the major features. In [12], 20 traffic measurement metrics were used from packet header features, then utilized a genetic algorithm to identify the major features. In [13], two stages classifiers are applied using twelve network features and include logical and physical addresses, IAT, etc. In [14], a set of attributes were utilized to distinguish IoT device traffic from non-IoT traffic include sleep time, average packet size, active volume, number of servers, DNS interval, unique DNS requests, most frequent port number, etc. In [15], 274 features were extracted from each TCP session. Random

forest classifier has been used to check the important features which found time to live was the most important. In [16], the two-stage with four classifiers were applied, three of them in the first stage to identify bag of port numbers, domain names, and cipher suites. The result, as well as six other attributes, was used in the second stage.

In [17–20], a DL has applied. In [17], four types of features have been extracted include the traffic volume, protocol features, statistics of packet length, and traffic direction. In [18], only two devices have been identified using IAT. On other hand, in [19] an IoT security method has been applied to detect anomalies trying a distinct number of features (297, 234, and 179). Whereas in [20, 21], the DL model used the TCP payload feature to identify ten distinct devices.

From the former studies, it is obvious that some authors have focused on increasing the identification performance regardless of computational time by taking more features or trying multi-stages and high computational models, while others have tried a set of features and used additional models at the preprocessing stage for the sake of selecting the important features. The contribution of this work is:

- Collect the traffic of six IoT home devices (about 1.2 GB at different times).
- The traffic of these devices is analyzed and scrutinized to give an early insight into the nature of some attributes and which ones are suitable for exploitation in ML algorithms as well as which ones are preferred to be ignored since they may reduce the performance of the model.

The remaining of this work is presented as follows: The device traffic collection is explained in Sect. 2. Section 3 describes the traffic analysis considering some protocols and features. Analysis discussion is clarified in Sect. 4. Finally, Sect. 5 addresses the study conclusion.

2 Traffic Collection

To collect the traffic of IoT devices, an access point is required so, Raspberry Pi is configured to work as a router and then wireshark network protocol analyzer is activated. Six IoT home devices are installed utilizing the Raspberry Pi's home network. In this paper, IoT devices have been chosen from three different manufacturers which are: aswar camera (Camera), Google home mini (Google_Mini), SonoFF Smart Switch (Switch), SonoFF Power Plug (Plug), SonoFF Power Strip (Strip), and SonoFF Bulb (Bulb).

3 Traffic Analysis

First of all, there is a gap between the traffic rates of the testbed devices. Camera and Google_Mini are both high-traffic devices in contrast to the testbed SonoFF devices. Therefore, ML algorithms are expected to be more suitable for identifying

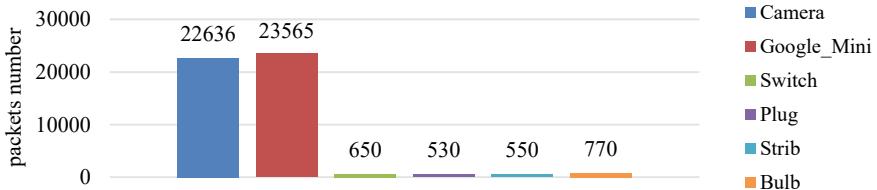


Fig. 1 Packets number during three hours

the collected traffic than DL algorithms. Figure 1 shows the packet number during three hours. There is a small difference in SonoFF devices traffic rates as well as in camera and Google_Mini traffic rates.

In this section, the collected traffic is analyzed and shed the light on some protocols and flow-based features.

3.1 Protocol List at the Initial Operation

While IoT devices are connected to the home network, a list of protocols from various network layers flow which includes EAPOL, DHCP, ARP, ICMPv6, MDNS, IGMPv2, IGMPv3, DNS, and NTP. After that, UDP and TCP packets payload are exchanged between devices and their servers. Generally speaking, not all protocols are present with every device but they are nearly the same for every device each time the traffic is collected but they may come in different sequences. Figure 2 demonstrates the protocol list in pooled traffic for 20 s. It is clear there are more differences in protocol number and type between devices of distinct manufacturers but small differences between SonoFF devices especially Plug, Strib, and Bulb. Besides, UDP payload only existed in Google_Mini (one packet) and Camera (120 packets).

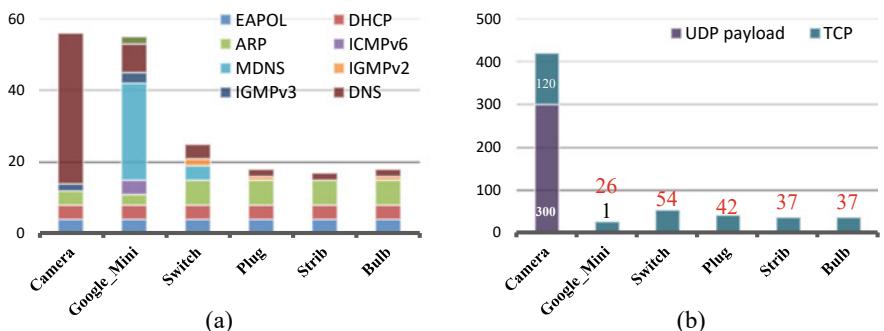


Fig. 2 Protocol list at the initial operation for 20 s

Table 1 DHCP protocol features

Testbed device	Vendor class identifier	Parameter request list item	Maximum DHCP message size	Host name
Camera	udhcp 1.19.4	1, 33, 3, 6, 15, 28, 51, 58, 59	576	Not exist
Google_Mini	dhcpcd-6.8.2:Linux-3.8.13 + :armv7l:Marvell	1, 3, 6, 12, 15, 28, 42	1500	exist
SonoFF devices	Not exist	1, 3, 28, 6, 15, 44, 46, 47, 31, 33, 121, 43	1500	exist

3.2 DHCP Protocol

In addition to the main function of the DHCP protocol of providing IP addresses to the connected devices, it can set a bunch of options that are useful for device configuration. The most important options are Vendor Class Identifier, Parameter Request List Item, Maximum DHCP Message Size, and Host Name, which are listed in Table 1 for all testbed devices. As shown, some of the features of DHCP can distinguish test device manufacturers like Parameter Request List Item and Maximum DHCP Message Size. Only Host Name can discriminate the testbed devices.

3.3 DNS and MDNS Traffic

Unlike smartphones and computers traffic, the DNS/ MDNS packets in IoT device traffic tend to be limited since IoT devices are designated to perform a specific function. DNS/MDNS query name is an important feature that is at least could categorize devices according to their manufactures. Table 2 shows the unique DNS Query Names count with examples for all testbed devices. Query Names are extracted from the whole collected traffic using python scapy library. As shown, only one Query Name exists for all SonoFF devices except the switch device that generates another one during normal operation. It is expected that number of Query Names of

Table 2 DNS query names

Testbed devices	Query names count: example
Camera	11: p2p1.cloudlinks.cn, p2p4.cloud-links.net, upg.cloudlinks.cn, etc.
Google_Mini	50: clients1.google.com, _googlecast._tcp.local, www.gstatic.com , eWeLink_1000aeaf54f.local, _googlerpc._tcp.local, etc.
SonoFF devices	1: as-dispd.coolkit.cc (during startup operation)
Switch	1: api.coolkit.cn (during normal operation)

Google_Mini are increased slightly since it depends on receiving voice commands from humans but here more commands were tried as well as connect SonoFF devices to receive orders from Google_Mini (e.g., eWeLink_1000aef54f.local). While in the case of camera, only eleven unique Query Names have been observed, and there is no expectation of an increase in number.

3.4 TCP Traffic

TCP traffic represents one of the most significant protocols of creating device fingerprints that could well distinguish device manufacturers and sometimes individual devices. The TCP packet lengths of the testbed devices traffic were extracted within three hours and took four statistical operations for the packet lengths within multiple ranges as shown in Fig. 3. TCP packet lengths of camera were almost zero or 1398, while in Google_Mini, they were distributed slightly wider than camera. Besides, there are small differences in TCP packet lengths of SonoFF devices as shown in Fig. 3c, d for Strip and Bulb, respectively.

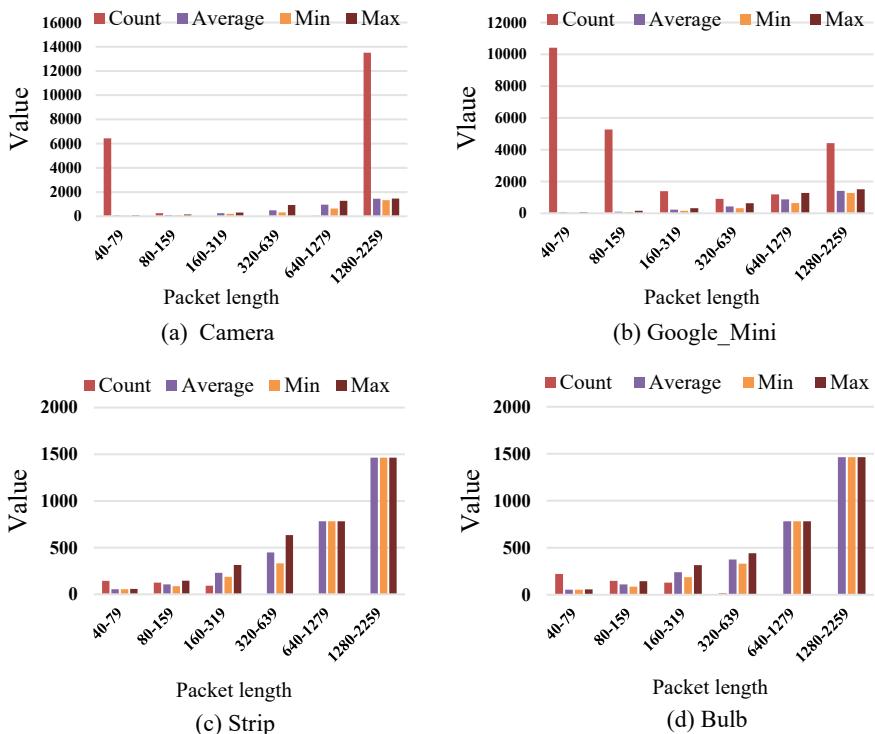


Fig. 3 Some statistical operations of TCP packet lengths within three hours

Other TCP packet attributes have been checked for the whole traffic like protocol type, remote servers (or IPs), and time-to-live regarded to TCP packets as summarized in Table 3. SonoFF devices are made conversations with just two remote servers using HTTPS protocol. Furthermore, only switch is made another conversation with a remote server using HTTP protocol (usually a few packets).

3.5 *Interarrival Time*

TCP IAT represents the time between two sequent packets. IAT of TCP packets has computed depending on timestamps (time since the previous frame in the TCP stream) as shown in Fig. 4. Since Camera and Google_Mini are high-traffic devices (short IATs), 500 TCP consecutive packets are considered while 50 TCP consecutive packets are taken for SonoFF devices (traffic is captured at different times). It is obvious the instability of IAT in the case of Google_Mini as well as sometimes tend to be like camera in IAT values, while SonoFF devices have introduced very close graphs especially plug and strip. Figure 4d shows the IAT of 50 TCP packets of bulb which were captured during a weak Internet connection (abnormal traffic caused shorter IATs).

4 Discussion

In this section, the analytical study on the traffic of the testbed devices is discussed to identify features that can be used for ML algorithms to produce high performance.

Although there are differences in the protocol list at the initial operation in all devices traffic, these protocols may not achieve high model performance as well as they are expected to be a high computational process. For example, If ten protocols are extracted for the first 20 non-repetitive packets and represent each protocol by (1) value if exist and by (0) value if not, it will produce a vector with 200 (10 features \times 20 packets) in length. This high dimensionality of the vector may decrease the identification opportunity of devices within the same manufacturer since the differences between them are little; e.g., if two MDNS packets exist in only Switch traffic, they may produce no effect with this high dimension vector. Besides, most of these protocols exist in other foreign devices, so it will be difficult to expand the proposed data in the future. To solve these issues, (i) a way must be found to reduce vector dimension while creating a device fingerprint; (ii) it requires focusing on the important protocols and assigning high weights for them; (iii) looking for other important features within protocols packets.

DHCP features and DNS/MDNS query names are strong features if they are added to the created vector in initial operation but they are not suitable if used alone due to lack of DHCP and DNS/MDNS protocol packets. The extracted TCP features are eligible to generate robust device fingerprints that are used to categorize devices

Testbed devices	Protocol type	Conversations with remote servers (or IP)	Time-to-live (device-server)
Camera	TCP (dynamic port)	32 servers, e.g., 91.220.202.6,	64-(34 to 59)
Google_Mini	TCP (dynamic port, HTTPS, and HTTP)	39 servers, e.g., storage.googleapis.com, etc.	64-(33 to 63 and 123)
SonoFF devices	HTTPS	2 servers: as-dispd.coollkit.cc, and ec2-52-220-5-109.ap-southeast-1.compute.amazonaws.com	128-(36 to 42)
Switch	HTTP	as-api.coollkit.cc	128-(57 to 60)

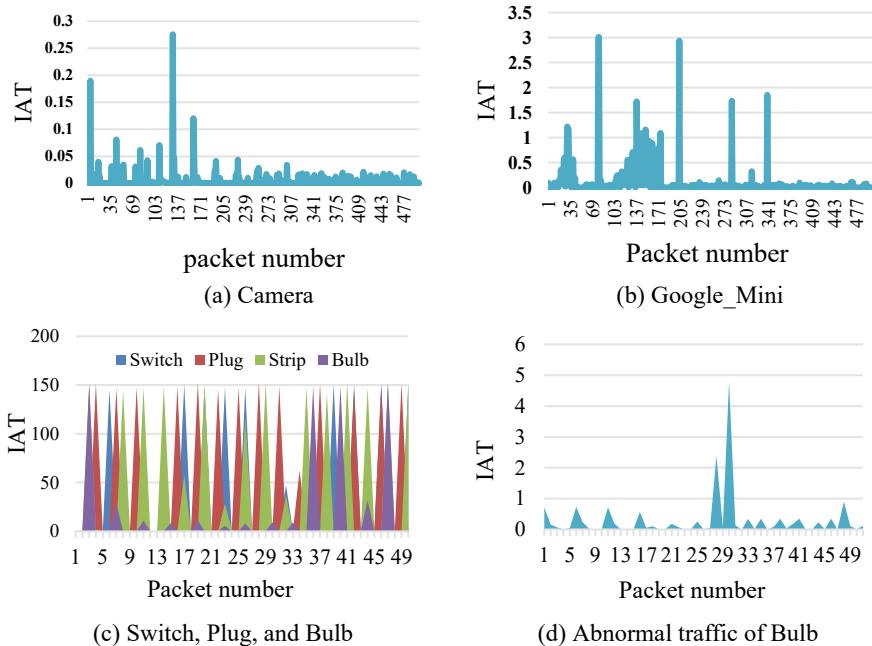


Fig. 4 IAT of TCP packets

according to their manufacturers. ML algorithms have to be tried to check if the mentioned TCP features are able to identify individual devices with high accuracy or not, and then determine whether or not other features need to be included. But, in general, TCP packet direction is a good feature to consider.

It is very clear that IAT is not suitable for identification purposes since it produces abnormal values in weak network environments. Therefore, it is expected that the results of some studies with low accuracy came from this feature while trying to perform some statistical operations to increase its weight.

5 Conclusion

This work provided an IoT device traffic analysis step to get an idea of preparing data for ML or DL algorithms. The analytical study was performed on approximately 1.2 GB of network traffic collected from six different testbed devices. Some protocols from various network layers and flow-based features were monitored and computed. From the visualizations of the extracted features and protocols, it is found that the generated protocol list and their related features in the initial operation needed some techniques to generate robust and moderate dimensional vectors to identify individual devices. Moreover, TCP features are expected to produce a very efficient model

especially if the target is identifying according to their manufactures. In the future, several ML techniques will be implemented, considering the features of TCP packets.

Acknowledgements We would like to commend Mustansiriyah University (www.uomustansiriyah.edu.iq) for their guidance in completing this work.

References

1. Qureshi A, Qureshi MA, Haider HA, Khawaja R (2020) A review on machine learning techniques for secure IoT networks. In: 2020 IEEE 23rd international multitopic conference (INMIC), pp 1–6. <https://doi.org/10.1109/INMIC50486.2020.9318092>
2. Chavis JS, Buczak A, Rubin A, Watkins LA (2020) Connected home automated security monitor (CHASM): protecting IoT through application of machine learning. In: 2020 10th Annual computing and communication workshop and conference (CCWC), pp 684–690. <https://doi.org/10.1109/CCWC47524.2020.9031162>
3. Kelly C, Pitropakis N, McKeown S, Lambrinoudakis C (2020) Testing and hardening IoT devices against the Mirai Botnet. In: 2020 International conference on cyber security and protection of digital services (cyber security), pp 1–8. <https://doi.org/10.1109/CyberSecurity49315.2020.9138887>
4. Bahizad S (2020) Risks of Increase in the IoT Devices. In: 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom), pp 178–181. <https://doi.org/10.1109/CSCloud-EdgeCom49738.2020.00038>
5. Guo H, Heidemann J (2020) Detecting IoT devices in the internet. IEEE/ACM Trans Netw 28(5):2323–2336. <https://doi.org/10.1109/TNET.2020.3009425>
6. Kumar A, Shridhar M, Swaminathan S, Lim TJ (2020) Machine learning-based early detection of IoT botnets using network-edge traffic. arXiv Prepr. arXiv2010.11453
7. Lin YC, Wang F (2018) Machine learning techniques for recognizing IoT devices. In: International computer symposium, pp 673–680. https://doi.org/10.1007/978-981-13-9190-3_74
8. Salman O, Elhajj IH, Kayssi A, Chehab A (2021) Data representation for CNN based internet traffic classification: a comparative study. Multimedia Tools Appl, pp 16951–16977. <https://doi.org/10.1007/s11042-020-09459-4>
9. Cheng W, Ding Z, Xu C, Wu X, Xia Y, Mao J (2020) RAFM: a real-time auto detecting and fingerprinting method for IoT devices. J Phys Conf Ser 1518(1):12043
10. Hamad SA, Zhang WE, Sheng QZ, Nepal S (2019) IoT device Identification via network-flow based fingerprinting and learning. In: 2019 18th IEEE international conference on trust, security and privacy in computing and communications/13th IEEE international conference on big data science and engineering (TrustCom/BigDataSE), pp 103–111. <https://doi.org/10.1109/TrustCom/BigDataSE.2019.00023>
11. Deng L, Feng Y, Chen D, Rishe N (2019) Iotspot: Identifying the iot devices using their anonymous network traffic data. In: MILCOM 2019–2019 IEEE military communications conference (MILCOM), pp 1–6. <https://doi.org/10.1109/MILCOM47813.2019.9020977>
12. Aksoy A, Gunes MH (2019) Automated IoT device identification using network traffic. In: ICC 2019–2019 IEEE international conference on communications (ICC), pp 1–7. <https://doi.org/10.1109/ICC.2019.8761559>
13. Hameed A, Leivadeas A (2020) IoT traffic multi-classification using network and statistical features in a smart environment. In: 2020 IEEE 25th international workshop on computer aided modeling and design of communication links and networks (CAMAD), pp 1–7. <https://doi.org/10.1109/CAMAD50429.2020.9209311>

14. Sivanathan A et al (2017) Characterizing and classifying IoT traffic in smart cities and campuses. In: 2017 IEEE conference on computer communications workshops (INFOCOM WKSHPS), pp 559–564. <https://doi.org/10.1109/INFCOMW.2017.8116438>
15. Meidan Y et al. (2017) Detection of unauthorized IoT devices using machine learning techniques. arXiv Prepr. arXiv1709.04647
16. Sivanathan A et al (2018) Classifying IoT devices in smart environments using network traffic characteristics. *IEEE Trans Mob Comput* 18(8):1745–1759
17. Bai L, Yao L, Kanhere SS, Wang X, Yang Z (2018) Automatic device classification from network traffic streams of internet of things. In: 2018 IEEE 43rd conference on local computer networks (LCN), pp 1–9. <https://doi.org/10.1109/LCN.2018.8638232>
18. Aneja S, Aneja N, Islam MS (2018) IoT device fingerprint using deep learning. In: 2018 IEEE international conference on internet of things and intelligence system (IOTAIS), pp 174–179. <https://doi.org/10.1109/IOTAIS.2018.8600824>
19. Bao J, Hamdaoui B, Wong, W-K (2020) IoT device type identification using hybrid deep learning approach for increased IoT security. In: 2020 International wireless communications and mobile computing (IWCMC), pp. 565–570. <https://doi.org/10.1109/IWCMC48107.2020.9148110>
20. Kotak J, Elovici Y (2020) IoT device identification using deep learning. In: Conference on complex, intelligent, and software intensive systems, pp 76–86
21. Obaid AJ (2021) Wireless sensor network (WSN) routing optimization via the implementation of fuzzy ant colony (FACO) algorithm: towards enhanced energy conservation. In: Kumar R, Mishra BK, Patnaik PK (eds) Next generation of internet of things. Lecture notes in networks and systems, vol 201. Springer, Singapore. https://doi.org/10.1007/978-981-16-0666-3_33

A Survey of Scheduling Tasks in Big Data: Apache Spark



Balqees Talal Hasan and Dhuha Basheer Abdullah

Abstract Today is the age of big data in which data is increased exponentially. Since the traditional computing systems unable to process these massive data, big data processing frameworks were developed. Apache Spark considers as one of the most relevant cluster computing frameworks for scalable data processing. The issue of task scheduling has been an active field of research in computing systems since its inception, and now in the age of big data, it is considered as one of the most important research fields. The main goal of this paper is to provide a comprehensive overview of the researches undertaken in the field of scheduling tasks in Apache Spark, and therefore, this study can be used as a starting point in the field of scheduling tasks in Apache Spark and as a benchmark to propose a novel improvement of job scheduling for Apache Spark.

Keywords Big data · Apache Spark · Scheduling algorithms · In-memory data processing

1 Introduction

Nowadays, the term “big data” has become the most trendy in the IT industry. It refers to enormous datasets whose volume, velocity, and variety make it difficult to manage and process using conventional database systems and applications [1].

Analysis of big data refers to the process of seeking insights from a huge amount of data. It is therefore important to make use of a processing framework to analyze such enormous data [2].

A variety of data processing frameworks have been developed as a result. Generally, data processing frameworks, such as MapReduce, Flink, Storm, Dryad, TensorFlow, Caffe, and Spark, are classified according to the type and condition of the data they are designed to process. Apache Hadoop MapReduce is a framework for batch processing, Flink and Storm are frameworks for streaming processing, Dryad is a

B. T. Hasan (✉) · D. B. Abdullah

Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

framework for graph processing, TensorFlow and Caffe are frameworks for deep learning. However, all frameworks listed above are not general computing frameworks because each will deal with certain data calculations. In contrast, Apache Spark considers as a general data processing framework that is commonly used in both industry and academia, it can support batch, interactive, iterative, and streaming computations [3, 4].

Big data is a developing field that has many issues that need to address. The vital problems in the big data area are: big data analysis, data management, job scheduling in big data, information privacy, and fault tolerance [2].

This paper presents a survey of job scheduling in Apache Spark. The purpose is to present a summary of research works that have been presented in this research field, highlights the advantages of each of the proposed algorithms, showcase the lacks of the proposed algorithms which can be exploited in future researches to enhance the performance of Apache Spark.

The remaining part of this survey is arranged in the following way: Sect. 2 introduces Apache Spark and job scheduling in Spark. Section 3 provides a review of the proposed scheduling algorithms. Finally, the “Conclusion and Future Work” section concludes the paper with future work.

2 Background

This section first introduces Apache Spark, followed by job scheduling in Apache Spark.

2.1 *Apache Spark*

Apache Spark is an open-source framework for cluster computing systems. Originally, Spark was developed by Zaharia at UC Berkeley as a research project in 2009 and open-sourced in 2010 under a BSD license and a very quickly acquired by Apache in 2013 [5].

Apache Spark offers a hybrid data processing framework that has the capability for handling batch and stream processing. Spark enables in-memory computation in which all the processing is done in-memory. Spark offers near-real-time stream processing capabilities in addition to batch processing [3].

The major abstraction in spark is the resilient data set (RDD) that offers fault tolerance through efficient storage of the lineage of transformations as an alternative to the data. The dataset is divided into partitions that are distributed through the cluster nodes, and it can be rebuilt using its lineage if any partition is lost [6].

2.2 Spark Job Scheduling

Each Spark application may contain many jobs. Each job may be made up of several stages. Each stage consists of one or more tasks that can be executed in parallel with other tasks [7].

In spark, the job scheduler depends on the directed acyclic graph (DAG). DAG is made of vertexes and directed edges. Any vertex is an RDD, and the edge is the procedure to be performed on the RDD [8].

There are two schedulers in Apache Spark that complement each other in scheduling Spark applications: DAG-Scheduler and Task-Scheduler. DAG-Scheduler achieves stage-oriented scheduling. It is responsible for translating the DAG into stages of tasks aggregated as a TaskSet and then sends TaskSets to the Task-Scheduler. The Task-Scheduler is in charge of submitting the tasks to the cluster nodes, running them, restarting if there are failures, and mitigating stragglers [9, 10].

Spark provides two choices as scheduling algorithms: First In-First Out (FIFO) and FAIR. By default, Spark supports the FIFO scheduling algorithm across jobs. The FAIR Scheduling algorithm prevents resource contention among jobs because it offers a roughly equal share of resources to all the tasks in a round-robin manner [11, 12].

3 Apache Spark Scheduler Optimization

For scheduling, Spark has a centralized scheduler which schedules the tasks in FIFO or FAIR fashion. Since this scheduler cannot satisfy the requirements of current data analytics, substitute proposed scheduling algorithms are used. This survey essentially focuses on the most important objectives that must be achieved in job scheduling and the optimized scheduling algorithms that have been proposed by researchers. Table 1 provides a summary of the research contributions in tasks scheduling in Apache Spark.

3.1 Data-Aware Scheduling

Data locality plays a significant role in gaining performance while running a Spark job. It determines how close the data is stored to the job processing it. To reduce the impact of network congestion among different nodes/racks, the scheduler will depend on data locality and schedules jobs close to data [13]. Since data locality considered as one of the main objectives of scheduling algorithms, some researchers address this issue.

For instance, there is a trend of jobs using only a subset of their data such as sampling-based approximate query processing systems. KMN utilizes the technique

Table 1 Summary of the research contributions in tasks scheduling in Apache Spark

Name	Merits	Demerits	Results	Workloads
KMN [14]	Minimize Cross-rack network utilization	Cluster utilization could be affected by the additional tasks	It uses only 5% more resources to reduce the average duration of a job by 81%	1. Conviva 2. Machine learning
Fu et al. [15]	This algorithm enhances the map task's data locality	The scheduler will wait for a scheme generation which involves additional time overhead	The network load and the job performance can be significantly enhanced	1. WordCount 2. Join 3. Sort
Symbiosis [17]	It reduces job completion times and Improves resource utilization	Symbiosis does not consider network usage thus adding additional delays to job processing	Compared with the current scheduler for Spark, job completion times were decreased by 11.9%	1. WordCount 2. PageRank
Firebird [16]	When dealing with various types of jobs, Firebird does not need system tuning	Firebird focuses only on scheduling of map tasks	In some cases, Firebird can be as faster as the default Spark up to 9 times	1. TPC-H 2. K-means 3. WordCount
ExpREsS [18]	It reduces energy consumption and satisfies the deadline constraints	ExpREsS requires the execution of profiling runs for multiple times	The results show improvement in the system's performance when ExpREsS has been utilized	1. Grep 2. Wordcount 3. Pagerank 4. ALS
EASAS [19]	EASAS can decrease energy consumption in Spark cluster	EASAS relies on greedy strategy that can lead to overloaded physical nodes	On average, EASAS reduces roughly (25–40%) of the overall energy usage of Spark apps	1. Sort 2. TeraSort 3. K-means 4. PageRank
A-Scheduler [22]	It schedules many jobs simultaneously with different policies based on their reliance on data	It deals with Streaming data only	It minimizes latency by 42% and increases throughput & energy efficiency by 21% and 13%, respectively, compared to the default Spark scheduler	Distributed real time event analysis

(continued)

Table 1 (continued)

Name	Merits	Demerits	Results	Workloads
DDPS [23]	It can prioritize high priority data items	DDPS only supports two priority levels: High and Low	The average latency of high priority messages has been reduced	Synthetic workload
Neptune [24]	Prioritizes the tasks dynamically to achieve low latency for stream jobs	Couroutines may increase memory burden on executors as they keep the task state in memory	The processing latencies for stream jobs are up to 3 times less	1. YahooStreaming 2. TCPH
Hetero Spark [27]	It includes the GPU accelerator in the Spark framework	It transfers data between CPU and GPU JVMs using Java RMI, which has a lot of overhead	On a range of machine learning applications, HeteroSpark shows up to $18 \times$ speed	1. Logistic regression 2. K-Means
RUPAM [26]	For each task, RUPAM selects a node that provides the best resource as well as will achieve the data locality	Unbalanced workloads have not been considered by RUPAM	Compared to the default Spark Scheduler, RUPAM will improve overall performance by 62.3%	1. TeraSort 2. Machine Learning 3. K-means
HetSpark [28]	It uses 2 types of executors: commodity executors and accelerated executors	In the event of unavailability of GPU or Processor resources, a GPU or processor job is indefinitely waiting	When apps perform linear tasks, commodity executors are preferred, however when applications perform computational tasks, GPU executors are preferred	1. BlackScholes 2. MonteCarlo
H-Scheduler [25]	It maximizes the advantage of heterogeneous storage and decreases job execution time	The data location and data reuse exploitation approach is not included by H-Scheduler	H-Scheduler decreases job execution time by up to 73.6%	1. Wordcount 2. Sort 3.Grep
Sparrow [30]	It offers near-optimal efficiency while overcoming centralized limitations	Sparrow cannot accept preemption when a high priority task comes at a worker executing a lower priority task	Sparrow is 12% superior to the median response times of the ideal scheduler	TPC-H

(continued)

Table 1 (continued)

Name	Merits	Demerits	Results	Workloads
Pigeon [31]	It reduces head-of-line blocking and avoids starvation for long jobs	In some cases the Pigeon approach may lead to load imbalance among worker groups	Trace-driven simulations and experimental findings indicate that Pigeon outperforms Sparrow	1. Yahoo 2. Cloudera 3. Google

of “late binding” which dynamically selects the input data subset according to the current cluster status. KMN is scheduling a few further tasks in the previous stage, and these additional tasks increase the probability of task outputs being spread through racks and select choices that best keep away from the congested links [14].

In [15], Fu et al. utilize the bipartite graph modeling to offer optimal scheduling of the data locality for both the map and reduce tasks. The proposed algorithm produces the best scheduling solution by converting the task scheduling problem into the bipartite graph problem.

3.2 Network-Aware Scheduling

Spark is not managing network resources explicitly. The current resource managers such as Mesos and Yarn essentially focus on memory and computing resources in a cluster. The main problem is that current schedulers have been built as isolated solutions that rely only on the resources of computing or networks. The use of computing and network resources can become unbalanced without coordination [16, 17].

Symbiosis [17] addressed challenges in coordination between computation-intensive and network-intensive tasks in a cluster, to make the usage of resources more balanced.

Firebird [16] exploits and implements software-defined networking (SDN) in Spark to improve task scheduling based on a cluster’s network status.

3.3 Energy-Aware Scheduling

Spark follows an “always-on” strategy to minimize the execution times of the applications. Thus, the cluster continuously uses all processing nodes, although the performance objectives of the workloads can be met with fewer resources. But a similar approach may lead to a considerable increase in the cluster’s energy consumption [18]. The following studies addressed this critical issue in the Spark cluster:

In [18], Maroulis et al. propose, ExpREsS, a scheduler that utilizes DVFS for reducing the energy usage so that it adjusts the clock speed of CPUs to more efficiently suit the apps running on the Spark cluster.

EASAS [19] dynamically assigns tasks to optimal executor with minimum energy utilization while fulfilling SLA which is defined as deadline constraints.

3.4 Scheduling in Spark Streaming

Apache Spark presented the idea of a micro-batch stream processing model, this model divides the data live stream into micro-batches of a pre-set time interval, then processes each micro-batch in parallel using a distributed set of tasks [20, 21]. However, to schedule micro-batches effectively to achieve low latency and increase the performance is very challenging, and there are many studies that are recently proposed to improve the scheduling of micro-batch stream jobs within Spark.

Cheng et al. [22] propose A-scheduler, an adaptive scheduling algorithm that dynamically schedules parallel micro-batch jobs using different policies based on the data dependencies. A-scheduler is implemented using two pools of jobs called independent and dependent. The independent pool uses priority-based sharing as a scheduling policy, whereas the dependent pool uses FIFO scheduling policy.

Ajila et al. [23] present the data-driven priority scheduler (DDPS) that enables users to assign a priority to all the different types of input data items. The scheduler guarantees the execution of high priority data items in time even when the system is under high or bursty input load. DDPS can be utilized in conjunction with dynamic resource allocation techniques.

In [24], Garefalakis et al. present neptune as an execution framework for unified stream and batch applications. Neptune's scheduler employs a locality- and memory-aware scheduling policy to decide which tasks to suspend and when.

3.5 Heterogeneity-Aware Scheduling

A trend in nowadays computing cluster is that heterogeneous hardware is deployed to meet diverse requirements of different big data workloads. The heterogeneity can emerge from nodes containing accelerators (e.g., GPUs, FPGAs) or nodes that are supplied with both HDDs and SSDs. This leads to nodes with different performance. However, the heterogeneity of applications and hardware do not taken into account by big data platforms which results in poor utilization of high-speed hardware [25, 26]. Several studies address the issue of heterogeneity in Spark cluster.

For instance, Li et al. [27] propose HeteroSpark, a heterogeneous architecture accelerated by GPU. The authors present a mechanism for GPU ability encapsulation in the form of callable functions and presenting them via APIs to the user applications.

Xu et al. [26] present RUPAM, a heterogeneity-aware Task-Scheduler, which schedules tasks based on data locality, task-level resource characteristics as well as node characteristics including network, storage, accelerators as well as the CPU and memory.

Hidri et al. [28] propose HetSpark which allows Spark to work with two types of executors: an accelerated executor and a commodity executor. The accelerated executors are authorized to utilize GPU together with CPU and memory, whereas the commodity executors utilize CPU cores and memory resources only.

Pan et al. [25] present a task scheduling technique called H-Scheduler for heterogeneous storage clusters. H-Scheduler categorizes the tasks by the data locality and storage device speed both together into four types: local SSD task, remote SSD task, local HDD task, and remote HDD task.

3.6 Decentralized Scheduling

The original data center schedulers were centralized. This approach has a near-perfect vision of workers and uses this vision to schedule arriving tasks to the available workers. However, scheduling latency becomes too long as the cluster sizes and workloads increase. Distributed schedulers have been proposed to solve this problem [29, 30].

Sparrow [30] implemented the load balancing technique which is known as the power of two choices in Spark. The power inquests two random workers and puts the task on the worker with fewer-queued tasks.

Pigeon [31] has been proposed for heterogeneous jobs as a distributed, hierarchical scheduler. In the Spark cluster, Pigeon divides workers into groups and delegates the task scheduling to a group master in each group.

4 Conclusion and Future Work

Apache Spark has been developed to resolve the problem of massive data processing. In order to enhance the performance of Apache Spark framework and use the computing cluster resources efficiently, different scheduling algorithms have been developed. This paper categorizes the job scheduling objectives that affecting the performance of the framework into six major categories (e.g., data-aware, energy-aware) and summarizes the contribution, shortcomings, results, and workloads used in each of the proposed algorithms. Our future work will include introducing a novel improvement of job scheduling in Apache Spark framework which can yield better performance.

References

1. Bibri SE (2019) The theoretical and disciplinary underpinnings of data–driven smart sustainable urbanism: an interdisciplinary and transdisciplinary perspective big data science and analytics for smart sustainable urbanism. Springer, pp 31–68
2. Seethalakshmi V, Govindasamy V, Akila V (2018) Job scheduling in Big Data. In: A survey 2018 international conference on computation of power, energy, information and communication (ICCPPEIC). IEEE, pp 23–31
3. Shaikh E, Mohiuddin I, Alufaisan Y, Nahvi I (2019) Apache Spark: a big data processing engine, pp 1–6
4. Tang S, He B, Yu C, Li Y, Li K (2018) a survey on spark ecosystem for big data processing
5. Salloum S, Dautov R, Chen X, Peng PX, Huang JZ (2016) Big data analytics on Apache Spark. *Int J Data Sci Anal* 1:145–164
6. Mozafari B, Ramnarayan J, Menon S, Mahajan Y, Chakraborty S, Bhanawat H, Bachhav K (2017) SnappyData: a unified cluster for streaming, transactions and interactive analytics. CIDR
7. Karau H, Warren R (2017) High performance Spark: best practices for scaling and optimizing Apache Spark. O'Reilly Media, Inc.
8. Yu J, Zhang Z, Sarwat M (2018) Geosparkviz: a scalable geospatial data visualization framework in the apache spark ecosystem. In: Proceedings of the 30th international conference on scientific and statistical database management, pp 1–12
9. Ganelin I, Orhian E, Sasaki K, York B (2016) Spark: Big data cluster computing in production. Wiley & Sons
10. Anon TaskScheduler (Spark 1.2.2 JavaDoc)
11. Zecevic P, Bonaci M (2017) Spark in action. Manning Publications Co.
12. Islam MT, Srirama SN, Karunasekera S, Buyya R (2020) Cost-efficient dynamic scheduling of big data applications in apache spark on cloud. *J Syst Softw* 162:110515
13. Ankam V (2016) Big Data Analytics. Packt Publishing, Birmingham
14. Venkataraman S (2014) The power of choice in data-aware cluster scheduling. In: Proceedings of 11th (USENIX) symposium on operating systems design and implementation (OSDI)
15. Fu Z, Tang Z, Yang L, Liu C (2020) An optimal locality-aware task scheduling algorithm based on bipartite graph modelling for spark applications. *IEEE Trans Parallel Distrib Syst* 31:2406–2420
16. He X, Shenoy P (2016) Firebird: Network-aware task scheduling for spark using SDNS. In: 2016 25th International conference on computer communication and networks (ICCCN). IEEE, pp 1–10
17. Jiang J, Ma S, Li B, Li B (2016) Symbiosis: Network-aware task scheduling in data-parallel frameworks. In: IEEE INFOCOM 2016—the 35th annual IEEE international conference on computer communications. IEEE, pp 1–9
18. Maroulis S, Zacheilas N, Kalogeraki V (2017) Express: energy efficient scheduling of mixed stream and batch processing workloads. In: 2017 IEEE international conference on autonomic computing (ICAC). IEEE, pp 27–32
19. Li H, Wang H, Fang S, Zou Y, Tian W (2019) An energy-aware scheduling algorithm for big data applications. *\\$park Cluster Comput*, pp 1–17
20. Wenig B, Damji JS, Das T, Lee D (2020) Learning spark: lightning-fast data analytics. O'Reilly Media, Incorporated
21. Penchikala S (2018) Big data processing with apache spark (Lulu. com)
22. Cheng D, Chen Y, Zhou X, Gmach D, Milojcic D (2017) Adaptive scheduling of parallel jobs in spark streaming. In: IEEE INFOCOM 2017—IEEE conference on computer communications. IEEE, pp 1–9
23. Ajila T, Majumdar S (2018) Data driven priority scheduling on spark based stream processing. In: 2018 IEEE/ACM 5th international conference on big data computing applications and technologies (BDCAT). IEEE, pp 208–210

24. Garefalakis P, Karanasos K, Pietzuch P (2019) Neptune: scheduling suspendable tasks for unified stream/batch applications. In: Proceedings of the ACM symposium on cloud computing, pp 233–245
25. Pan F, Xiong J, Shen Y, Wang T, Jiang D (2018) H-scheduler: storage-aware task scheduling for heterogeneous-storage spark clusters. In: 2018 IEEE 24th international conference on parallel and distributed systems (ICPADS). IEEE, pp 1–9
26. Xu L, Butt AR, Lim S-H, Kannan R (2018) A heterogeneity-aware task scheduler for spark. In: 2018 IEEE international conference on cluster computing (CLUSTER). IEEE, pp 245–256
27. Li P, Luo Y, Zhang N, Cao Y (2015) Heterospark: a heterogeneous cpu/gpu spark platform for machine learning algorithms. In: 2015 IEEE international conference on networking, architecture and storage (NAS). IEEE, pp 347–348
28. Hidri KK, Bilas A, Kozanitis C (2018) HetSpark: a framework that provides heterogeneous executors to Apache Spark. *Procedia Comput Sci* 136:118–127
29. Delgado P, Dinu F, Didona D, Zwaenepoel W (2016) Eagle: a better hybrid data center scheduler. *Tech Rep*
30. Ousterhout K, Wendell P, Zaharia M, Stoica I (2013) Sparrow: distributed, low latency scheduling. In: Proceedings of the twenty-fourth ACM symposium on operating systems principles, pp 69–84
31. Wang Z, Li H, Li Z, Sun X, Rao J, Che H, Jiang H (2019) Pigeon: an effective distributed, hierarchical datacenter job scheduler. In: Proceedings of the ACM symposium on cloud computing, pp 246–258

Unmanned Ground Vehicle Prototype Development for Search Evacuation and Defense Based on IoRT



Hiba A. Gumar, Baraa M. Albaker, and Mohammed N. Al-Turfi

Abstract Internet of robotic things (IoRT) applications are increased rapidly in the last few years due to the important services that are provided especially in facilitating human daily life. In particular, when it comes to protecting human lives through enabling people to reach difficult, dangerous, and uninhabited places for different reasons. This paper presents an unmanned ground vehicle (UGV) prototype mainly developed for extracting and evacuating injured persons from danger zones to the nearest safe point to get the necessary treatment as well as defending the lives of the injured persons during their transportation. The developed prototype is equipped with five degrees of freedom dual robotic arms and a door that has a treadmill. When an injured person is found, the door opens and the arms pull the victim to the treadmill to carry him inside the vehicle. First-person viewing camera and wide-angle surveillance camera are integrated and deployed for monitoring UGV road and its surrounding. A smartphone application is adopted to monitor and control the robot via the Internet. The implemented system successfully achieves different evacuation scenarios with high availability, integrity, and reliability.

Keywords UGV · IoRT · Military vehicle · Search and rescue robot · Casualty extraction · Robotic arm · Node MCU

1 Introduction

IoRT is recently considered a hot topic that represents the merging of IoT systems and robotics technologies. This is due to its importance in facilitating the life of people. One of the most critical and influential applications in robotics technology is in the search and rescue field that concerning saving lives. Robots allow reaching places that rescuers cannot reach and thus keep them away from danger; as well, they work tirelessly and increase human capabilities. After combat and disasters, there will be many fatalities owing to treatable traumatic injuries, while it can be

H. A. Gumar (✉) · B. M. Albaker · M. N. Al-Turfi
College of Engineering, Al-Iraqia University, Baghdad, Iraq

avoided if treated properly by medical treatment promptly [1, 2]. This time is known as “The Golden Hour of Trauma” [3, 4], which suggests the person’s survival rate increases significantly if it is treated timely. Therefore, the American Ministry of Defense gave priority to transporting the injured within an hour or less, which led to a marked decrease in the level of deaths [5, 6]. Nevertheless, some dangers fall on first responders who deal with the injured [7], as they must enter war zones and may encounter explosives or enemy troops targeting them. Moreover, in disasters, responders have to enter dangerous places to extract the injured [8], which makes it difficult to send them to these areas. For these cases, robots can add great contribution to save valuable human lives of both injured responders and personnel. The main contribution of this paper to develop a small-scale UGV prototype that enables various sensing and actuating devices with new features in the way of carrying the casualty and integrating the Internet of robotic things (IoRT) technology to make the unmanned ground vehicle that applicable for combat and evacuation missions. Rest paper organized as Sect. 2 is describing the related works, Sect. 3 is describing the proposed work, Sect. 4 the methodology and the implementation of the proposed work, Sect. 5 is describing the results and dictations, and finally the conclusion in Sect. 6.

2 Related Works

The robotics used in this area can be categorized by its functions into defense, search, extract, and evacuate [9]. Defensive robotics provide professional remote services that are used for military operations to enhance the existing capabilities of soldiers to keep them as safe and out of danger as possible. Ground robots increase military superiority by allowing troops to gain advantages over the enemy. Platforma-M and military assistance and surveillance system (MASS) are examples of this kind of robot [10]. In the search field, robots attempt to search and allocate the position of injured personnel, such as Remotec Wolverine [11], iRobot PackBots [12], NIFTi UGV [13], and ICARUS [14]. In the extraction field, a robot will attempt to carry injured persons from life-threatening zones to safe places. Their large volume categorizes this kind of robot, and they are usually equipped with more complex and costly equipment than the previous type. Also, extracting wounded persons is a more complicated task because of the necessity to interact with the injured persons properly without harming them. In this robot category, a remote operator uses a multi-joystick to manually control the robot. The concern with the design of some robots is how to effectively carry and transport the injured persons. Sometimes, this requires the injured to climb into the stretcher. However, some robots may provide an effective and autonomous way to carry the injured and solve the problem when the injured is unconscious or incapacitated. Systems like iRobot Valkyrie [12, 15] and Robotic Extraction and Evacuation (REX) are examples of this kind of robot [15, 16]. Other robots such as BEAR [17, 18] and cRONA [19] used two-armed robots to carry the injured persons. Robotic arms may provide a good solution to pick up and carry the

unconscious wounded individual. However, the wounded individual safety is a matter of concern. Also, several assumptions should be made to use this kind of two-armed robot including both legs of the wounded individual are present and the individual does not suffer from neck or head trauma [9]. To extract wounded individuals safely from a danger zone, a robot must capable of effectively extracting wounded individuals properly without harming them and endangering their lives. Also, it must be able to protect the injured person's body during evacuation and transportation to a safe place. Life Support for Trauma and Transport (LSTAT) is an example of this category. It has a stretcher with a sensory subsystem and a robotic snake-like manipulator [12]. The Robotic Evacuation Vehicle (REV) is another example that is the larger transport half of the marsupial pair REX and REV. Upon reaching a combat zone, REV will deploy a ramp and send REX into the field to extract a wounded soldier [15, 16].

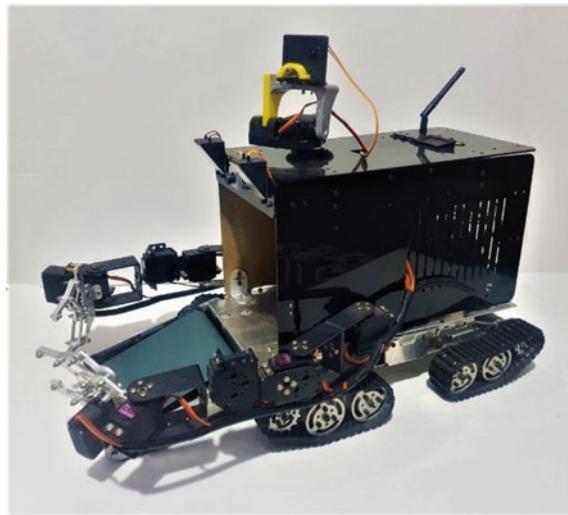
This work focuses on searching, extracting, and evacuating the injured person in safe ways without causing damage during the operation, as well as defending the injured during their transportation. The work addressed in this paper involves developing a robot prototype that is capable to search for injured persons, then extract them outside the danger zone, and finally evacuate the patient to a medical assistance location. The robot is an unmanned ground vehicle that uses IoRT technology to control the vehicle through a smartphone application via the Internet. It searches for the injured person using two cameras, one to see the road and the other gives a clear wide-view of the environment surrounding the robot. The streaming video is sent to the smartphone application and the robot's movement is controlled through the same application. The vehicle is equipped with two arms and a door that has a treadmill for safe extraction.

3 Proposed Work

As stated in the previous section, all researches focus on developing a robot that is capable to handle only one function (defense, search, extraction, and evacuation). In the extraction field, the robots take unsafe ways to carry and transport the injured persons, and in some scenarios to effectively extract and evacuate the person; two robots must be deployed for extraction and evacuation missions. The objective of the proposed prototype is to design a UGV that can do all functions above. This is maybe vital and essential in many scenarios that need to effectively extract and evacuate the injured person in time. The prototype can search for a possible injured person, and extract and evacuate the injured to a medical assistance location safely.

The robot has two sideways arms that can safely carry the injured person from an uninjured part of his body and with the help of a rotating tape included in the door that rotates when the door is open to pull the person inside the vehicle. Also, the robot has a defense tool that can defend itself against attacking enemies during personal transportation. The proposed UGV is based on the IoRT [20] using NodeMCU, and

Fig. 1 Proposed UGV prototype



it can be remotely controlled via the Internet using a smartphone application. The proposed robot is shown in Fig. 1.

4 Methodology

Several controllers are used in our prototype to handle all the functions required as stated earlier. NodeMCU is used and connected to the Internet via LAN. A tablet is connected to the Internet via WiFi and used to send requests to the NodeMCU by an application to control the actions of the robot. The simplified block diagram of the system is shown in Fig. 2.

Autodesk Maya program was used to design the 3D model that required for the project such as the cover, the base, and the tools that are used for holding and controlling the cameras, the laser, then the parts of camera holding and controlling were printed with a 3D printer with a filament polylactic acid (PLA) type. The base of the vehicle was made from aluminum and cut by aluminum CNC (controlled cutting) machine. The cover was made from acrylic and cut by acrylic CNC machine. Two NodeMCUs were used to connect the other components of the vehicle as shown in Fig. 3.

The control commands are sent to the four motors that are responsible for the movement of the vehicle. Also, two additional control commands are used for controlling the two servos (the capacitor that was added to each servo for making the movement of each servo more softy) which are responsible for the movement of a defense tool. In this case, a laser is used to simulate the defensive tool. Esp32 camera is used to send streaming video to the smartphone application. Another control signal

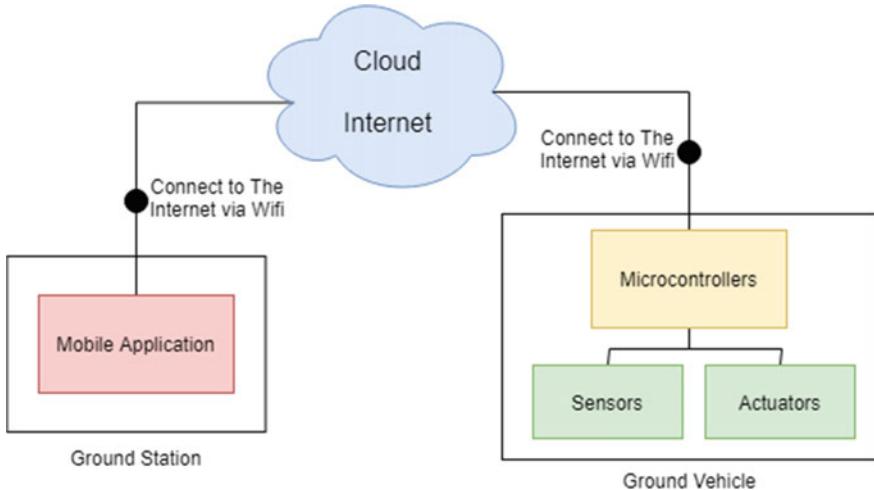


Fig. 2 Simplified block diagram of the proposed system

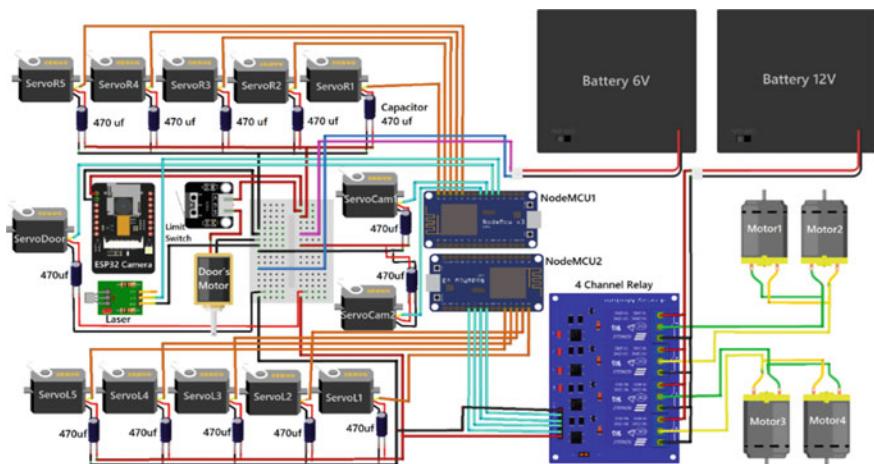


Fig. 3 Circuit schematics for the proposed robot

is used to control the door servo to open or close it. When the door is open, there is a rotating tape that rotates clockwise toward the top to facilitate the lifting of the injured person inside the vehicle. Also, two 5 degrees-of-freedom (DoF) robotic arms are integrated into the UGV, and each of them has five servos. The arms are used to raise the injured person to the vehicle with the help of the rotating tape and that for ensuring maintaining the safety of the injured person while raising the person and ensuring that his health situation does not worsen. The data flow diagram of the proposed system is shown in Fig. 4.

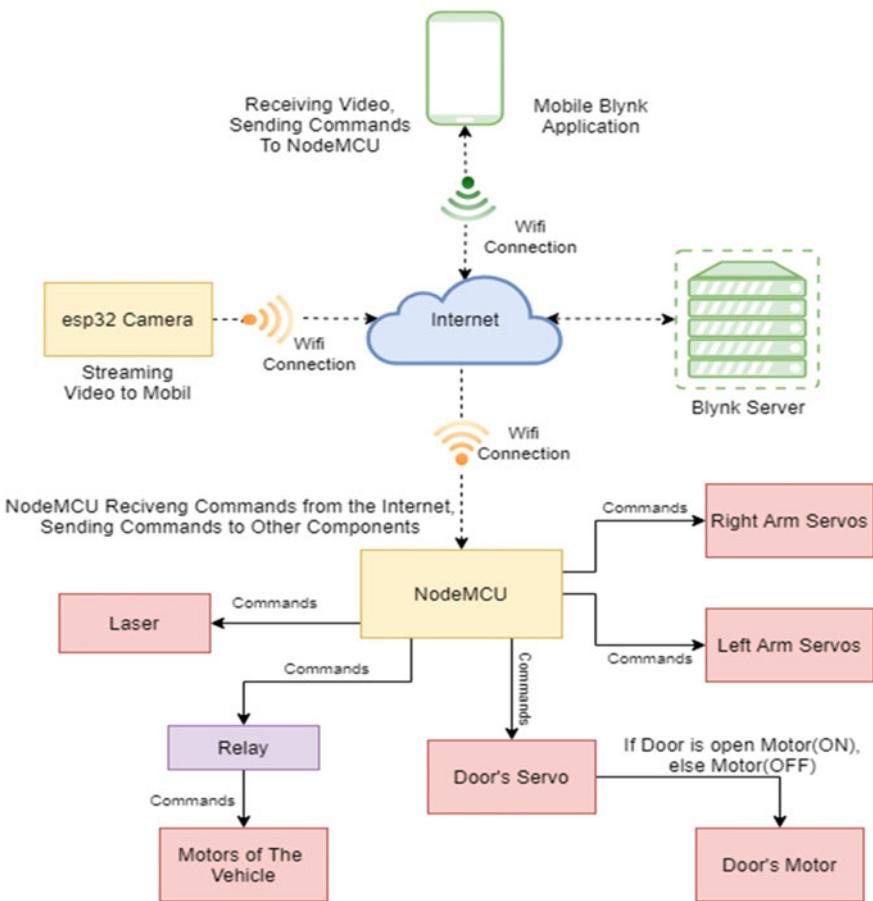


Fig. 4 Data flow diagram of the proposed system

5 Results and Discussions

After analyzing, the proposed work was highlighted the application and implementation of the UGV in terms of search, evacuate, and defense, and testing its accuracy to responding command, its ability to carry objects and insert them into the vehicle, and frequent movement, it was seen that the whole vehicle works perfectly. Blynk platform was used to control the movement of the whole robot. It works in real-time, and during the testing, the measuring time of each command response will take 1 s. The cost of the whole vehicle (including the electronics, 3D printing parts, acrylic, and aluminum pars) was 2000 US\$. The robot was controlled manually by the platform as depicted in Fig. 6.

During the implementation, the joystick in the middle was used to control the movement of the vehicle forward, back, left, and right and was seen the vehicle went

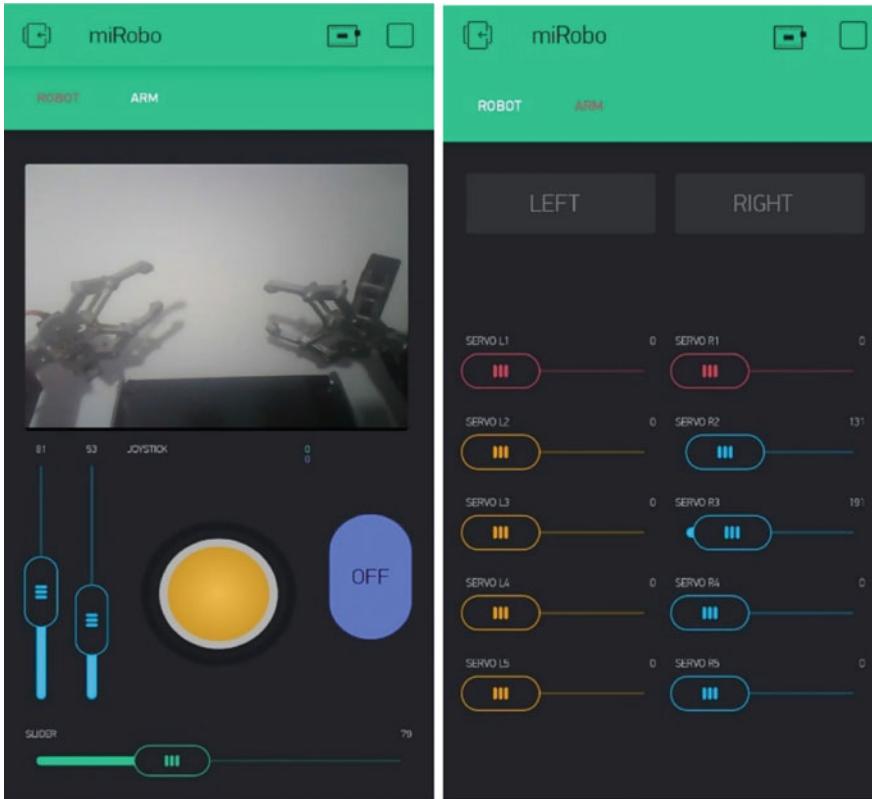


Fig. 5 Blynk interface showing **a** Robot tap, and **b** Arm tap, for controlling the movement of the robots and servos

well even off-road. The vertical slides in the left that are shown in ROBOT Tap (Fig. 5a) were used to control the movement of the camera and the laser up and down left and right. The laser (which represents the defense tool) can be turned on and off using the button on the right and the small video streaming monitor displayed online video from the esp32 camera that was used to search for objects. The servos in the ARM tap that shown in (Fig. 5b) are used to control the two robotic arms servos of the vehicle for gripping the objects that find during the searching and with the help of the rotating bar that works automatically (because of using a limited switch) when the door was open, the object can be easily lifted in the vehicle. The door can be open using the slide at the bottom. The process flow of the proposed system is shown in Fig. 6.

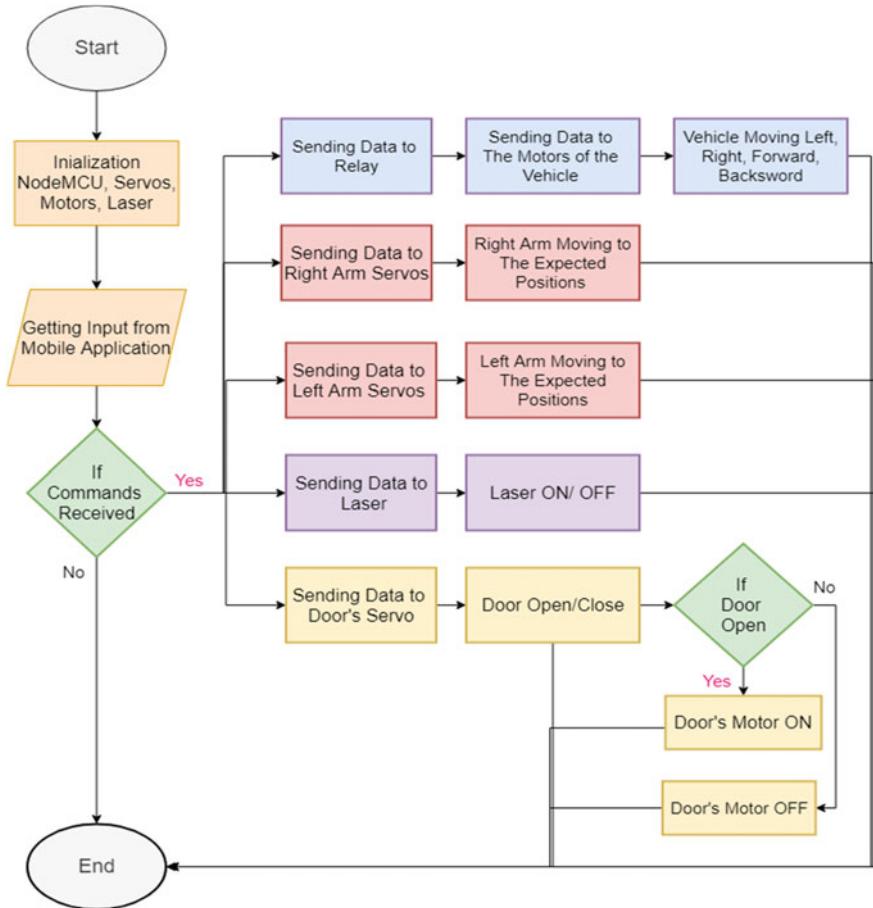


Fig. 6 Process flow of the proposed system

6 Conclusions

The Internet of robotic things is a term launched recently and used in many different fields. In this work, we focused on developing IoRT-based military vehicle prototype that is capable of handling several functions onboard. These functions include searching, extraction, and evacuation of injured persons. This is to transport injured persons from danger zones to safe places. The integration of these functions into one system allows a more powerful extraction and evacuation tool that can be useful in many scenarios than robots of single onboard carried function. A software application was adopted and used to operate the whole prototype remotely over the Internet. The proposed system was tested, and the robot successfully follows commands set by the remote operator as desired.

References

1. Dickey NW (2015) Combat trauma lessons learned from military operations of 2001–2013. Defense Health Board
2. International Federation of Red Cross and Red Crescent Societies (IFRC) (2009) World disasters report 2009: Focus on Early Warning, Early Action. Geneva
3. Dinh MM, Bein K, Roncal S, Byrne CM, Petchell J, Brennan J (2013) Redefining the golden hour for a severe head injury in an urban setting: the effect of prehospital arrival times on patient outcomes. *Injury* 44(5):606–610
4. Harmsen AMK, Giannakopoulos GF, Moerbeek PR, Jansma EP, Bonjer HJ, Bloemers FW (2015) The influence of prehospital time on trauma patients outcome: a systematic review. *Injury* 46(4):602–609
5. U.S. Army Medical Research and Material Command (2017) Unmanned systems teaming for semi-autonomous casualty extraction, SBIR-STTR. <https://www.sbir.gov/sbirsearch/detail/1319095>
6. Kotwal RS, Howard JT, Orman JA, Tarpey BW, Bailey JA, Champion HR, Mabry RL, Holcomb JB, Gross KR (2016) The effect of a golden hour policy on the morbidity and mortality of combat casualties. *JAMA Surg* 151(1):15–24
7. Chapman PL, Cabrera LD, Varela-Mayer C, Baker MM, Elnitsky C, Figley C, Thurman RM, Lin C-D, Mayer LP (2012) Training, deployment preparation, and combat experiences of deployed health care personnel: key findings from deployed U.S. Army combat medics assigned to line units. *Mil Med* 177(3):270–277
8. Department of Homeland Security (2015) First responder guide for improving survivability in improvised explosive device and/or active shooter incidents. <https://www.dhs.gov/sites/default/files/publications/First>
9. Williams A, Sebastian B, Ben-Tzvi P (2019) Review and analysis of search, extraction, evacuation, and medical field treatment robots. *J Intell Robot Syst* 96:401–418
10. Aashraya A, Munaswamy P (2020) IoT based military robot using raspberry Pi3. *Eur J Molecular Clin Med*
11. Buckstone K, Judd L, Orlowski N, Tayler-Grint M, Williams R, Zauls E (2013) Warwick mobile robotics? Urban search and rescue robot
12. Gilbert G, Turner T, Marchessault R (2007) Army medical robotics research
13. Kruijff-Korbayova I, Freda L, Gianni M, Ntouskos V, Hlavac V, Kubelka V, Zimmermann E, Surmann H, Dulic K, Rottner W, Gissi E (2016) Deployment of ground and aerial robots in earthquake-struck Amatrice in Italy (brief report). In: 2016 IEEE international symposium on safety, security, and rescue robotics (SSRR), pp 278–279
14. De Cubber G, Doroftei D, Rudin K, Berns K, Matos A, Serrano D, Sanchez JM, Govindaraj S, Bedkowski J, Roda R, Silva E, Ourevitch S, Wagemans R, Lobo V, Cardoso G, Chintamani K, Gancet J, Stupler P, Nezhadfar A, Tosa M, Balta H, Almeida J, Martins A, Ferreira H, Ferreira B, Alves J, Dias A, Fioravanti S, Bertin D, Moreno G, Cordero J, Marques MM, Grati A, Chaudhary HM, Sheers B, Riobo Y, Letier P, Jimenez MN, Esbri MA, Musialik P, Badiola I, Goncalves R, Coelho A, Pfister T, Majek K, Pelka M, Maslowski A, Baptista R (2017) Search and rescue robotics—from theory to practice. InTech
15. Gilbert GR, Beebe MK (2010) United States department of defense research in robotic unmanned systems for combat casualty care
16. Watts R, Rowe P, Gilbert G (2004) TATRC and TARDEC collaborative robots program
17. Theobald D (2010) Mobile extraction-assist robot. US Patent 7,719,222, B2
18. Atwood T, Klein J, VECNA's battlefield extraction-assist robot BEAR. Rob Mag. <http://www.botmag.com/articles/04-25-07vecnabear.html>
19. Hu J, Lim Y-J (2014) Robotic first responder system and method. US Patent 20140150806 A1
20. Byeong M, Sonya K, JongSuk C (2020) Organizing the internet of robotic things: the effect of organization structure on users' evaluation and compliance toward IoT service platform. In: IEEE/RSJ international conference on intelligent robots and systems

Data Mining Analysis Models Based on Prospective Detection of Infectious Disease



Ahmed J. Obaid

Abstract Data analytics techniques are popularly acceptable and widely preferred by researcher for analysing the data. This includes descriptive analysis using statistical methods, classification techniques to segregate data, prediction to predict and various other techniques using data mining techniques. Data analysis has its various applications widely applied in healthcare industry. Contagious diseases are a major concern for authorities; it runs rampant in places with less medical facilities. In order to curb them, they need data of these diseases. To make the strategical decision and for necessary measures, the authorities use different data mining techniques to discover the knowledge from the data. In this paper, the authors have analysed the data using statistical methods and linear regression models and time series methods. The data for contagious diseases for 5 years (2011–2015) is pooled from government website. As a result of this research strategical knowledge is generated that can be used for taking the necessary measures.

Keywords Data analysis · Regression · Statistical techniques · Time series · Health care · Contagious disease

1 Introduction

Today the world is changing very fast. Human lifestyle has dramatically changed. Technology has taken another leap, and new concepts and techniques are introduced which have been very useful. Technology is used for many purposes like education, finance, management, and health as shown in Fig. 1. Health sector is important for human lives. If quality of health is improved, then quality of life is also improved. To achieve that, diseases have to be controlled. But we see rise of many contagious diseases like malaria and diphtheria which affect human lives. Contagious diseases are those which spread rapidly from one person to another either by direct contact,

A. J. Obaid (✉)

Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Kufa, Iraq

e-mail: ahmedj.aljanaby@uokufa.edu.iq



Fig. 1 Applications of data analytics

indirect contact or droplet contact. According to NCBI, “infectious or communicable disease can be defined as an illness caused by another living agent, or its products, that can be spread from one person to another. An emergency condition can be defined as a state of disarray that has occurred during or after a regional conflict, or a natural disaster (i.e. flood, earthquake, hurricane and drought)”. Their statistics also says “infectious disease during an emergency condition can raise the death rate 60 times in comparison with other causes including trauma. Greater than 40% of deaths in emergency conditions occur secondary to diarrheal illness with 80% of those involving children less than 2 years of age”. There is rapid rise of these diseases worldwide. For example, according to Centre for Disease Control and Prevention (CDC), in 2016 in USA, there are 9272 new cases for tuberculosis (TB), 53,850 new cases of salmonella, 36,429 cases of lyme disease and 375 new cases of meningococcal disease. Now according to WHO, there is a new threat of Ebola erupting in Africa. In 1995, 17 million of 52 million deaths were caused by infectious diseases.

To improve health, we need to reduce the spread of these diseases. There is large amount of data on these diseases. But due to the lack of analysis of this data, there is little progress in preventing and reducing the contagious diseases. Here data analytics helps to analyse the data for diseases and take actions according to results. When there is sufficient data about people affected by diseases, it is collected and used with various techniques to obtain results and take actions accordingly. This paper too emphasizes on studying the contagious diseases which affect small children (under 5 yrs.) and use techniques like linear regression to analyse the data.

2 Literature Review

Vijayashree and Narayana Iyengar [1] used data mining techniques to predict heart diseases where they have found that neural networks use supervised and unsupervised learning, decision tree algorithm uses ID3 algorithm, Naïve Bayesian classifier uses probability, and genetic algorithm uses inheritance, mutation, selection and crossover for classification and prediction of heart disease. Results demonstrate that weighted associative classifier (WAC) gives more accurate results in predicting the heart disease. Umadevi and Snehapriya [2] have done a survey of various techniques and tools used for heart disease diagnosis. Their survey shows that Bayesian classifier, decision tree, neural networks and support vector machines are widely used for heart disease predictions. Further, they have compared the tools such as Rapid-Miner, Orange, Tanagra, MATLAB, KNIME and their usage. Oswal and Shah [3] in their paper tell us about the various studies of data mining in health issues and its applications in health sector. They have analysed the various applications of data mining in healthcare sector especially in predictive analysis which leads to a proper medical diagnosis. It has been found by them that decision tree, Naïve Bayes and neural networks are the widely used techniques for applications like CAD, acute lymphatic leukaemia (ALL) and various biomedical applications. The research done by Shinde and Priyadarshi [4] also deals with heart disease. The diagnosis of heart disease has been done using Naïve Bayes algorithm with an accuracy of 88.33% and 86.66 on changing the testing data set. Khemphila and Boonjing [5] have used artificial neural networks (ANN) with feature selection for diagnosis of heart disease and found that feature selection is helping in increasing the accuracy level of classification and the role of information gain (IG) in feature selection. Princy and Thomas [6] have used KNN and ID3 algorithms for detecting the risk rates of heart disease with a higher accuracy level. Sah and Sheetalani [7] have reviewed various classification techniques and found that the ensemble method of SVM and KNN gave a better accuracy level than conventional models. Jothi et al. [8] have reviewed various data mining techniques like SVM, decision tree, k-nearest neighbour and have suggested that single methods do not suffice the need of all types of detections. So, a hybrid model with the available data set needs to be developed for each type of diagnosis. Patel and Hardik [9] in their research have used step-by-step approach for data mining such as classification, clustering and association. Sanati-Mehrizy et al. [10] have studied and found that data mining is predominantly used in various medical applications such as hearing aid data set, cancer cell mining and DNA speculation. Durairaj and Ranjani [11] have done a comparative study of data mining techniques in healthcare applications, and the study shows that there was 97.77% of accuracy for cancer predictions and around 70% for estimating the success rate of in vitro fertilization (IVF) treatment. Ramageri [12] has done a survey and study on basic concepts of data mining and the different techniques that can be used for business and healthcare industry. Naidu and Rajendra [13] have proposed a new algorithm called maximal frequent itemset algorithm (MAFIA) in which K-means clustering is also integrated and classification is done using ID3 algorithm for heart disease detection.

Table 1 Data set of contagious diseases

Disease/year	Diphtheria	Measles	Diarrhoea	Malaria
2011	0.1	1.3	91.1	8
2012	0.1	0.8	95.8	2.7
2013	0	3.3	76.5	20.2
2014	0	3.7	89.5	3.5
2015	0	3.4	94.4	2.3

which gave 85% accuracy. Farhad et al. [14] where there were 86.25% cases were predicted correct by decision tree C4.5 algorithm which has been used for detecting the success rates of delivery in pregnant women and possibilities of type of delivery. Kaushar [15] has done a study of various tools that are used for implementing data mining techniques such as R and Python and suggested that the tool selection is dependent on the type of application. From these research works, we observe that many researches have been done in health sector using data mining and still there are chances of improving the efficiency of methods [16, 17]. In our work, we have incorporated the basic statistical techniques to predict contagious diseases.

3 Methodology

In this experiment, authors collected the data from government website called ncbi.gov.in; it is where all health data is stored and updated. The experiment is done on years of 2011–2015, that is, total of five years. The following is the data set (Table 1).

The experiments are implemented using R tool. R tool is used to do programming in R language. R is a programming language. It is a software used for statistical analysis, graphical representation and reporting. It is used here for linear regression to obtain results. This language has integrated development environment (IDE) and is suitable for number of languages including Python. It is free of cost and maintained under GNU open-source licence. It can run on any modern operating system.

4 Algorithms and Techniques

The techniques used in this paper are descriptive analysis using mean, standard deviation, covariance and also to use linear regression and time series regression to find the correlation between the diseases and to infer the rate of diseases in order to take preventive measures.

i. Mean

An average is the central measure of a data set. The average is measured by mean, median and mode. Mean is the sum of all numbers divided by the number of numbers given in Eq. 1.

$$\text{Mean} = \frac{\sum_{i=1}^n a_i}{n}$$

Let there be n observations a_1, a_2, \dots, a_n

$$\text{Mean} = \frac{a_1 + a_2 + \dots + a_n}{n} \quad (1)$$

E.g.: From the above table,
8, 2.7, 20.2, 3.5, 2.3

$$\text{Mean} = \frac{8 + 2.7 + 20.2 + 3.5 + 2.3}{5} = 7.34$$

ii. Standard Deviation

After obtaining the mean using Eq. 1, now we measure the variance first. Variance is mean of squared deviations. Obtaining the variance, the square root of the variance is standard deviation. The standard deviation is a very good measure of dispersion and is the one to use when the mean is used as the measure of central tendency as given in Eqs. 2 and 3. Let there be n observations a_1, a_2, \dots, a_n .

From (1)

$$a = \frac{a_1 + a_2 + \dots + a_n}{n}$$

Let $b_1 = (a_1 - a)^2, b_2 = (a_2 - a)^2$ and so on

$$\text{Variance} = \frac{b_1 + b_2 + \dots + b_n}{n} \quad (2)$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}} \quad (3)$$

For example—Using the above data, mean = 7.34.

We have deviations—0.66 -4.64 12.86 -3.84 -5.04.

Sq. of deviation—0.4356, 21.5296, 165.3796, 14.7456, 25.4016

$$\begin{aligned} \text{Variance} &= \frac{0.4356 + 21.5296 + 165.3796 + 14.7456 + 25.4016}{5} \\ &= 45.32416 \end{aligned}$$

$$\text{Standard Deviation} = \sqrt{45.32416} = 6.732$$

iii. Time Series Regression

When data is in series of particular time intervals; it is called time series data. It is basically used in statistical analysis or trend analysis.

iv. Covariance

Covariance is basically relation between two variables. It means that covariance measures the change between two variables. A positive covariance means both variables move in same direction while a negative covariance means both move in opposite directions.

$$\text{Cov}_{a,b} = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{n - 1} \quad (4)$$

on. The following steps show how covariance is calculated. The formula for covariance is as follows:

In this formula, a represents the independent variable, b represents the dependent variable, N represents the number of data points in the sample, x -axis represents the mean of the a , and y -axis represents the mean of the dependent variable b .

v. Linear Regression

Linear regression is a statistical model to find the relationship between the variables in which one may be a target and the others the predicting variables. The simple linear regression is of the form

$$Y = B_0 + B_1 * x \quad (5)$$

where Y is the target to be predicted.

5 Results and Discussions

The analysis is based on statistical computation made using mean, standard deviation and covariance. Table 2 shows the mean and standard deviation calculated for each disease from 2011 to 2015. Both of them are calculated from Eqs. 2 and 3. The mean

Table 2 Mean and standard deviation of disease

Year: 2011–2015		
Disease/year	Mean	Standard deviation
Diphtheria	0.04	0.055
Measles	2.5	1.34
Diarrhoea	89.46	7.668
Malaria	7.34	7.54

Table 3 Covariance—difference, sum squares, mean squares

Year	Difference	Sum sq	Mean sq
2012	1	3577	3577
2014	1	0	0
2012:2014	1	195	195
2013	1	5399	5399
2015	1	300	300
2013:2015	1	0	0
2011	1	6715	6715
2015	1	0	0
2011:2015	1	1	1
2012	1	3577	3577
2015	1	38	38
2012:2015	1	157	157
2011	1	6715	6715
2014	1	0	0
2011:2014	1	1	1

and standard deviation are highest for diarrhoea and least for diphtheria. Diphtheria has mean 0.04 and SD IS 0.055. Measles is slightly more with mean of 2.5 and SD 1.34. Diarrhoea, the highest mean of 89.46 and SD of 7.668. Malaria has significant mean of 7.34 and SD of 7.54.

Table 3 shows covariance of the years. Here difference, sum square and mean square are calculated. The results show the difference as 1 for all the years. The sum of square and mean square are same. For 2012 and 2014, 2012 has sum and mean square of 3577 and 2014 has value 195. Year 2013 has 5399 value of sum and mean square while 2015 has value 300. Similarly, year 2011 has value 6715 and 2014 has value 0, while year 2012 has 3577 and 2015 has 38 value of mean square and sum square.

Covariance: Covariance is used for the comparison of two or more parameters. Here we use it to compare diseases and find their relationship. We take two diseases, e.g. diarrhoea and malaria, and compare their data for all five years. As shown in figure, the parameters here are the years. The variation of the percentage of these two diseases in all these years is shown. To calculate covariance, the data should be in form of data frames. Data frames are tables or two-dimensional array structure where columns are values and rows are set of values for each column. We use it to tabulate the data so it becomes convenient to study and analyse it. The columns are years and rows are diseases, the values are percentage of disease in each year. In R, to get covariance, the code requires the use of data frames because it is easier to get relationship between two variables when they are in tabulated form. Thus, we create data frames. After creating data frames, it is used in code of covariance called

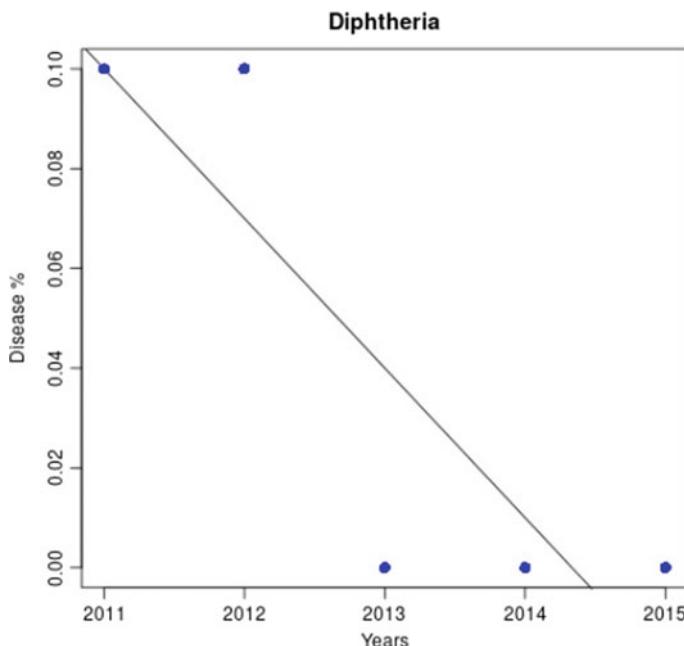


Fig. 2 Linear regression graph of diphtheria

ANCOVA. Combination of different years of the five is taken and comparisons are made. Thus, we get relationship of diarrhoea and malaria of all five years.

The above results are graphically expressed using linear regression. The graphs are calculated using R tool as shown in the following figure. The result shows the linear flow of the diphtheria disease. Diphtheria has the lowest of percentage value compared to all other diseases. The highest was 0.1% (Fig. 2).

Figure 3 shows the graphical analysis of malaria diseases from the given data set, and the graphical results show that there is an inclination from the years 2011 to 2015. The year 2013 shows maximum number of cases reported for malaria. Whereas years 2012 and 2015 are showing the lowest number of cases reported for malaria. Malaria shows little association to regression compared to other diseases. Except for 2013, it decreases from 2011 to 2015.

There is a fluctuation of occurrence of measles where it got reduced in 2012 but has increased to a greater extent in 2014. Though it has started decreasing in 2015, the eradication is yet to be done which is clearly depicted in Fig. 4.

Diarrhoea is a disease of high concern when compared to all the other diseases among children. It has the highest percentage of occurrence and has increased after 2014. Figure 5 illustrates the percentage of diarrhoea affected children against the year from 2011 to 2015. The data collected has not given any correlation between the diseases but to an extent there was an observation of occurrence of diarrhoea in children whenever they were affected with one or more of the other three diseases.

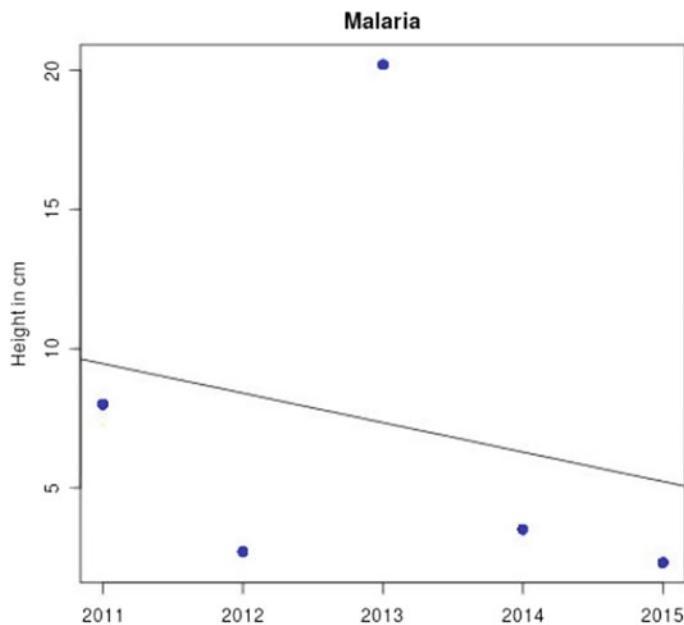


Fig. 3 Linear regression graph of malaria

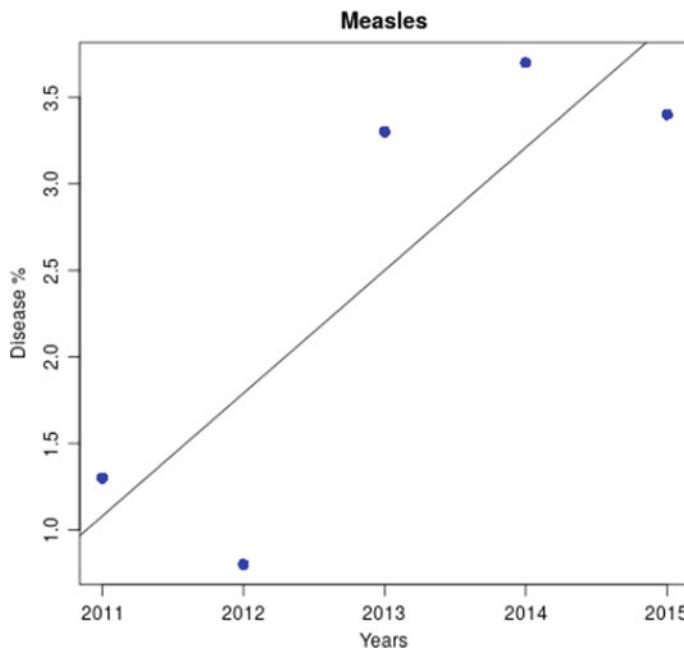


Fig. 4 Linear regression graph of measles

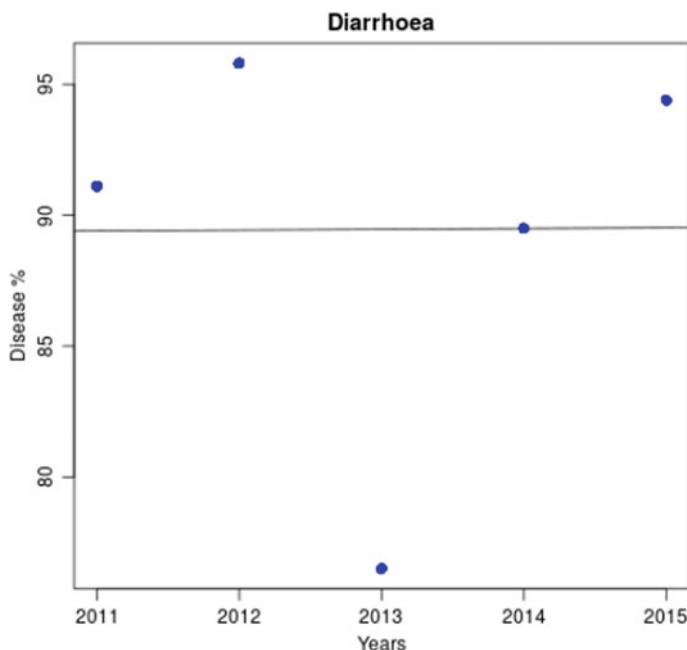


Fig. 5 Linear Regression Graph of Diarrhoea

This observation has been given by the doctors with whom the consultation was taken when working on this research work. Malaria always had a correlation with diarrhoea.

The graphical representation of multi-time series of the above-mentioned diseases is given in Figs. 6 and 7 which describes the increase and/or decrease with respect to time which is year in this case for a period of 5 years.

Diphtheria has the lowest values compared to all other diseases. From 2011 to 2012, it is constant 0.1%, and from 2013 to 2015, it is zero, i.e. nil. Also from Tables 1 and 2, we can observe that measles time series shows that first there is decrease in 2012. It is lowest in 2012 and increases from there, highest is 2014. Diarrhoea shows little randomness but its values are always high, even the lowest is 76.5. This shows that diarrhoea has affected a lot than other diseases.

6 Conclusions

The data and the results which calculates the mean and standard deviation have shown the analysis and characteristics of the diseases over the five-year period. Some diseases like diphtheria are very low which is a good sign. Diseases like diarrhoea are highest. This indicates that diarrhoea is the disease which affects most of the

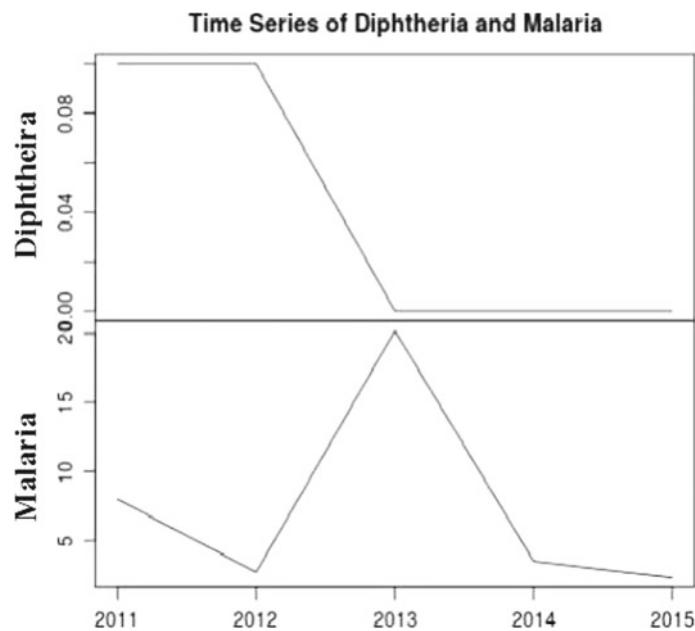


Fig. 6 Time series of diphtheria and malaria

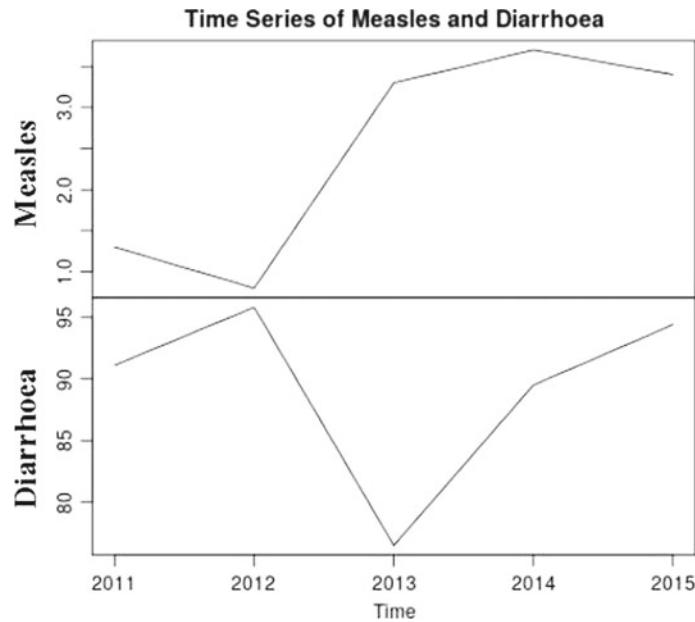


Fig. 7 Time series of measles and Diarrhoea

children. It has highest percentage value for affected children. So priority should be given to diarrhoea, and preventive measures should be taken for it. The other disease which shows significant rise is malaria, in 2013 it was 20.2. Though there is big difference compared to diarrhoea, it is still relatively higher than other diseases. Measles highest point is 3.5. So it is still within limits, and it shows it is not a big threat. This is also a good sign as measles is very low, so children are immune to it. This analysis facilitates the healthcare department to take preventive measures and curb the diseases. The government data is thus studied and given in simplified form. The study of analysing the data of diseases in the fundamental descriptive form has led to further exploration of predictive analysis using data mining techniques.

7 Limitations and Future Scope

The results obtained shows that the data is not linear, so linear regression does not work. The data is selected only for five years from 2011 to 2015. It only emphasizes on a handful of contagious diseases pertaining to only Maharashtra. The source of the data, i.e. NCBI, has vast amount of data and different kinds of data like mortality rate, still born, vaccination, diseases but here the data is only selected for some contagious diseases in Maharashtra. Since it is not linearly regressive, the percentage of children affected cannot be predicted by this method. So alternative means should be used for prediction of diseases. Here linear regression is used to study the disease. Other efficient methods can also be used here. The data is of only five years (2011–2015), and it can be extended to 10 years also. Future prediction of 2016, 2017, 2018 and 2019 can be done. This is just a portion of data from government data sets. Similar to this, study can be done for mortality rate of new born babies, vaccination, etc. This can also be extended to whole of India.

References

1. Vijayashree J, Narayana Iyengar N (2016) Heart disease prediction system using data mining and hybrid intelligent techniques: a review. *Int J Bio-Sci Bio-Technol* 8(4):139–148
2. Umadevi B, Snehapriya M (2017) A survey on prediction of heart disease using data mining techniques. *Int J Sci Res (IJSR)* 6(4):2228–2238
3. Oswal S, Shah G (2017) A study on data mining techniques on healthcare issues and its uses and application on health sector. *Int J Eng Sci Comput* 7(6):13536–13538
4. Shinde SB, Priyadarshi A (2015) Diagnosis of heart disease using data mining technique. *Int J Sci Res (IJSR)* 4(2):2301–2303
5. Khemphila A, Boonjing V (2011) Heart disease classification using neural network and feature selection. In: 2011 21st International conference on systems engineering IEEE, pp 406–409
6. Thomas J, Princy RT (2016) Human heart disease prediction system using data mining techniques. In: 2016 International conference on circuit, power and computing technologies [ICCPCT]. IEEE
7. Sah RD, Sheetalani J (2017) Review of medical disease symptoms prediction using data mining technique. *IOSR J Comput Eng* 19(3):59–70

8. Jothi N et al (2015) Data mining in healthcare—a review. *Procedia Comput Sci* 72:306–313
9. Patel S, Patel H (2016) Survey of data mining techniques used in healthcare domain. *Int J Inf Sci Tech (IJIST)* 6(1/2):53–60
10. Sanati-Mehrizi R et al (2013) A study of application of data mining algorithms in healthcare industry. In: 120th ASEE annual conference &exposition: american society for engineering education. Paper ID #6905
11. Durairaj M, Ranjani V (2013) Data mining applications in healthcare sector: a study. *Int J Sci Technol Res* 2(10):29–35
12. Ramageri BM, Data mining techniques and applications. *Indian J Comput Sci Eng* 1(4):301–305
13. Mounika Naidu P, Rajendra C (2012) Detection of health care using datamining concepts through web. *Int J Adv Res Comput Eng Technol* 1(4):45–50
14. Oabid AJ, AlBermany S, Alkaam NO (2020) Enhancement in S-box of BRADG algorithm. In: Solanki V, Hoang M, Lu Z, Pattnaik P (eds) Intelligent computing in engineering. Advances in intelligent systems and computing, vol 1125. Springer, Singapore. https://doi.org/10.1007/978-981-15-2780-7_80
15. Abdulbaqi AS, Obaid AJ, Mohammed AH (2021) ECG signals recruitment to implement a new technique for medical image encryption. *J Discrete Math Sci Crypt*. <https://doi.org/10.1080/09720529.2021.1884378>
16. Abdulbaqi AS, Obaid AJ, Mohammed AH (2021) ECG signals recruitment to implement a new technique for medical image encryption. *J Discrete Math Sci Crypt* 24(6):1663–1673. <https://doi.org/10.1080/09720529.2021.1884378>
17. Lakshmi G, Ghonge M, Obaid AJ (2021) Cloud based IoT smart healthcare system for remote patient monitoring. *EAI Endorsed Trans Perv Health Technol*. <https://doi.org/10.4108/eai.15-7-2021.170296>

Intelligent Parameter Tuning Using Deep Q-Network for RED Algorithm in Adaptive Queue Management Systems



Ayman Basheer, Hassan Jaleel Hassan, and Gaida Muttasher

Abstract Network traffic is growing with every passing day, making it more critical than ever to deal with the exploding amounts of Internet traffic and reduce delays; thus, self-learning network management systems are a must for efficient network management; such systems are called active queue management systems or AQM. This paper discusses the use of machine learning to auto-tune the parameters of AQM algorithms by training a deep reinforcement learning model to balance the queuing delay and throughput to get the maximum possible network score which is also known as the reward system. Deep Q-Network (DQN) is used as the foundation to control and auto-tune the parameters of the RED algorithm. Results from the NS3 simulation suggest that the DQN algorithm has better network reliability and is thus preferable to the RED active queue control algorithm.

Keywords Adaptive queue management · AQM · RED · Network congestion · Network traffic management · Deep Q-network · Reinforcement learning

1 Introduction

It is necessary to prevent high Internet failure levels for the packets. If a package is lost before it hits its target, it loses all the energy it has expended in transit. This condition will, in severe cases, trigger congestion to collapse [1]. Active queue management (AQM) has arisen as the sophisticated network control tool for selectively sending and receiving packets for effective management when it comes to queuing networks [2]. Unlike passive queue such as First-In-First-Out (FIFO), AQM implements the smart drop of network packets to minimize network congestion by adjusting the parameters of AQM, such as the possibility of packet-drop adaptation to the environment. Online Innovation Task Force (IETF) common usage of and approved AQM schemes [3].

A. Basheer (✉) · H. J. Hassan · G. Muttasher
Department of Computer Engineering, University of Technology, Baghdad, Iraq

H. J. Hassan
e-mail: 60012@uotechnology.edu.iq

Machine learning (ML) has developed into an essential technology in the industries and our quality of life. In fact, deep learning (DL) appears to outperform numerous ML methods in diverse fields such as effective data coding and modeling artifacts (unsupervised learning), as well as usual classification and prediction employment (supervised learning) [4]. DL has also been extended to reinforcement learning (RL), which is an ML type that looks at how the program agent decides to take appropriate action on such states to get the highest total reward [5]. In this paper, we suggest an AQM (RED) implemented deep reinforcement learning framework for effective network control and research the trade-off between queuing latency and throughput. Our program is built using DQN. Based on the current state consisting of dequeue rate, enqueue rate, drop rate, and avg_queue_lengh, it chooses a drop or non-drop operation for a packet at the departure point. Once an event is chosen, compensation is determined on the basis of the many factors to be explained in the following.

2 Previous Work

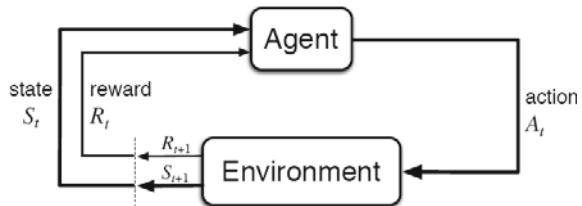
Bouacida et al. introduced learn queue AQM algorithm in [2018], which focused on wireless networking reinforcement learning. By dynamically modifying a buffer size utilizing Q-learning in a specified period, they change the Q-table and refine the Q-function strategy; however, check their method for just two and three scenarios deployed [6].

Bisoy et al. in [2017] proposed a scheme with a shallow neural network containing one of three neurons in a single hidden layer to resolve the nonlinearity of the networking framework and the queuing latency while not dealing with the trade-off between delay and throughput performance [7].

Reinforcement learning-queuing delay limitation (RL-QDL) AQM algorithm suggested in [2007] by Vučević et al. “RL agents provide topology details from the bandwidth broker that handles resource management and QoS provisioning based on what QoS requirements are met in egress routers (ERs). This supports class-based queuing (CBQ) by endorsing three separate classes: expedited forwarding (EF), guaranteed forwarding (AF), and best effort (BE) track to provide end-to-end QoS to customers with specific service types” [8]. In [as of 2018]. With respect to network scheduling algorithms, Chen et al. suggested automated computation offloading strategy focused on DRL by implementing a double DQN on the edge node. Comparing with standard algorithms, their solution implied the optimum trade-off between task latency and drop [9].

Xu et al. applied DRL to network trace engineering in [2018] by implementing an actor-critical approach with a replay of prioritized experiences. Authors contrasted their algorithm with the commonly used baseline solutions, such as load balance (LB), network utility maximization (NUM), and shortest path (SP). And they concluded that their model outperforms the base model [10].

Fig. 1 Markov decision process



3 Proposed System

We first give an explanation of DQN, which is the baseline of our system. Then we describe the design of our system in terms of the state, action, and reward, and the algorithm is followed by focusing on how to give a reward to the agent in detail.

3.1 Deep Q-Network

In this model, the algorithm reads the environment as steps in a discrete-time sequence where at each step, the model reacts to a given state by selecting an action from a predefined set of possible actions called action space. A reward is generated, and a new state is created; by iterating this interaction, each sequence or trajectory observation is suitable for expressing itself as a finite Markov decision process (MDP) (Fig. 1).

3.2 DQN-Based AQM System Design

In this section, we introduce the state, action, and reward function design for our AQM algorithm based on the DQN.

Step 1: Process of Selecting Action

With respect to the state of RL, we consider four elements: dequeue rate, enqueue rate, drop rate, and Avrg_Queue_Len. At each time step t , state s_t is defined as $s_t = \{\text{dequeue rate}, \text{enqueue rate}, \text{drop rate}, \text{and Avrg_Queue_Len}\}$ which is an input of multilayer perceptron (MLP) consisting of three hidden layers of 16–32–16 neurons for each layer. For selecting an action, the main Q-network is used, and it returns two probabilities as an output (drop/non-drop probability). To find better action on a particular state, we use explore/exploit strategy, which means that the agent takes action based on its own selection (exploit), or sometimes takes a random action uniformly based on a certain probability (explore). For the explore/exploit strategy, we are starting from a highly random probability of action. The exploring probability is set at 90 percent based on the round of the episode at the first episode of the network

simulation, and it diminishes to 0 percent through the episode. Section 3.2 explains the selection process for action.

Step 2: Reward Engineering

Actions are evaluated by a system called the reward function; it is critical to ensure a proper balance when designing the reward function and preventing an infinite packet drop or non-drop state by optimizing the trade-off between delays in queuing and drop rate. We refer to learn Queue's reward function as a baseline [6, 11, 12].

$$\text{Reward} = \text{clip}((\gamma * \text{delay_rerword}) + ((1 - \gamma) * \text{enqueue_reward}), -1, 1) \quad (1)$$

Reward systems consist of two components which are: delay_rerword and enqueue_reward for queuing delay and packet drop-rate, respectively, and the γ is scaling factor used to balance between delay_rerword and enqueue_reward.

delay_rerword is

$$\text{delay_rerword} = \text{desiredQueueDelay} - \text{current_delay} \quad (2)$$

when desiredQueueDelay is expected delay and the default is 0, and the current_delay is

$$\text{current_delay} = \text{Avrg_Queue_Len}/\text{dequeue_Rate} \quad (3)$$

where = Avrg_Queue_Len is current queue length in bytes, and dequeue_Rate is the average dequeue rate per sec

$$\text{dequeue_Rate} = \text{dequeueCounter}/\text{timeDiffrence} \quad (4)$$

where dequeueCounter is the packet numbers will not enter the queue, where timeDiffrence is set by user. enqueue_reward is described in:

$$\text{enqueue_reward} = (\text{min_delay} - \text{desiredQueueDelay}) * \text{enqueue_rate} \quad (5)$$

where min_delay is defined as:

$$\text{min_delay} = \text{avrgQueueLenInBytes}/\text{band_in_bytes} \quad (6)$$

where avrgQueueLenInBytes is the current size of queue used in the device measured in bytes, and the band is the data rate of the link connecting the two nodes, and enqueue_rate is defined as:

$$\text{enqueue_rate} = \text{enqueueCounter}/(_enqueueCounter + \text{dropCounter}) \quad (7)$$

where dropCounter is the number of dropped packets, and respectively, enqueueCounter is the enqueued packets.

Step 3: Training Process

Since we use DQN as our model, at each time stage t , the algorithm stores entities in replay memory in the form of tubules of $\text{et} = (\text{st}, \text{at}, \text{rt}, \text{st} + 1)$. When the number of replay memory experiences approaches the mini-batch size, the agent randomly selects samples of the memorable experiences in a consistent manner. In the first step, we initialize the model weights, which assign the weights of the layers from a Gaussian distribution [13] and minimize the loss using the optimizer Adam [14] to train the model.

Figure shows the flowchart of the DQN based AQM training process. In the flowchart, t_{curr} is current time step t , and C is the target update step to update the target network periodically (Fig. 2).

Fig. 2 Flowchart of DQN-based AQM training

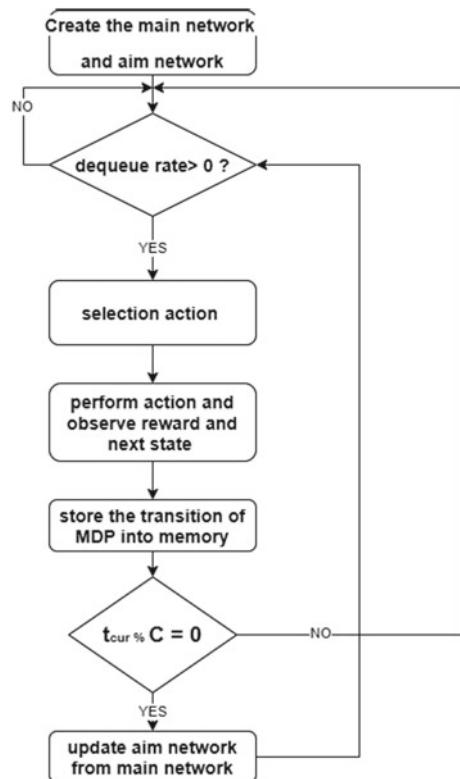


Table 1 Packet delay (RTT) of two algorithms

Algorithm	DQN-RED	RED
Average RTT in a millisecond	24.275	25.1083

Table 2 Drop rate of two algorithms

Algorithm	DQN-RED	RED
Average packets dropped per sec	3.18	3.28

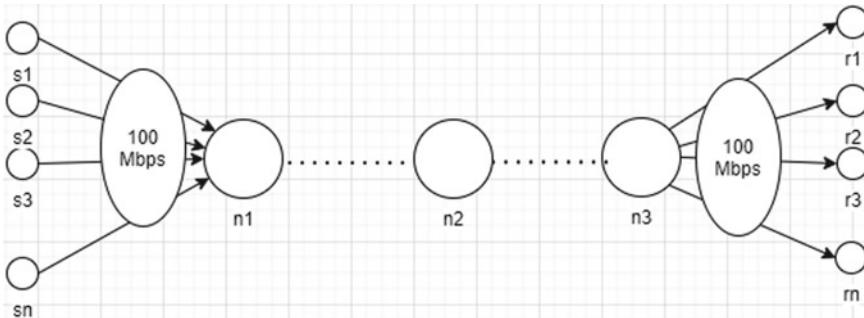


Fig. 3 Simulated network structure

4 Simulation Analysis

This section validates the validity and performance of the NS3 simulation experiment of the designed DQN algorithm, and the simulation uses the typical single-bottleneck network topology as shown in Fig. The network has n senders ($S_1 \sim S_n$), receivers ($r_1 \sim r_n$), and one router (n_2). The delay and bandwidth between sender (n_1) and (n_2) are 100 Mbps and 0.1 ms, and the bandwidth and delay between each receiver and (n_2) are 2 Mbps and 5 ms too. To compare, we analyzed the RED algorithm and DQN algorithm's queue length, throughput, delay, and packet loss rate under changing load, respectively. The performance of the algorithm, the simulation time is 120 s. Figure shows the queue length of the RED and the DQN algorithms. As can be seen in Tables 1 and 2, the average delay of the standard RED is larger than that of the DQN improved variant. So, the DQN algorithm reduces the drop probability and reduces the delay (Figs. 3, 4, and 5).

5 Conclusion

To boost the parameter settings of the RED algorithm, allow the algorithm to achieve improved network efficiency, the RED algorithm selects the correct parameters according to the learning process. This paper discusses an improved RED algorithm

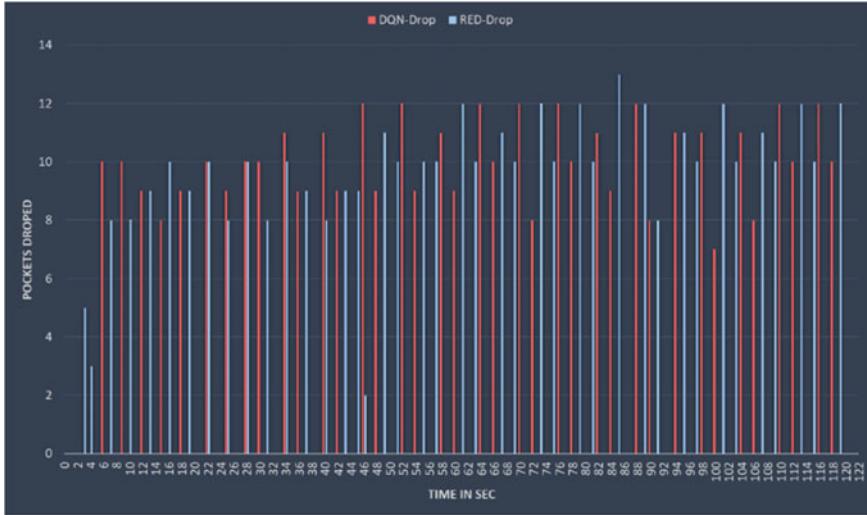


Fig. 4 Plotting drop rate/sec for each sec in the simulation



Fig. 5 Plotting pocket delay measured by the RTT for each sec in the simulation

known as DQN. The Q-learning algorithm is used in this method to pick the highest likelihood parameter for the decrease of packets. The parameters of the RED algorithm are immune to faults and can anticipate complex shifts in network networks. The learning structure gets the optimum network quality management system. This will tailor the highest likelihood of packet drop for the algorithm, ensuring better efficiency of the network while preventing congestion. The simulation tests validate

the DQN algorithm's benefits, which can be applied in the network to preserve reliability, minimize latency, increase the throughput, etc. The overall network output of the DQN algorithm is better than that created by the RED algorithm.

References

1. Athuraliya S, Low S, Li V, Yin Q (2001) REM active queue management. *IEEE Netw Mag* 15(3)
2. Kuhn N, Natarajan P, Khademi N, Ros D (2016) Characterization guidelines for active queue management (AQM). In: Internet engineering task force (IETF), RFC 7928. [Online]. Available: <https://tools.ietf.org/html/rfc7928>
3. Baker F, Fairhurst G (2015) IETF recommendations regarding active queue management. In: Internet engineering task force (IETF), RFC 7567. [Online]. Available: <https://tools.ietf.org/html/rfc7567>
4. Mohammadi M, Al-Fuqaha A, Sorour S, Guizani M (2018) Deep learning for IoT big data and streaming analytics: a survey. *IEEE Commun Surv Tutor*. [Online early access]. Available: <https://doi.org/10.1109/COMST.2018.2844341>
5. Sutton RS, Barto AG (2018) Reinforcement learning: an introduction, 2nd ed. MIT Press
6. Bouacida N, Shihada B (2018) Practical and dynamic buffer sizing using LearnQueue, *IEEE Trans Mob Comput*. [Online early access]. Available: <https://doi.org/10.1109/TMC.2018.2868670>
7. Bisoy SK, Pandey PK, Pati B (2017) Design of an active queue management technique based on neural networks for congestion control. In: 2017 IEEE international conference on advanced networks and telecommunications systems (ANTS). [Online]. Available: <https://doi.org/10.1109/ANTS.2017.8384104>
8. Vucevic N, Perez-Romero J, Sallent O, Agusti R (2007) Reinforcement learning for active queue management in mobile All-IP networks. In: 2007 IEEE 18th international symposium on personal, indoor and mobile radio communications. [Online]. Available: <https://doi.org/10.1109/PIMRC.2007.4394713>
9. Chen X, Zhang H, Wu C, Mao S, Ji Y, Bennis M (2018) Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning. *IEEE Internet Things J*. [Online early access]. Available: <https://doi.org/10.1109/JIOT.2018.2876279>
10. Xu Z, Tang J, Meng J, Zhang W, Wang Y, Liu CH, Yang D (2018) Experience driven networking: a deep reinforcement learning based approach. In: IEEE INFOCOM 2018—IEEE conference on computer communications. [Online]. Available: <https://doi.org/10.1109/INFOCOM.2018.8485853>
11. Gadicha AB et al (2021) *J Phys Conf Ser* 1963:012141
12. Bansal R et al (2021) *J Phys Conf Ser* 1963:012170
13. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, vol 9, pp 249–256. [Online]. Available: <http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf>
14. Kingma DP, Ba J (2017) Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980v9](https://arxiv.org/abs/1412.6980v9). [Online]. Available: <https://arxiv.org/abs/1412.6980>

Secure Smart Contract Based on Elliptic Curve in Property Exchange Applications Using Blockchain



Noor Sabah

Abstract As we live in a fast and rapidly developing society with goring businesses, more and more people chasing after the dream of a better financially independent life stray from the large corporations and organizations to form startups and small businesses, increasing the demand and need putting more pressure on employees of such institutes to prepare and manage contracts, papers, and verifications, the more businesses activities, means more needed employees to manage their assets by adding more to the humanly mistakes that may made. So, to reduce time and provide more confident, organized and secure labor new strategy has been employed. In this paper, blockchain technology has been adopted for secure smart contract that implemented it in a real-estate scenario. The main purpose is to create a distributed ledger, reducing the need for centralization procedures and adding more security and integrity by implementing elliptic-curve cryptography as a signature generation/verification and encryption/decryption protocol.

Keywords Smart contract · Blockchain · Elliptic curve · Hash function · Digital signature · Non-repudiation

1 Introduction

Smart contracts are useful frameworks for equal fulfillment of agreements between known and unknown persons. Contribute to smooth contact without any problems and minimum confidence requirement is usually implemented. Through incorporating the capacity of physically intelligent contracts into the digital world with blockchain technology. The best medium for digital contracts seems to be Blockchain. This idea led to the invention of intelligent contracts [1]. Smart contracts are self-executing systems that operate on the blockchain and are able to implement rules, effects, and estimates of increasing blockchain operation. In 1994, Nick Szabo formally introduced the idea of intelligent contracts. Smart contracts can take any type of

N. Sabah (✉)
University of Anbar, Anbar, Iraq

input data, perform input calculations based on the smart contract protocols, and can execute the decisions based on the current output conditions.

This groundbreaking principle automates the execution of contract agreements and improves openness without any broker or trustworthy third party (TTP), since each node of the blockchain complies with the smart contract protocols defined. All contracts are held in a linear sequence in blockchain. The participating nodes cannot, however, modify and eradicate the chances of assault by changing the contracts contained in the blockchain. Intelligent contracts allow both true and invalid transactions that may occur through a blockchain to be dealt with [2]. Blockchain technology has been established by cryptocurrency Bitcoin since 2008. A blockchain is a digital ledger which publicly records transactions after nodes are verified. In order to sustain publicly in a safe and trustworthy environment, the value of smart contract adoption of the blockchain technologies becomes a focus area for growth. Intelligent and irreversible contracts are trackable. All transaction details are stored in an intelligent contract and immediately enforced.

The nodes verify single transaction and cryptographic hash function secures single transaction. A transaction is connected by the previous hash value transaction. When the transaction has been implemented, one can change or adjust the ledger; however, the transaction can be accessed publicly, giving the network clarity. Blockchain technology and intelligent contract integration provide considerable flexibility for developing and designing and implementing certain real-world problems in less cost and in less time, without traditional third-party system [3].

In order to maintain the appropriate degree of confidence, blockchain technologies and smart contracts should be used. Events are stored as blockchain transactions that deliver highly secure, normal, trustworthy, and reputable knowledge [4]. Events are stored, and a variety of advantages can be provided by intelligent blockchain contracts that can be listed below [5, 6]. A smart contract is a self-executing contract which uses blockchain technology for the digital execution, verification, or facilitation of contract performance or negotiation. Since the blockchain technology is introduced to stable and decentralized systems, intelligent contracts can promote legitimacy of transactions between contracting parties without the need for third parties as seen in conventional contracts [7]. Suppose you want to sell or for renting your apartment to someone.

This technique helps to reduce the demands of any trusted third party in order to use these services. The intelligent agreement while at the same time safeguarding the system from malicious attacks, the agreements are nothing more than resident blockchain text programmers that can execute them. As the customer may enter into the intelligent contract arrangement by utilizing the blockchain technology's specific addresses [8]. You can simply post an intelligent contract on blockchain network when assuming want to sell or rent your apartment to someone.

2 Previous Related Works

Daniel [9] studies smart contracts' implementation by state-of-the-art blockchain technology that can serve as component technology for a computing paradigm as service-oriented computing (SOC) in the blockchain for foster reusing and increasing cost-effectiveness and describes the conceptual underpinnings of this new landscape are more integrated than one would expect and that smart contracts, to some extent, may indeed be interpreted as elementary pieces, that is, services, of a blockchain-based, service-oriented computing paradigm.

Terzi [10] describes the implementations of BC technology in two real-life supply chain scenarios. The first one intended to logging and tracing products by utilizing smart contracts to identify the ingredients of food products and how to uniquely identify the food product throughout its shipment from the factory to the customer who purchases it. The essential aspects of this route are the transparency of the process and the verification of transport. The second reasoning is intended to deal with the authentication works for users holding BC identities. A user's identity is essential to secure the network, allowing only permissioned parties to access it and perform the data.

Christidis [11] examine whether they give a good fit for the Internet of things (IoT) sector. Blockchains allow having a distributed peer-to-peer network where non-trusting members can interact with each other without a trusted intermediary in a verifiable manner. The paper reviews how this mechanism works and looks into smart contracts-scripts that reside on the blockchain, allowing for the automation of multi-step processes. The paper then describes how a blockchain–IoT combination facilitates sharing services and resources, leading to creating a marketplace of services between devices and automating several existing time-consuming workflows in a cryptographically verifiable manner. The paper concludes that the blockchain–IoT combination is powerful and can cause significant transformations across several industries, paving the way for new business models and novel, distributed applications.

Karamitsos [12] has studied the influence of smart contract with the various components for the implementation in case applied within estate industry and descript the advantages utilizing smart contract and blockchain technology for real estate are as: Different parties can modify database, trustless among entities and parties, advantage of disintermediation in case of the transactions can be independently verified and automatically validated, and transactions advantage especially that separate transactions between the parties seeking to improve the efficiency of the invoicing process.

Zyskind [13] proposed the architecture that intended to protect the personal data and ensures users own and control their data by a lightweight decentralized blockchain data management architecture. The proposed method implemented the off-chain data storage in order to improve the efficiency and utilizing the protocol that turns a blockchain into an automated access-control manager that does not require trust in the third party [14–16].

3 Blockchain Categories

Most governmental document systems are still based on physical records; some have been moved to digital recording using a centralized approach. A centralized system has multiple problems compared to a decentralized system.

Giving us the following advantages with blockchain compared to traditional systems:

1. Distributed systems allow the data to be mirrored across multiple nodes; fake, forged, and corrupted records can be detected or rejected quickly by comparing the data across nodes.
2. Smart contracts minimize the dependency on the third party for user authentication and information verification and allow the exchange of assets with minimum effort and high.

These systems can be compared and described according to their characteristics in terms of power processing, data transfer, DOS attack, etc.

As can be seen in Table 1, moving to decentralized systems is more beneficial than other systems. As a decentralized system that will explain the use of blockchains and smart contracts in a secure form.

4 The Proposed Smart Contract Model

As an example, to illustrate such a system's usage, the proposed system as a real-estate department can be described and explained in detail. Nodes are created in multiple trusted locations, whether it be government locations, banks, or any large corporation. A trusted node is placed with a party that benefits directly or indirectly from ensuring secure and fast transaction, like banks and real-estate firms.

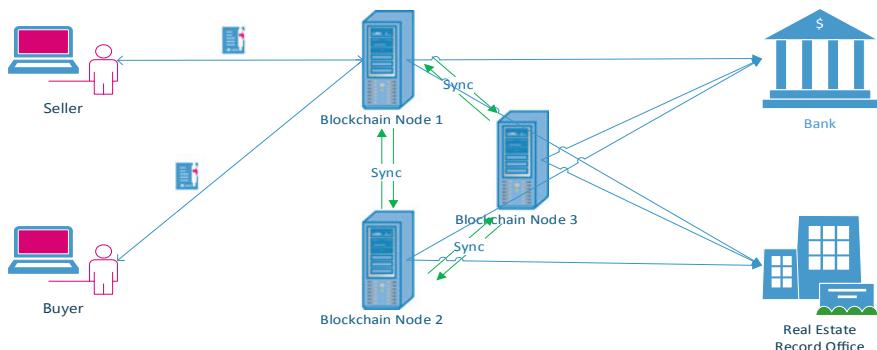
They all suffer if a record is corrupted or faked, so they all work together to keep the system safe. Users have communicated with nodes to create and sign contract, and the rest is done within the node, where the node verifies the information provided and does the necessary. Banks and real-estate offices communicate with nodes to facilitate the asset exchange. A simple scenario for the bank activities on the blockchain with signing a smart contract can be shown in Fig. 1.

Where each user must have a bank account and an account with the real-estate office to allow the exchange of assets. Also, it must have an account on the system with his information preset and verified. The following phases are the main steps for establishing the smart contract between the seller and buyers sequentially.

Note: Workflow can be seen in Fig. 2.

Table 1 Comparing centralized and decentralized system characteristics

Feature	Centralized	Decentralized
Data storage locations	Data kept one place	Data are copied to multiple places
Processing power	All the processing is done in one node, requiring large and powerful systems to handle all workloads	Processing power can be distributed among multiple smaller nodes
Networking, communication, and data transfer	All is handled by one location. Requires enormous infrastructure to handle all traffic	Traffic can be distributed to multiple locations
Delay and responsiveness	If a delay happens, it can compromise the whole system and reduce responsiveness	Traffic is routed to multiple nodes, reducing delay and increasing responsiveness
Physical intrusion attacks	Easier to protect from physical intrusion	Harder to protect multiple locations from physical intrusion
Denial of Service Attacks	If the central node cannot be accessed, the whole system fails	The system is mirrored to multiple points. If one fails, the rest can still operate
Data integrity Insurance	If data is changed maliciously, it is hard to detect and recover	If data is changed maliciously on one Node, it can be verified and repaired by comparing it to other nodes
Data redundancy	Off location, backups are needed to ensure data safety against corruption	Each node acts as a backup and no need for independent backups

**Fig. 1** Proposed smart contract system

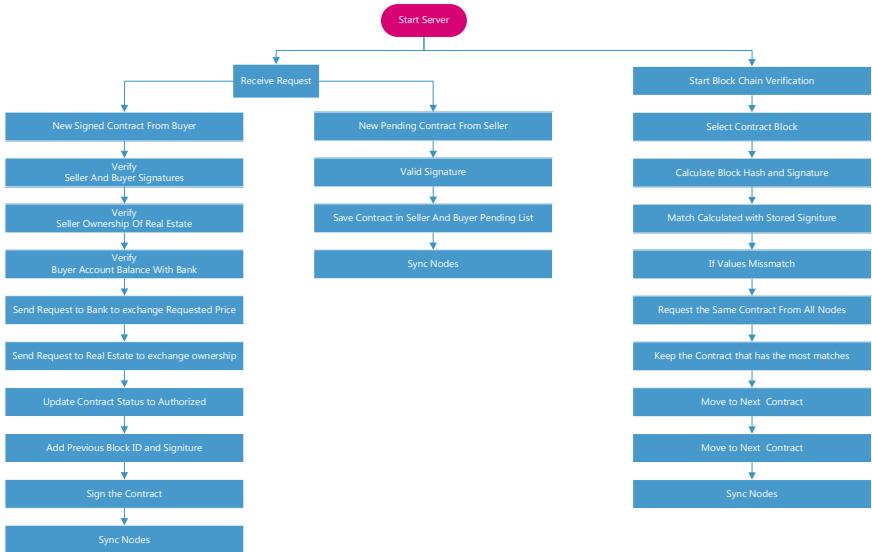


Fig. 2 Workflow for the proposed system

Registration Phase

User registration steps are also known as a wallet:

1. Choose ID and password
2. Fill bank account information
3. Fill real-estate account information
4. Generate Public/Private key pair Using Elliptic Curve
5. Encrypt Information with AES-256
6. Send information to be saved in Blockchain nodes.

Contract Creation

Seller Side

1. Connects to Nearest Node
2. Sign In to Blockchain Account
3. Create New Contract
4. Fill Information:
 - a. Seller ID
 - b. Seller Public Key
 - c. Seller Bank Account
 - d. Seller Real-Estate Account
 - e. Real Estate to Sell
 - f. Price Requested
 - g. Buyer ID

- h. Buyer Public Key
5. Sign Message Information with Elliptic Curve using Seller Private key
6. Save Signature in Message
7. Send Contract to Node.

Buyer Side:

1. Connect to Nearest Node
2. Connects to Nearest Node
3. Sign In to Blockchain Account
4. Request Pending Contracts
5. Open Selected Pending Contract Form
6. Verify Seller Signature
7. If the Contract Accepted Sign Message Information with Elliptic Curve using Buyer's Private key
8. Send Contract to Node.

Sing a Contract:

To sign a contract by seller or buyer:

1. Make sure all contract Information is filled
2. Convert Contract information to the serialized form
3. Calculate the Information Hash using Sha256
4. Generate User Signature by Encrypting the Hash using Elliptic Curve and the user's Private Key
5. Attach the public key to the Contract
6. Attach the Signature to the Contract.

Verifying a Contract

To verify the contract signature by the seller or buyer:

1. Make sure all contract Information is filled
2. Convert Contract information to the serialized form
3. Calculate the Information Hash using Sha256
4. Extract the other party public key attached to the Contract
5. Extract the other party signature attached to the Contract
6. Decrypt the Signature using the elliptic curve and the public key extracted earlier
7. Match the decrypted Hash with the generated Hash
8. If hashes match, the Contract is verified.

Sign a contract in the Trusted Node:

1. Convert Contract to the serialized form
2. Calculate Hash value using SHA256
3. Generate Signature by encrypting the Hash using the elliptic curve and the Node's Private key
4. Attach the Signature to the Contract.

Verify a Signed contract in Trusted Node:

1. Convert Contract to the serialized form
2. Calculate Hash value using SHA256
3. Decrypt the attached Signature using the elliptic curve and the Node's Public key
4. Match the calculated Hash and the extracted Hash.

Adding a Contract:

To add a contract to the blockchain in the trusted node:

1. **Verify Contract Assets**
 - a. Make sure all contract Information is filled
 - b. Verify seller signature
 - c. Verify buyer signature
 - d. Verify seller and buyer bank account ownership
 - e. Verify seller and buyer real-estate account ownership
 - f. Verify seller has ownership of the real estate included in Contract
 - g. Verify buyer has the bank balance to satisfy the requested price.
2. **Exchange Assets**
 - a. Send a request to the bank to exchange balance from buyer to seller
 - b. Send a request to the real-estate office to exchange ownership from seller to buyer
 - c. Verify the successful completion of the task
3. Mart the Contract status as authorized
4. Attach the ID and Signature of the previous Block in blockchain to the Contract
5. Sign a contract in the Trusted Node
6. Add the Contract to the blockchain
7. Sync the blockchain with other nodes.

Blockchain Verification in Trusted Node:

1. Set the processing index to the first Node in the blockchain
2. Get the block hash by Decrypting the block signature using Trusted Node's public key
3. Verify a Signed contract in Trusted Node:
4. If Values mismatch:
 - a. Request the same block from all nodes
 - b. Verify a Signed contract in Trusted Node:
 - c. Select the Contract with most matches as the authentic Contract
 - d. Update the Contract value
 - e. Go to Next index in blockchain.

5 Conclusion

1. Integrate the blockchain and smart contract to build a stable intelligent contract. So to gain tremendous versatility and to build and plan and to execute actual usage cases without the conventional third-party framework (TPP), with less time and with less expense.
2. The distribution of protection criteria that are an essential part of the deal is a substantially more efficient way to secure a smart agreement.
3. In addition to protecting the contact line or storage position, the introduction of protection criteria in a smart contract provides greater a guarantee of secrecy, validity, and legitimacy of electronically transmitted contracts.
4. A correctly protected device may be sued for preserving its own credibility. Because the program retains credibility, arrangements may be made to preserve legitimacy from the point of entry. Therefore, the document's validity depends on device stability. It is a compelling point that will be taken into consideration when developing protection devices.
5. The proposed approach provides continuous end-to-end protection during the entire life cycle of an electronic contract. Transaction management and automated signature systems must be provided in order to secure important online contracts.
6. Electronic signatures can be used to verify the validity of the document to be submitted using a digital illustration of the signature that cannot be checked. It is permanently attached, like handwritten signatures (wet signatures).
7. Public key electronic signatures (PKIs) provide solid security in the field of data validity, author credibility, and non-repudiation to deter forgery.

References

1. Buterin V (2014) A next-generation smart contract and decentralized application platform. White Paper
2. Szabo N (1997) The idea of smart contracts. Nick Szabo's Papers and Concise Tutorials
3. Mohanta BK, Panda SS, Jena D (2018) An overview of smart contract and use cases in blockchain technology. In: 2018 9th International conference on computing, communication and networking technologies (ICCCNT). IEEE
4. Hull R, Batra VS, Chee YM, Deutsch A, Health FT, Vianu V (2016) Towards a shared ledger business collaboration language based on data-aware processes. In: Proceedings of international conference on service oriented computing (ICSOC)
5. Marino B, Juels A (2016) Setting standards for altering and undoing smart contracts. In: International symposium on rules and rule markup languages for the semantic web. Springer, Cham
6. Saleem MA (2017) The impact of socio-economic factors on small business success. Geografia-Malaysian J Soc Space 8(1)
7. Nzuvu S (2019) Smart contracts implementation, applications, benefits, and limitations. J Inf Eng Appl 9(5):63–75

8. Gao Z, Xu L, Chen L, Shah N, Lu Y, Shi W (2017) Scalable blockchain based smart contract execution. In: 2017 IEEE 23rd international conference on parallel and distributed systems (ICPADS). IEEE, pp 352–359
9. Daniel F, Guida L (2019) A service-oriented perspective on blockchain smart contracts. IEEE Internet Comput PP(99)
10. Terzi S, Zacharaki A, Nizamis A, Votis K (2019) Transforming the supply-chain management and industry logistics with blockchain smart contracts. In: The 23rd Pan-Hellenic conference
11. Christidis K, Devetsikiotis M (2016) Blockchains and smart contracts for the internet of things. IEEE Access
12. Karamitsos I, Papadaki M, Barghuthi NBA (2018) Design of the blockchain smart contract: a use case for real estate. J Inf Secur 9:177–190
13. Zyskind G, Nathan O, Pentland A (2015) Decentralizing privacy: using blockchain to protect personal data. In: IEEE security and privacy workshops, San Jose, pp 180–184
14. Sasank TS et al (2021) J Phys Conf Ser 1879:032124
15. Meshram C, Ibrahim RW, Obaid AJ, Meshram SG, Meshram A, Abd El-Latif AM (2020) Fractional chaotic maps based short signature scheme under human-centered IoT environments. J Adv Res. <https://doi.org/10.1016/j.jare.2020.08.015>. ISSN 2090-1232
16. Oabid AJ, AlBermany S, Alkaam NO (2020) Enhancement in S-Box of BRADG algorithm. In: Solanki V, Hoang M, Lu Z, Patnaik P (eds) Intelligent computing in engineering. Advances in intelligent systems and computing, vol 1125. Springer, Singapore. https://doi.org/10.1007/978-981-15-2780-7_80

Performance Investigation of Short Channel Impacts and Analog/RF Figure of Merits (FOMs) of SOI-FinFET



Nishant Srivastava and Prashant Mani

Abstract FinFETs are significant and at front runner among integrated circuits (ICs) because of their significant scalability and reduced impact of short channel effects (SCEs). The present paper provides a study on the device structure and traits of device, namely silicon on insulator (SOI-FinFET). On optimizing the Fin aspect ratio value which is AR = Fin height/Fin width, the goal is to ameliorate short channel effects, self-heating problems as well as to examine the width quantization effect. The considered approximations are related to SOI-FinFET structure as well as internal capacitance coupling. Various parameters impact such as channel doping attenuation or charge trapping within insulator is not accounted, since the present paper focuses to compute the effect of geometrical dimensions and other relevant factors upon SOI-FinFET behavior and operation. The device channel is kept either undoped or gently doped. The electrical performance metrics like DIBL, subthreshold slope are extracted using Silvaco TCAD simulator. The analog/RF figure of merits (FOMs) like transconductance (gm), output conductance (gd) and gate capacitance are examined for SOI-FinFET via 3-D ATLAS device simulation.

Keywords Bulk · Silicon on insulator (SOI) · Subthreshold swing · Ultra-low-power circuit · FinFET · Short channel impacts

1 Introduction

From past decades, the shortening of MOSFET channel length is the key to enhance device performance as well as power efficiency. The hostile scaling led to concerned short channel effects (SCEs) causing a significant rise in off-state current as well as device standby power. Additionally, the channel length scaling led to weaker gate control upon the device channel. To bring an effective gate control, addition of second gate aids existing gate, thus placing the proposed device upon insulator. Thus, device body will be the bottom gate. These two options demonstrate a better performance to

N. Srivastava · P. Mani (✉)

SRM Institute of Science and Technology, NCR Campus, Modinagar, Ghaziabad 201204, India

scale down the conventional MOSFETs. The usage of SOI structure within the device channel scaling enhances device speed and ON-state current. The utilization of SOI attenuates the source and drain junction capacitances with exclusion of depletion regions extending inside substrate. Such substrates cause dielectric isolation, thus lowering the parasitic effects in bulk devices. One of the authors worked on usage of back-gated and thin SOI with buried oxide (BOX) to reduce threshold voltage and lowering voltage operation [1]. A new MOSFET with upper drift region double step partial SOI (UDDS-PSOI) is proposed to handle electric field impact in vicinity of drain [2]. It will boost the device's breakdown voltage. To improve both static as well as dynamic power, a FinFET device based on SRAM architecture with back-gate biasing is provided which also increased SRAM efficiency, device reliability and stability [3]. Additionally, it improved the write/read operation [3].

Scaling the CMOS device into millimeter regimes raises the power dissipation manifold because of rising leakage current. The leakage current increases due to reduced threshold voltage. Further, the scaling reduces drain-induced barrier lowering (DIBL), temperature impact, narrow width effect, impact of gate-induced drain leakage, as well as gate oxide tunneling [3, 4]. The traditional MOSFETs suffer from high parasitic capacitance values, more latch-up and a lot of surface mobility degradation because of direct scaling. The suggested solutions include: (i) operating device on low values of temperature, (ii) using fully depleted silicon on insulator (FDSOI) [5]. In comparison, the thin film SOI-MOSFETs causes reduction in SCEs, with excellent latch-up immunity and minor DIBL effect [6]. It is the thinning of SOI film which controls the SCEs over the junction. The SOI-MOSFETs suffer from floating body problems due to built-up charges inside the silicon film which causes a breakdown, kink impact, bipolar transistor action as well as self-heating effects [7]. The SOI device has a lower carrier transport efficiency because of heterogeneous distribution of electric fields within the channel having peak value near the drain causing hot-carrier effect. Over these, the SOI-FinFET device is advantageous: as they (i) operate at a very low supply voltage [8], (ii) show lower threshold voltage sensitivity to gate length [9], (iii) show suitability at low standby power (LSTP) applications [10], (iv) provide raised ion and reduced I_{off} value [11], (v) lower the subthreshold swing, and (iv) possess lower SCEs and leakage levels in FinFETs [12].

To resolve width quantization issue in FinFET, a new bottom part FinFET is designed in present paper. And, lowering of active fin height tends to improve SCE impact, reduce the power dissipation—static and dynamic and alleviate self-heating problems. The paper is organized as follows: In section II, the device design and structure are discussed. In section III, the simulation and analysis of the device are discussed with a focus on I_d vs. V_{gs} characteristics and analog and RF performances of an SOI-FinFET. Finally, the paper conclusion is drawn.

2 Device Design and Structure

Figure 1a illustrates cross-sectional schematic of SOI-FinFET's architecture. The

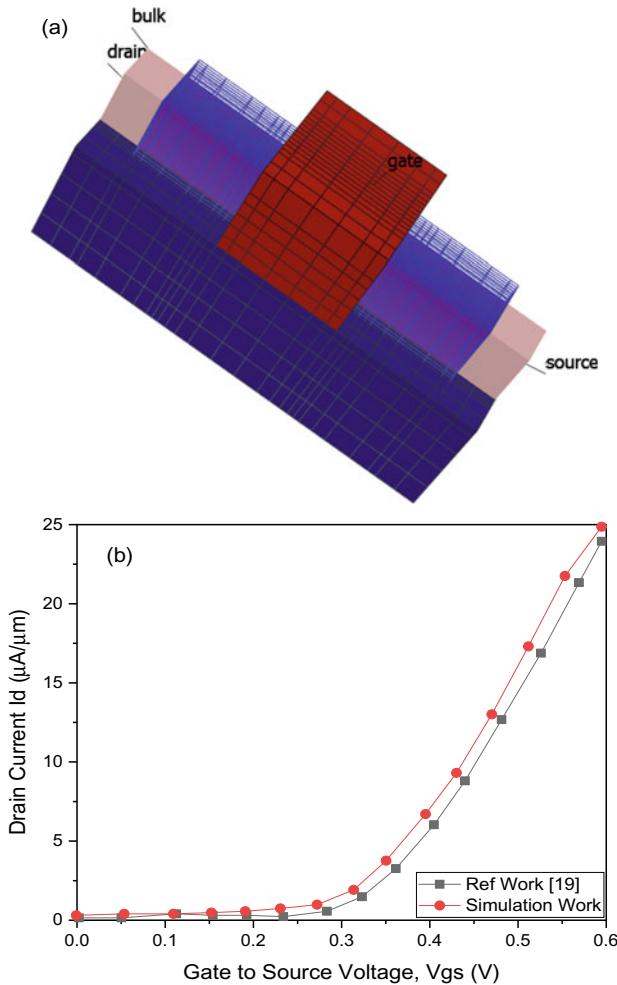


Fig. 1 **a** Cross-sectional diagram of silicon on insulator FinFET **b** experimented drain characteristic

proposed device gate is fabricated, whereas source, as well as drain, are the implants. The undoped channel ceases the device from experiencing coulomb scatter because of impurities as well as excessive RDF. Consequently, the device carrier mobility becomes higher compared to planar-MOSFETs [13]. A very lightly doped channel is fabricated because of device-to-device threshold voltage changes. The FinFET geometrical characteristics give an increased gate capacitance for similar oxide thickness in comparison with planar MOSFET [14]. The device gate geometry has a significant role to control FinFETs performance metrics. The 'Fin width' represented as W_{fin} , as illustrated in Fig. 1. The lower values of W_{fin} can provide a better gate control upon channel due to a smaller subthreshold swing as well as lowering of drain-induced barrier lowering (DIBL). Numerous values of Fin aspect ratios

(AR), where AR = Fin height/Fin width, are examined, to compute its impact upon the device channel controllability, SCEs, on-state current (I_{on}), as well as buried-insulator-induced barrier lowering (BIIBL). The value of DIBL rises on getting AR < 1.5 [15]. In comparison with the traditional MOSFET, a vertical arrangement in FinFET device makes it look a compact device. On the other hand, the design indicates fin width should be near half value of channel length. Thus, to keep optimized, AR gives a noteworthy challenge to control processing besides designing flexibility [16–20]. The present paper focuses on analysis of the SOI-FinFET device, its capacitances and internal structure, respectively.

A 3-D schematic representation of single material SOI-FinFET structure is demonstrated in Fig. 1a. The dimensions of the device are as follows: Gate length (LG), fin thickness (TFin) and fin height (H_{fin}) are 26 nm, 10 nm as well as 15 nm. The value of bulk silicon substrate is considered as $1 \times 10^{15} \text{ cm}^{-3}$ for p-type doped, value of source/drain regions is $1 \times 10^{20} \text{ cm}^{-3}$ for n-type doped, and active and inactive fin doping are p-type of $1 \times 10^{15} \text{ cm}^{-3}$ and $1 \times 10^{17} \text{ cm}^{-3}$. The simulations of the device are carried out in 3D-ATLAS simulator at temperature of 300 K. Basically, the device simulations count drift-diffusion transport model and quantum correction within inversion layer. A Lombardi constant voltage temperature (CVT) model is utilized to study an impact of transverse field in addition to doping as well as temperature-dependent mobility metrics. Further, an impact of surface scattering as well as acoustic phonon scattering is considered. Additionally, the Shockley–Read–Hall (SRH) model, carrier recombining impact as well as band-to-band Auger effect are accounted for determining the carrier lifetime and density. The narrowing of silicon bandgap evaluates inherent carrier concentration within heavily doped regime. Both Newton as well as Gummel techniques are accessed simultaneously to compute numerical solutions. Other various parameters considered are: ratio of source/drain length = 40 nm, doping density $N_D = 10^{20} \text{ cm}^{-3}$, supply voltage (V_{dd}) is 0.7 V and gate work-function = 4.85 eV.

3 Simulation and Analysis

The analysis carried in the following section portrays how SOI-FinFET characteristics can be optimized for certain design goals with an appropriate selection of geometric dimensions. Additionally, different structural factors for evaluation and improvement of SOI-FinFET subthreshold swing are identified. The thickness of buried oxide marks a trade-off between thermal and electrical metrics of device. The thinner BOX leads to better heat dissipation; deep electric field penetration lowers down depletion layer to remove parasitic capacitance, generating a higher speed and reducing the power consumption. On contrary, the thinner BOX raises parasitic coupling capacitance in substrate.

3.1 ID Versus VGS Characteristic

SOI-FinFET design demands an analysis of its dimensions to reduce SCEs, DIBL and control the subthreshold slope. To acquire this, the aspect ratio needs to be in suitable range. The present section investigated SOI-FinFET device I-V characteristics and DIBL in subthreshold regions. The domination of drain current over diffusion at subthreshold region illustrates device switching ability. Another concern is to manage the leakage generally for low-power applications.

Figure 2 represents device *I*-*V* characteristics at different aspect ratios (Ars). The leakage current reduces on increasing the AR value. It gives chance to have an optimal AR to achieve lower SS and leakage values.

The paper further studied an impact of SOI-FinFET AR ($H_{\text{fin}}/\text{fin}/W$) upon DIBL (difference in threshold voltage fetched at $VDS = 25 \text{ mV}$ and $VDS = 0.5$). It is further normalized via difference in drain voltage. The DIBL is evaluated via Eq. (13), such that V_{th} (threshold voltage) for the geometric AR. Figure 3 depicts DIBL of $59/\text{mV}$ at AR value of 1.5. It gets further reduced by increased AR value. From Fig. 3, it can be demonstrated that lowering AR reduces the impact of DIBL. On engineering SOI-FinFET architecture, optimization of thickness and length of channel, fin height and fin thickness is a fundamental metric to raise ON-current and lower leakage current. It can be acquired via superiority in electrostatic control upon channel. The device results reveal that AR tends to be modified further for a high-performance and lower-power demanding subthreshold applications. With $\text{AR} > 2$, the active area heightens, and thus, the device driving capability raises via better electrostatic integrity and a larger vertical channel. The device capacitance and geometrical analysis reveal that a rise in the aspect ratio H_{fin}/W from value 2-to-3 tends to raise conductivity levels

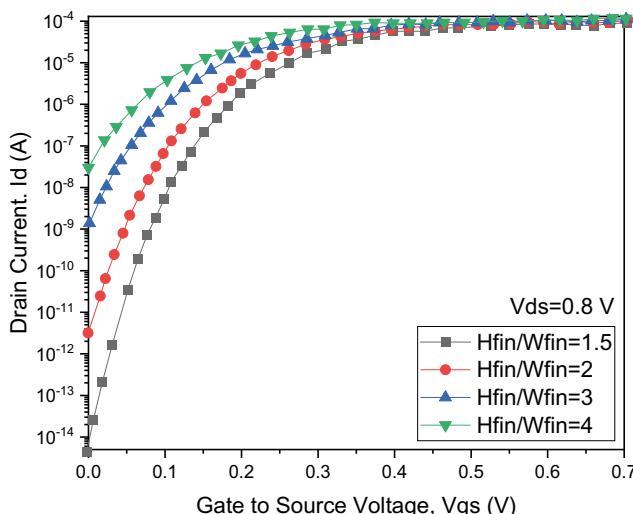


Fig. 2 $I_d - V_{\text{gs}}$ characteristics of the SOI-FinFET at different value of $H_{\text{fin}}/W_{\text{fin}}$ ratio

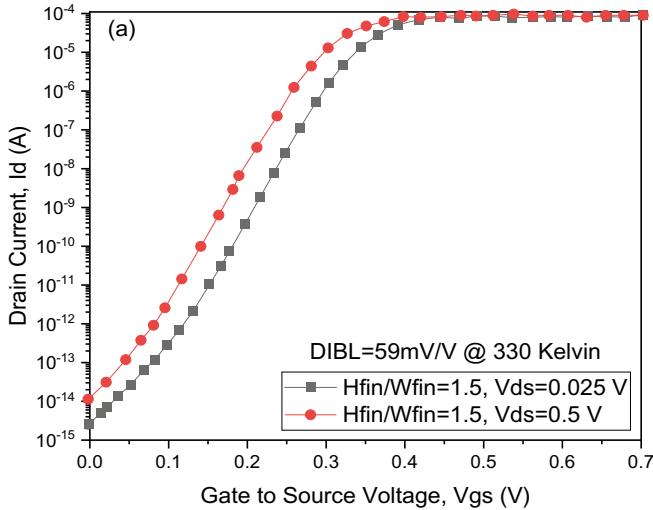


Fig. 3 $I_d - V_{gs}$ characteristics of SOI-FinFET showing DIBL dependency on drain current at different aspect ratios (AR = 1.5, 2, 3, 4)

of device and lower down DIBL to give an outstanding subthreshold. It resembles to a wider channel close to vertical gates to improve the electrostatic of the device. It is a higher aspect ratio that lowers DIBL and subthreshold swing. Hence, FinFET device structure ($AR > 2$) exhibiting significant characteristics which owe to a more uniform potential inside channel [17].

The authors showed that tall fins lower SCEs and demonstrate better performance. It concludes that FinFET electrostatics can further be enhanced after a reduction in the fin width [17–20]. The tall and narrow fins show excellent subthreshold presentation at ideal subthreshold slope and DIBL. The higher aspect ratio for tri-gate FinFETs is constructed with fin widths < 5 nm at AR of 13. The proposed device reveals an outstanding performance with decrease off-state leakage and increased electrostatic gate control over the channel. The AR tuning and trade-offs are summarized in terms of area/layout efficiency, complexity in manufacturing fins uniformly with narrow width and design efficiency. The FinFETs at higher AR show superiority in layout efficiency at lower area penalty. One of the studies established the result by concluding that high values of fin AR can provide layout efficiency compared to planar MOSFET. The narrower width with higher aspect ratio can control channel in a better manner, thus reducing leakage (W_{fin} is scaled to prevent SCEs). A wider transistor (i.e., large W_{fin}) can provide more current. However, smaller W_{fin} can efficiently suppress the off-state leakage current values. Thus, the FinFET efficacy gets raised on associating different operative channels for multi-parallel fin fusion.

3.2 Analog and RF Performances of an SOI-FinFET

The transconductance (gm) variation with drain current of SOI-FinFET at $V_{ds} = 0.7$ V is shown in Fig. 4. Transconductance computes gain of amplifier. As Aspect ratio increases gm increases. This increase in gm is result of rise in V_{ds} . It is perceived higher $gm \sim 78 \mu\text{S}$ for aspect ratio = 4.

Next, the $G_d - V_{ds}$ curves are acquired for different W_{fin}/L_g values and are plotted in Fig. 5. The output conductance (gd) variation with drain to source voltage (V_{ds}) of SOI-FinFET at $V_{GS} = 0.5$ V is depicted in this figure. At higher V_{gs} values, the device has higher variation in terms of G_d values. This increases the device heating and SCEs. The device persists in deep triode region for $V_{gs} = 1$ V with the highest drain current obtained for aspect ratio 4. For $V_{gs} = 0.5$ V, high ID is obtained for a higher aspect ratio because of the increase in H_{fin} . As seen from the output appearances, drain current is rising with increase in aspect ratio. Hence, it is beneficial in CMOS analog circuits for high gain. Thus, the device V_{gs} values need to be lowered down to reduce the heating and SCEs impact, which in turn reduces the width of the device. Although C_{gg} does not influence the DC execution fundamentally, it directly impacts the AC execution just like the delay and switching time of the circuit (Table 1).

The gate capacitance (C_{gg}) is an association of gate-to-source capacitance (C_{gs}) and gate-to-drain capacitance (C_{gd}). The metric gd is plotted with respect to gate-to-source voltages SOI-FinFET at $V_{ds} = 1.0$ V as demonstrated in Fig. 5. It is shown that the metrics C_{gs} and C_{gd} raise with an increment in V_{gs} value until saturation. In Fig. 6, it is shown that both C_{gs} and C_{gd} increase the AR via increasing active fin height. The higher fringing field lines originating from gate edges increase with an

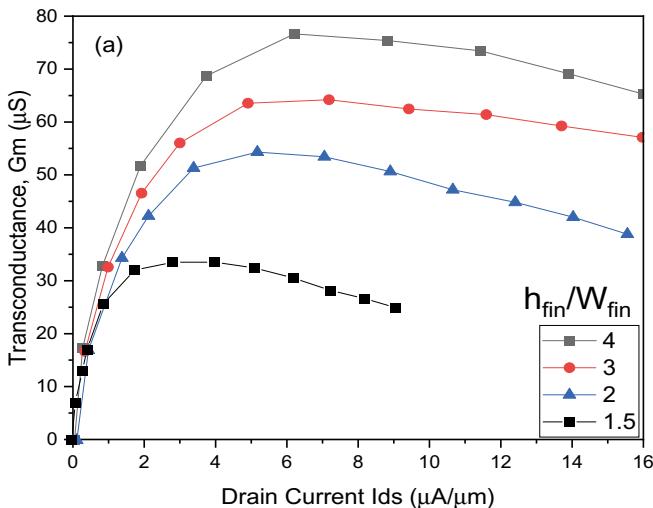


Fig. 4 Transconductance (Gm) characteristics of SOI-FinFET for different aspect ratios (AR = 1.5, 2, 3, 4)

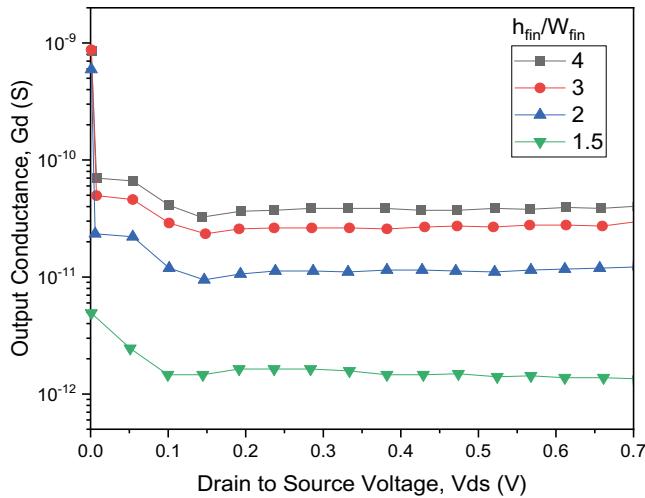


Fig. 5 Plot of device drain conductance (g_d) versus drain voltage (V_{ds}) at different values of aspect ratio (1.5, 2, 3, 4)

Table 1 h_{fin} and W_{fin} ranges of 3D SOI-FinFET for simulation

S. no.	Device parameters (H_{fin}/W_{fin})	Transconductance values (μS)	Output conductance (S)
1	4	76.65906	7.01e-11
2	3	63.54067	4.96e-11
3	2	54.29643	2.34e-11
4	1.5	25.58153	2.45e-12

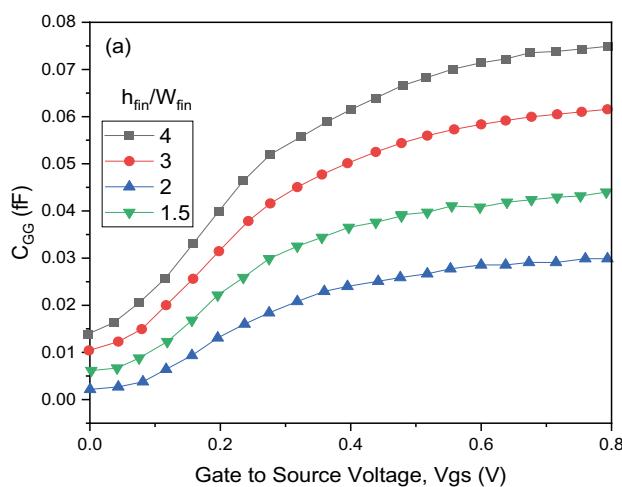


Fig. 6 Total gate capacitance (C_{gg}) versus (V_{gs}) at different vale of H_{fin}/W_{fin} ratio

increase in fin height. The fin structures reduces short channel effects in the device and smaller gate capacitance, resulting in improved delay in the circuit.

4 Conclusion

The structure, as well as internal capacitances of proposed SOI-FinFET device, is studied extensively in the present work. The electrical and analog/RF performance metrics at different aspect ratios are studied. Also, the DIBL of SOI-FinFET device is studied to propose an aspect ratio analysis for approximating subthreshold swing related to capacitive coupling model. The influence of the silicon bulk thickness, BOX and the aspect ratio on proposed device is also demonstrated. It is found that devices with a higher aspect ratio ($H_{\text{fin}}/W_{\text{fin}}$) have an improved drain-induced barrier lowering (DIBL) value. Our results depict that DIBL of proposed device is $\sim 20/\text{mV}$ that is attained at AR $\approx 3/\text{higher}$. In future studies, a complete device model can be developed after analysis of other device parameters. The present study analysis can help to analyze device suitability for designing ultra-low-power circuits. On determination of the optimum values of the aspect ratio equal to or greater than 3 is better for analog as well as RF circuit applications like capacitive parameters, transconductance and output conductance.

References

1. Choi YK, Asano K, Lindert N, Subramanian V, King TJ, Bokor J, Hu C (1999) Ultra-thin body SOI MOSFET for deep-sub-tenth micron era. In: International electron devices meeting 1999. Technical Digest (Cat. No. 99CH36318). IEEE, pp 919–921
2. Hisamoto D, Lee WC, Kedzierski J, Takeuchi H, Asano K, Kuo C, Hu C (2000) FinFET-a self-aligned double-gate MOSFET scalable to 20 nm. IEEE Trans Electron Devices 47(12):2320–2325
3. Morris DH, Avci UE, Young IA (2019) Intel Corp, U.S. Patent Application 15/992, 080
4. Wadhwa G, Raj B (2019) Design, simulation and performance analysis of JLTFT biosensor for high sensitivity. IEEE Trans Nanotechnol 18:567–574
5. Hisamoto D, Lee WC, Kedzierski J, Takeuchi H, Asano K, Kuo C, Hu C (2000) FinFET-a self-aligned double-gate MOSFET scalable to 20 nm. IEEE Trans Electron Devices 47(12):2320–2325
6. Wadhwa T, Kakkar D, Wadhwa G, Raj B (2019) Recent advances and progress in development of the field effect transistor biosensor: a review. J Electron Mater 48(12):7635–7646
7. Mani P, Pandey MK (2012) Silicon on insulator MOSFET development from single gate to multiple gate. Int J Adv Res Comput Sci Softw Eng 2(6)
8. Shrivastava M, Baghini MS, Sachid AB, Sharma DK, Rao VR (2008) A novel and robust approach for common mode feedback using IDDG FinFET. IEEE Trans Electron Devices 55(11):3274–3282
9. Park T-S, Cho HJ, Choe JD, Han SY, Park D, Kim K, Yoon E, Lee J-H (2006) Characteristics of the full CMOS SRAM cell using body-tied TG MOSFETs (bulk FinFETs). IEEE Trans Electron Devices 53(3):481–487

10. Chaudhry A, Kumar MJ (2004) Investigation of the novel attributes of a fully depleted dual-material gate SOI MOSFET. *IEEE Trans Electron Devices* 51(9):1463–1467
11. Subramanian V, Mercha A, Parvais B, Loo J, Gustin C, Dehan M, Decoutere S (2007) Impact of fin width on digital and analog performances of n-FinFETs. *Solid-State Electron* 51(4):551–559
12. Narendra V, Rai S, Tiwari S (2016) A two-dimensional (2D) analytical surface potential and subthreshold current model for underlap dual-material double-gate (DMDG) FinFET. *J Computational Electron* 15(4):1316–1325
13. Narendra V, Rai S, Tiwari S, Mishra RA (2016) A two-dimensional (2D) analytical subthreshold swing and transconductance model of underlap dual-material double-gate (DMDG) MOSFET for analog/RF applications. *Superlattices Microstruct* 100:274–289
14. Agarwal A, Sharma RL, Mani P (2021) Simulation and performance analysis of electrical properties of nano scale surrounding gate MOSFET. *Ann Romanian Soc Cell Biol*, pp 2102–2110
15. Liu Y, Masahara M, Ishii K, Sekigawa T, Takashima H, Yamauchi H, Suzuki E (2004) A highly threshold voltage-controllable 4T FinFET with an 8.5-nm-thick Si-Fin channel. *IEEE Electron Device Lett* 25(7):510–512
16. Pei G, Kedzierski J, Oldiges P, Ieong M, Kan E.C.-C. (2002) FinFET design considerations based on 3-D simulation and analytical modeling. *IEEE Trans Electron Devices* 49(8)
17. Wadhwa T, Wadhwa G, Bhardwaj TK, Kakkar D, Raj B (2020) Design and performance analysis of symmetrical and asymmetrical triple gate dopingless vertical TFET for biorecognition. *Silicon*, pp 1–9
18. Yanagi SI, Nakakubo A, Omura Y (2001) Proposal of a partial-ground-plane (PGP) silicon-on-insulator (SOI) MOSFET for deep sub-0.1- μm channel regime. *IEEE Electron Device Lett* 22(6):278–280
19. Lolivier J, Widiez J, Vinet A, Poiroux T, Dauge F, Previtali B, Deleonibus S (2004) Experimental comparison between double gate, ground plane, and single gate SOI MOSFETs. In: Proceedings of the 30th European solid-state circuits conference (IEEE Cat. No. 04EX850). IEEE, pp 77–80
20. Subramanian V, Mercha A, Parvais B, Loo J, Gustin C, Dehan M, Collaert N, Jurczak M, Groeseneken G, Sansen W, Decoutere S (2007) Impact of fin width on digital and analog performances of n-FinFETs. *Solid State Electron* 51(4):551–559

IoT-Based Surveillance and Face Detection BOT



Veral Agarwal, Nipun Tyagi, and Rachit Patel

Abstract In this paper, an IoT-based surveillance and face detection bot is presented. Security and surveillance are an important aspect in today's world. For more security, people tend to install a multiple camera, owing to which the cost increases significantly which also does not ensure required security. In order to overcome this problem, an IoT-based surveillance and face detection bot named "Vijayata" has been developed which promises to give more security and comfort to users. Vijayata gives users a 360-degree surveillance with intruder detection system (INDS) using artificial intelligence, Internet of things, and computer vision. The bot movement is controlled, and surveillance is monitored on its cloud-based application. This will reduce the number of cameras to be installed and also will increase the security of the specific area by using face detection and numerous sensors. Security is the vital concern for military, home, and industrial establishments. Vijayata will help in providing solutions to these establishments using surveillance and detecting intruders, covering the major problems of security.

Keywords Artificial intelligence · Internet of things · Face recognition · Image processing · Intruder detection · Surveillance bot · Cloud application · Web application

V. Agarwal
Nagarro Software Pvt Ltd, Gurgaon, Haryana 110038, India

N. Tyagi
Cisco Systems India Pvt Ltd, Bangalore, Karnataka 530068, India

R. Patel (✉)
Department of Electronics & Communication Engineering, ABES Institute of Technology,
Ghaziabad, Uttar Pradesh 201009, India

1 Introduction

Surveillance is the basic technique for the close observation of any person entering in any specific area; for that, we need to plant the electronic equipment to detect the presence of any Intruder. The surveillance can be of many types like computer, telephones, cameras, biometric, and RFID. Surveillance can be used by many private and government organizations for monitoring the wide range of activities [1–3].

CCTV has grown million dollars business today, and daily many cameras are being installed in every nook and corner for more visibility and detection of intruders. Billions of rupees were spent on new technologies related to the intruder detection system (INDS). In recent years, CCTV is introduced in various fields [4–6]. Furthermore of CCTV is establishing constantly for more accurate, sharp, crisp, saturated visualization of incidents and situations. Nowadays, CCTV comes with 1080pixel with dual motor head that is similar to Vijayata. However, Vijayata comes with added advantage of mobility. It comes with better placement of cameras [7]. The planted CCTVs are not much effective. It often comes across issues of bad visibility of face which cannot help in finding the person using that footage. To resolve these problems, Vijayata helps by performing many operations for detecting the intruder's face and notifying the user immediately. IT-based intelligent-service robots are advanced now as an alternate to the traditional work robot [8, 9]. During paper, a real-time cloud-based surveillance bot is going to be developed using Internet of things (IoT) with image processing and cloud-based application for interface of Vijayata. The proposed bot is capable of 360-degree surveillance and taking pictures of specific area. The bot movements can be controlled, and the surveillance will be monitored using its application/web server remotely [10]. Cloud is used for the storage and accessing of the data and bot. Vijayata is camera equipped robot with IoT that provides and actively manage communication with user. The interface for communication is application/web server for ease of use [10, 11]. It can be used for surveillance of areas like warehouse, factories, homes, offices, and military. Below is the functional specification of Vijayata. The bot has Raspberry Pi camera that is mounted on arrangement of servo motors that provide the 360-degree rotation and vertical movement to the camera to detect face of the person. Here in Vijayata sensors also play an important role by providing values that define the person which is detected or not [9].

2 Components

This bot consists of major components as follows, and their mechanism is defined in their respective sections.

2.1 *Raspberry Pi 4*

BOT is controlled and monitored remotely to look at the activity and obtain notifications when motion is detected. The photographs and videos are transferred on to a cloud server, when the cloud isn't available then the info is stored locally on the Raspberry Pi and sent when the connection resumes. Therefore, advantages like these make BOT ideal for surveillance [2, 3]. BOT does human detection by using Raspberry Pi Cam (RPI-Cam) with open-source computer vision (OpenCV) software which handles the image processing, sends captured pictures is that it breaks the video per/frame, and does analysis with the data stored in the database via Wi-Fi [4].

2.2 *Raspberry Pi CAM*

BOT is installed with Raspberry Pi camera, which will capture the photographs and video of the area which is to be surveillance. These captured images and videos are transferred to Raspberry Pi for processing. The camera can rotate 360 degrees with the help of servo motor, which will ensure more security and efficient than that of traditional surveillance devices used.

2.3 *Esp32*

The ESP32 developer board is used for interfacing the sensors we used in this bot, 1 A relay is also attached to ESP32 for switching on/off the Raspberry Pi 4 for power consumption reduction.

2.4 *Sensors, Motors, and Drivers*

The sensors and drivers are listed as follows with their mechanism in bot:

1. Passive infrared sensor (PIR): It is used in bot for human presence detection.
2. Ultrasonic Sensor: calculating the space supported the time required and here it is using for object detection. We measure the precise distance from our bot, and this will be calculated supported by this formula:

$$\text{Distance} = \frac{1}{2} T \times C \quad [C = \text{Speed of Sound and } T = \text{Time}]$$

3. Motor Driver- We are using L298N in bot for controlling the motors, and those are responsible for the movement of bot that is controlled by the web/mobile application.
4. Servo Motor: A servo motor is a device which will push or rotate an object with great precision. Here we are using MG995 servo motor which comes with a metal shaft that offers you more torque and provides 10 kg/cm at 4.8 V and 12kgcm at 6 V. We are attaching this with a camera for the vertical motion for detecting the face of the person.
5. Stepper Motor: It is used for the 360-degree rotation of the camera that is mounted on the arrangement of servo and stepper.

3 Working Mechanism and Organization

This system comprises of four distinctive operations that work altogether by sharing immediate live streaming data continuously and working upon that.

3.1 Bot Movement

The bot has been driven by the user, wherever the user wants to surveillance from one location to another. The DC motors are responsible for the bot movement, and these motors are controlled by the motor driver. Motor driver used is L293d which can control the speed of motors by adjusting the voltage level and control motors according to the user's command.

3.2 Face Detection Mechanism

The face detection mechanism can be understood by the sequence diagram in Fig. 1; in the sequence diagram, we are trying to show how bot detects the face of any person immediately and send alert to user.

Face detection sequence diagram: According to this diagram, initial values are from PIR and ultrasonic sensor to ESP32. Ultrasonic measures distance of the object from bot and PIR sensor detects radiated heat from human body. If any person is detected, then it starts moving camera toward face of the person and camera sends video continuously to Raspberry Pi. Further that video will be processed and analyzed if the detected person is guest or intruder from user database which is collection of images entered by user. If the detected person is intruder, the notification will be generated to the user immediately else it starts giving 360-degree surveillance as per user's request.

Figure 2 shows flow of working of face detection. When the BOT is in power on mode, it must ensure that Internet should be in connected state, if not one should

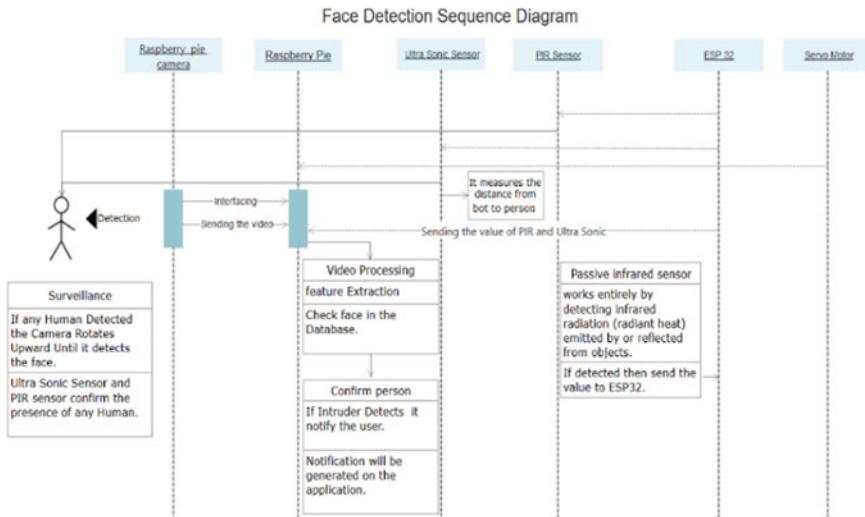


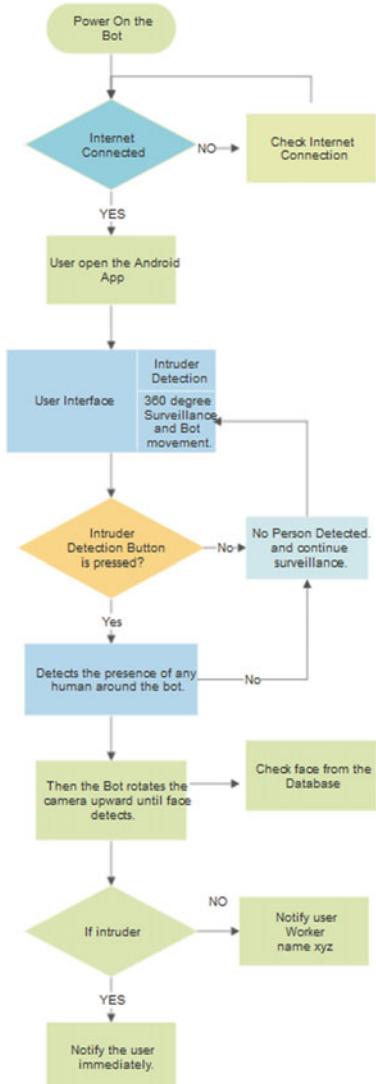
Fig. 1 Face detection sequence diagram

re-connect to establish Internet connectivity after which user will be able to manage/control BOT through mobile application. The camera presented on BOT can move 360 degrees and continuously scan area. As soon as any human presence is detected, camera captures the image of human, which is then verified by the database. If the person is intruder, then it sends alert else it notifies person. If no human presence is detected, BOT continues its surveillance. If the BOT is not operating for more than 30 min, it enters into power-saving mode.

3.3 360-Degree Surveillance

The bot has a feature that enables the user to surveillance in 360 degrees precisely, the user can keep an eye at every nook and corner of the specific area. The dual servo motor enables the camera to rotate the full 360-degree horizontal view and 96-degree vertical view. The user has the option of driving the bot remotely in a specific area the user only needs a good Internet speed for operating this bot. Figure 3 shows the CAD design of servo and stepper motor arrangement for the camera. This arrangement provides the user 96 degree vertical and 360-degree horizontal view of any specific arrangement.

Fig. 2 Flowchart of face detection mechanism



3.4 Application Development

Vijayata can be monitored and with a smartphone application. When user opens Vijayata application/web server the user needs to login with its user id and password generate at the time of sign up. Then the user directly gets connected to Vijayata. Now, user can surveillance, detect any intruder as well as update the guest list by uploading the picture of the visitor/guest coming into that respective area. User login windows in which the user needs to logins for accessing the bot and main page of

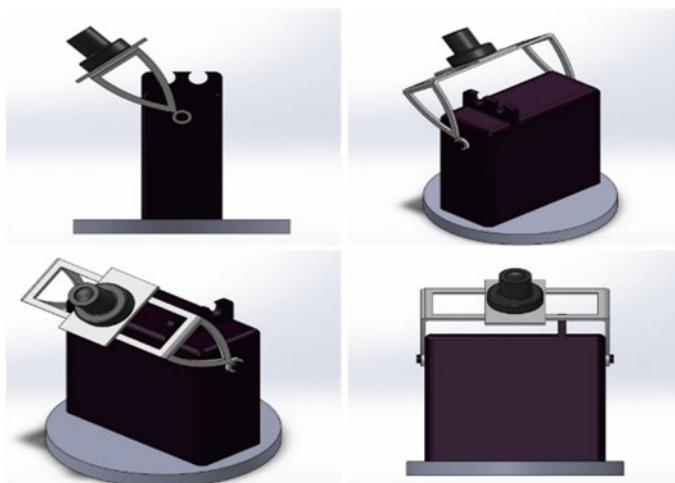
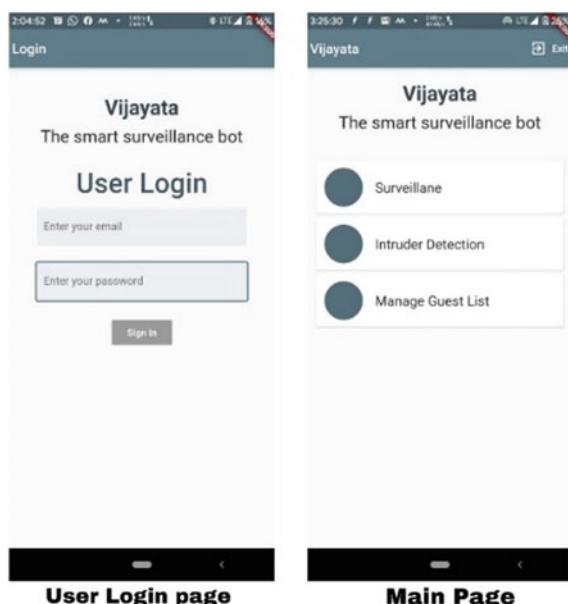


Fig. 3 CAD design of camera arrangement

the application on which the user can select what is wanting to do surveillance or intruder detection (Fig. 4).

Fig. 4 Application layout



4 Comparison

Parameter	Traditional system [3, 4]	Vijayata (Surveillance BOT)
Cost	Cost of the system is more than BOT	Cheaper than traditional system
Security	Chances of security breach are more than BOT, which makes it less secure	Chances of security breach are very less as compared to traditional surveillance system, which makes it highly secure
Feature	*Cannot do face detection *Does not perform real-time alert to the user *Cameras installed can rotate fixed angle *Camera will cover the area where it is installed	*Face detection can be done using BOT *BOT provides real-time alert to user *Camera can rotate 360° horizontal and 96° vertical *Bot can cover multiple areas by moving

5 Results

In this proposed paper for more understanding of Vijayata purpose and results, we have done analysis based on the data collected by the bot to the cloud about the presence of Intruder and guest and extract that data and apply data visualization. In Fig. 5, we are trying to show data analysis which represents the intruder and guest presence in the specific area in counts. Figure 5 shows the curves in which the dark blue curve indicates the guests counts and the sky-blue curve shows the intruder counts. If any intruder is detected, then the notification appears in the notification window of the user's mobile phone. The Vijayata captures the image of the person and matches it with stored images, and Fig. 6 shows the image how is detecting the face and eyes of the person, if any intruder is detected, then the notification appears in the notification window of the user's mobile phone.

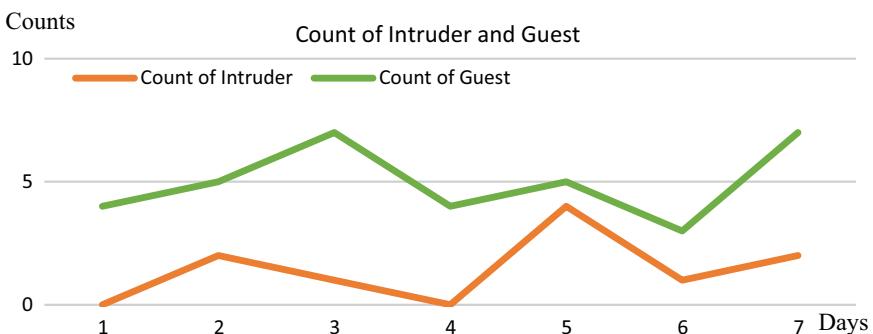


Fig. 5 Experimental results

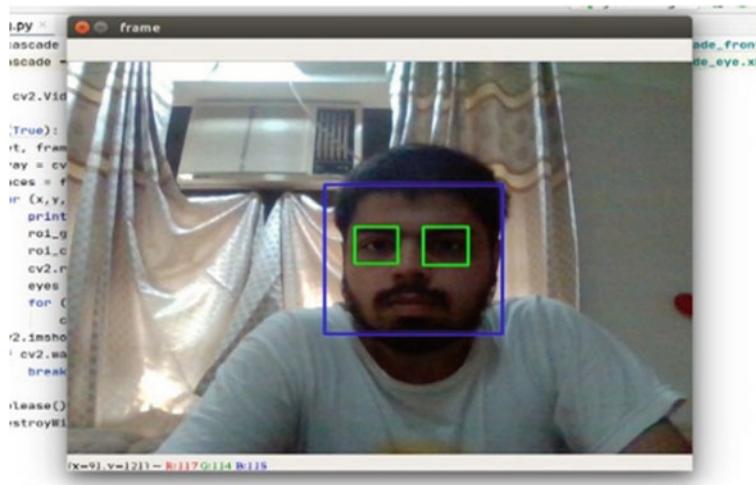


Fig. 6 Face detection by Vijayata (backend image)

6 Conclusion

Security and surveillance are very important aspect in today's time. The traditional security method for security involves many cameras, which leads to higher costs and the possibility of a security breach. To overcome these problems, we have developed "Vijayata" a "AI/IOT-BASED FACE DETECTION AND SURVEILLANCE BOT." Vijayata gives you the 360-degree surveillance with intruder detection system (INDS) using artificial intelligence, Internet of things, and computer vision. Using sensors like PIR, ultrasonic sensors, servo motor, Raspberry Pi, and Raspberry Pi Cam are used for face detection. Security is one of the prime concerns, and our BOT promises in reducing the number of cameras to be installed and safeguard the specified area where surveillance is done. Military, industry surveillance is few of the major security concerns areas, where security plays a crucial role to detect intruder and Vijayata, which will help in providing a solution for this. BOT proves to be cost-effective and cost-efficient by covering the major security concerns.

References

1. Daya AA, Salahuddin MA, Limam N, Boutaba R (2020) BotChase: graph-based bot detection using machine learning. IEEE Trans Netw Serv Manage 17(1):15–29. <https://doi.org/10.1109/TNSM.2020.2972405>
2. Ghouse Z, Nishika H, Nihar R (2017) Military robot for reconnaissance and surveillance using image processing. Int Res J Eng Technol 4(5)
3. Singh D, Zaware P, Nandgaonkar A (2017) Wi-Fi surveillance bot with real time audio & video streaming through Android mobile. In: 2017 2nd IEEE international conference on recent trends

- in electronics, information & communication technology (RTEICT), Bangalore, pp 746–75. <https://doi.org/10.1109/RTEICT.2017.8256696>
- 4. Mahamuni, Chaitanya Vijaykumar. "A military surveillance system based on wireless sensor networks with extended coverage life." 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC). IEEE, 2016.
 - 5. Shin MS, Kim BC, Hwang SM, Ko MC (2016) Design and implementation of IoT-based intelligent surveillance robot. *Stud Inf Control* 25(4)
 - 6. Ashok Kumar M, Thirumurugan T (2018) Integrated IOT based design and android operated multi-purpose field surveillance robot for military use. In: International conference for phoenixes on emerging current trends in engineering and Management (PECTEAM 2018), Advances in engineering research (AER), vol 142
 - 7. Saravana Kumar K, Priscilla P, Germiya K Jose, Balagopal G (2015) Human detection robot using PIR sensors. *Int J Sci Eng Technol Res (IJSETR)* 4(3)
 - 8. Verma H, Verma G, Yarlagadda J, Sharma A, Banarwal S (2018) Ardudroid Surveillance Bot. In: Muttoo S (ed) System and architecture. Advances in intelligent systems and computing, vol 732. Springer, Singapore.
 - 9. Hasan M et al (2019) A smart semi-automated multifarious surveillance bot for outdoor security using thermal image processing. *Adv Netw* 7(2):21–28
 - 10. Rajendran VG, Jayalalitha S, Radhakrishnan S, Arunbhaarat S (2020) Webpage controlled surveillance bot using Raspberry Pi. In: 2020 Fourth international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC), Palladam, India, pp 549–553. <https://doi.org/10.1109/I-SMAC49090.2020.9243320>
 - 11. Joshi A, Nagarjun CS, Srinivas R (2017) The DRASB—disaster response and surveillance bot. In: 2017 Second international conference on electrical, computer and communication technologies (ICECCT), Coimbatore, pp 1–8. <https://doi.org/10.1109/ICECCT.2017.8117959>

Atmospheric Turbulence Effects on Bit Error Rate in Lognormal and Negative Exponential Channel in FSO Link



Priyanka Bhardwaj, Manidipa Roy, and Sanjay Kumar Singh

Abstract With the increase in the population worldwide, there is a need of a well-ordered communication system which can manage the demands of the users. Radio communication system is not able to manage the over increasing demand of bandwidth. Optical wireless or FSO is one such solution to the problem. It has many advantages like large bandwidth, spectrum is unlicensed, high security, and ease of installation. But due to atmospheric turbulences, the link quality is compromised. One of the major issues faced in free space optical links is the induced fading due to the channel which in turn deteriorates the performance of the link. Different channel models are used for different turbulence conditions. Lognormal and negative exponential are two turbulence channel models discussed in this paper. For low turbulence condition and where propagation distance is less, we use lognormal model. Negative exponential model is employed where we have strong turbulence conditions and the link length spans a few kilometers. Various modulation techniques are used in a FSO link. In this paper, we have used BPSK modulation technique. Effects of atmospheric turbulence on the channel link are observed considering parameters such as refractive index structure parameter C_n^2 and mean irradiance I_0 . The change in the value of bit error rate due to change in the values of these parameters has also been listed.

Keywords FSO · Lognormal channel · Negative exponential channel · Atmospheric turbulence · BPSK · Bit error rate (BER)

P. Bhardwaj (✉) · M. Roy · S. K. Singh
ABES Engineering College, Ghaziabad, Uttar Pradesh, India
e-mail: priyanka.bhardwaj@abes.ac.in

M. Roy
e-mail: manidipa.roy@abes.ac.in

1 Introduction

FSO is a cableless communication system which uses optical carriers to transfer the message signal. It employs line of sight (LOS) technique for transferring message from transmitter to receiver. FSO has solved the problem of bandwidth deficiency. It provides large bandwidth. The installation of FSO communication system is very easy. Also, it is immune to radio frequency interference. The data rates provided by this system are comparable to optical fiber communication system data rates, thus making it a preferable choice.

FSO has applications in wide domain, for example, in broadcasting, educational institution applications, satellite communication, and military applications [1, 2].

2 Channel Modeling

Due to the gradients of pressure and temperature, there are irradiance variations in the received signal. This leads to beam broadening and distortion of received spectrum. There are different levels of turbulence conditions such as weak (S.I. < 1), moderate (S.I. = 1) and strong (S.I. > 1). Different statistical channel models for transmission in free space are developed which are Gamma-Gamma model, lognormal model, K-channel distribution, I-K channel model, negative exponential model, etc. In this paper, we have discussed lognormal model and negative exponential model [3–6].

2.1 Lognormal Channel

Lognormal model is used for weak turbulence conditions. The lognormal probability density function describes the fading statistics and scintillation for weak turbulence. It is very simple in terms of mathematical calculations. The probability density function of received optical signal in lognormal channel model is given by:

$$p(I) = \frac{1}{2\pi\sigma_i^2} \frac{1}{I} \exp\left\{-\frac{(\ln(I/I_0) - m_i)^2}{2\sigma^2}\right\}, I \geq 0 \quad (1)$$

where I_0 is the normalized irradiance arrived at receiver end; σ_i^2 is the log irradiance variance.

The main advantage of using lognormal modeling in free space optical communication is that it makes mathematical calculations very easy. As mentioned earlier, this model is applicable for weak turbulence only. It is used for propagation distance less than 100 m.

Refractive index structure parameter (C_n^2) is used for quantifying the amount of variation of refractive index in the medium. For a horizontally propagating field via turbulent medium, the refractive index structure parameter and log irradiance variance are related as:

$$\sigma_l^3 = 1.23C_n^2k^{7/6}L_p^{11/6} \quad (2)$$

where wave number is k , and L_p is horizontal distance of optical irradiance, C_n^2 is refractive index structure parameter, and σ_l^2 is log irradiance variance. Therefore, there is a direct relation between σ_l^2 and C_n^2 .

2.2 Negative Exponential Channel

This channel model is used for strong atmospheric turbulences where the length of the link extends to many kilometers, and thus, the number of independent/autonomous scatters becomes higher. In this case amplitude of the signal indicates Rayleigh distribution this in turn extends to negative exponential statistics for signal intensity. This is given by:

$$p(I) = \frac{1}{I_0} \exp\left[-\frac{I}{I_0}\right], \quad I_0 > 0 \quad (3)$$

where I_0 is the mean received irradiance. Here the value of the scintillation index is 1, i.e., S. I. $\rightarrow 1$ [7].

3 BPSK Modulation Scheme

There are different modulation schemes which can be used in free space optical communication like OOK, BPSK, and QPSK. In this paper, we have used BPSK as the modulation technique at the transmitter side.

Binary phase-shift keying (BPSK) or more popularly known as phase-shift keying (PSK). This is two-phase modulation techniques. In this scheme, binary 0 and binary 1 of the message. These two are shown by two separate phase states in the carrier signal.

Binary-1 is shown with the actual carrier, and binary 0 is represented by carrier having a phase shift of 180° . The two phases of the carrier signal are shown as:

$$S_1(t) = A_c \cos(2\pi f_c t) \quad (4)$$

$$S_0(t) = A_c \cos(2\pi f_c t + \pi) \quad (5)$$

where A_c is the amplitude of the carrier signal and f_c is the carrier frequency. Instantaneous time is denoted by t . The signal $S_0(t)$ stands for the carrier signal when binary 0 is transmitted and the signal $S_1(t)$ stands for the carrier signal when binary 1 is transmitted [8, 9].

4 Simulation and Discussion

Bit error rate (BER) helps us in determining the performance of the FSO link. Ratio of number of bits received to the total number of bits transferred gives the bit error rate. The bits of the message signal get altered due to the atmospheric turbulences. For a better performance, the BER should be as minimum as possible [10–13]. In the following sections, we have plotted the simulation results for evaluating the changes in BER due to changes in some specific parameters.

BER analysis for lognormal turbulence model

From Fig. 1, we can observe the effects of changing values of log irradiance variance (σ_l^2) on the bit error rate (BER) in the received signal. From Eq. (2), σ_l^2 is directly proportional to C_n^2 , which is used for characterizing for the amount of refractive index fluctuations. We have listed down the values of BER for different values of σ_l^2 at two different values of SNR. Table 1 shows the increasing value of BER as

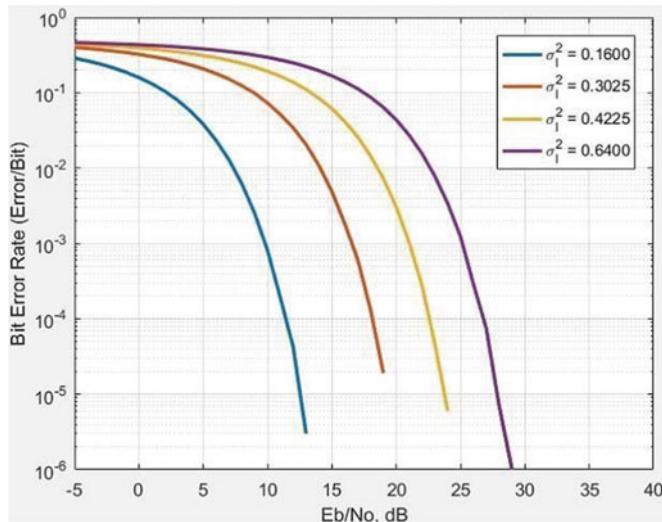


Fig. 1 Plot for BER versus SNR for different values of in lognormal channel model

Table 1 Variation of BER with σ_l^2 for SNR = 5db in lognormal channel

σ_l^2	BER ($\times 10^{-3}$)
(0.40)2 = 0.1600	38.25
(0.50)2 = 0.3025	207.50
(0.65)2 = 0.4225	311.60
(0.80)2 = 0.6400	380.20

Table 2 Variation of BER with σ_l^2 for SNR = 11db in lognormal channel

σ_l^2	BER($\times 10^{-3}$)
(0.40)2 = 0.1600	0.19
(0.50)2 = 0.3025	51.31
(0.65)2 = 0.4225	165.30
(0.80)2 = 0.6400	271.90

σ_l^2 rises, which indicates more variations in the atmosphere. Table 2 also shows the similar trend on increasing the value of σ_l^2 . Tables 1 and 2 are tabulated when the value for SNR is 5db and 11db, respectively.

BER analysis for negative exponential model

I_0 is the mean of receiver optical irradiance, i.e., $E[I]$. The plot in Figure 2 is obtained by varying the values of I_0 . It was observed that as the received optical irradiance increases, which means lesser was the effect of strong turbulent atmosphere on the

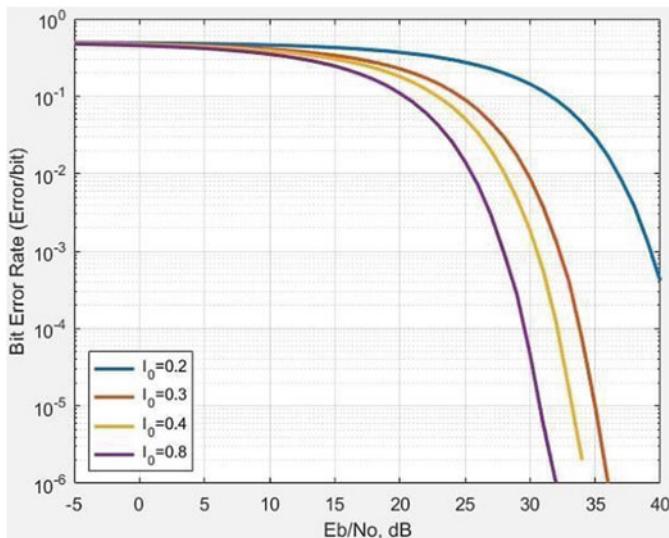
**Fig. 2** Plot for BER versus SNR for different values in negative exponential channel model

Table 3 Variation of BER with I_0 for SNR = 20db in negative exponential channel

I_0	BER($\times 10^{-3}$)
0.8	109.6
0.4	179.5
0.3	226.1
0.2	368.3

Table 4 Variation of BER with I_0 for SNR = 28db in negative exponential channel

I_0	BER($\times 10^{-3}$)
0.8	1.03
0.4	10.53
0.3	29.52
0.2	198.90

signal received, the BER was low. This variation in BER can be confirmed from the values tabulated in Tables 3 and 4. Tables 3 and 4 are tabulated when the value for SNR is 20 db and 28 db, respectively.

5 Result and Conclusion

The FSO link considered in this paper is tested in different turbulent conditions. There are many modulation techniques that can be deployed at transmitter, but for less complexity, we considered the BPSK modulation for multiple channel models. The main parameter used for the analysis is BER on a common value of SNR.

For weak turbulence in medium, lognormal model was selected, and from the simulations, it was substantiated that with strong irradiance fluctuations signified by increase in the value of σ_l^2 the bit error rate also increases.

On the other hand, for strong atmospheric turbulence, negative exponential channel was used and simulation results implied that more irradiance fluctuations in medium would result in low received optical irradiance I_0 and high erroneous signal (more BER for entire range of SNR).

References

1. Henninger H, Wilfert O (2010) An introduction to free space optical communication. Radio Eng 19(2), June 2010
2. Khalighi MA, Uysal M (2014) Survey on free space optical communication: a communication theory perspective. In: IEEE communications surveys & tutorials, vol 16, no. 4. Fourthquarter, pp 2231–2258

3. Parikh J, Jain VK (2011) Study on statistical models of atmospheric channel for FSO communication link. In: 2011 Nirma university international conference on engineering. Ahmedabad, Gujarat, pp 1–7
4. Jarangal, E, Dhawan D (2018) Comparison of channel models based on atmospheric turbulences of FSO system
5. Arnon S (2013) Effects of atmospheric turbulence and building sway on optical wireless communication systems. Opt Lett 28(2):129–131
6. Ricklin J, Hammel S, Eaton F, Lachinova S (2016) Atmospheric channel effects on free space laser communication. J Opt Fiber Commun Rep 3:111–158
7. Tannaz S, Ghobadi C, Nourinia J, Mostafapour E (2018) The effects of negative exponential and K-distribution modeled FSO links on the performance of diffusion adaptive networks. In: 2018 9th international symposium on telecommunications (IST). Tehran, Iran, pp 19–22
8. Choyon AKMSJ, Chowdhury R (2020) Performance comparison of free-space optical (FSO) communication link under OOK, BPSK, DPSK, QPSK & 8-PSK modulation formats in presence of strong atmospheric turbulence. J Opt Commun. <https://doi.org/10.1515/joc-2019-0250>
9. Barua B (2011) Evaluate the performance of FSO communication link with employing OOK and BPSK as modulation technique under turbulent condition
10. Kono Y, Pandey A, Sahu A (2019) BER analysis of lognormal and gamma-gamma turbulence channel under different modulation techniques for FSO system. In: 2019 3rd international conference on trends in electronics and informatics (ICOEI). Tirunelveli, India, pp 1385–1388
11. Nazrul Islam AKM, Majumder SP (2015) Analytical evaluation of bit error rate performance of a free- space optical communication system with receive diversity impaired by pointing error. J Opt Commun 36(2), Jun 2015
12. Sharif, Majumder SP (2016) Analytical SISO bit error rate evaluation of a satellite to ground link under the influence of log-normal atmospheric turbulence. In: 2016 3rd international conference on electrical engineering and information communication technology (ICEEICT). Dhaka, pp 1–4
13. Garcia-Zambrana A (2007) Error rate performance for STBC in free-space optical communications through strong atmospheric turbulence. IEEE Commun Lett 11:390–392

Simulation and Design of Mach–Zehnder Interferometer



Priyanka Bhardwaj, Manidipa Roy, and Sanjay Kumar Singh

Abstract In this paper, some light has been put on the emerging nanophotonics communication technology by emphasizing an important nanophotonics device called Mach–Zehnder Interferometer. Comprehensible research and study were performed for thorough understanding and intendment of the device that includes its structure, types, working, etc. Further scrutiny on the device was performed to elucidate the various aspects of its operation, the basis of which were the simulation results obtained from the designing of the Mach–Zehnder Interferometer on the software Lumerical DEVICE. Constructive and destructive interference patterns were obtained by subsequent simulations, and their energy intensities were plotted. Mach–Zehnder Interferometer also finds applications in fields of modulation and sensing as variable ratio power splitter and pressure sensor, respectively, among other essential appositeness. Both variable ratio power splitter and pressure sensor are studied in depth to understand its implementation and suitability in various optical devices. Some drawbacks and future aspects of the technology based on MZI have also been concluded.

Keywords Constructive and destructive interference · Lumerical mode · Pressure sensor · Mach–Zehnder Interferometer nanophotonics · Twin core fiber · Variable ratio power splitter

1 Introduction

Nanophotonics is the study of optics, i.e., behavior of light on a very minute scale of nanometers. This branch of optics deals with light waves that range in hundreds of nanometers in wavelength, and the interaction aspect of light with objects of

P. Bhardwaj (✉) · M. Roy · S. K. Singh
ABES Engineering College, Ghaziabad, Uttar Pradesh, India
e-mail: priyanka.bhardwaj@abes.ac.in

M. Roy
e-mail: manidipa.roy@abes.ac.in

nanometer-scale objects is also analyzed here. With the advent of technology, there has been a fuming need of high speed digital communication. The optical communication domain is known to cater for this inevitable need of the world by providing larger bandwidth of data transfer over longer distances as compared to the conventional communication systems that include coaxial cables. Although, the most significant advantage of nanophotonics remains its ability to transmit larger amounts of information in a unit time over other transmission media, some other advantages including immunity to electromagnetic interference, security as the electronic emission cannot be intercepted casually, which makes it more desirable. The transfer of data through the fiber optics cables has provided us with multiple advantages still there is an imminent need to venture into nanophotonics.

Working with light in the nano-scale enables us to benefit from its desirable properties including low noise, high speed, and low voltage and power. Nanophotonics devices play an important role in data communication as they may serve as a media for the same and also convert optical energy into electrical energy and electrical energy into optical energy. Mach–Zehnder is one of the nanophotonics devices used in many fields.

1.1 Working Principle of Mach–Zehnder Interferometer

The Mach–Zehnder Interferometer was developed by Ludwig Mach and Ludwig Zehnder. The method involves two beam splitters that can split and recombine the two beams and can produce two outputs. It is a simple device that demonstrates interference by division of amplitude. In MZI, the coherent LASER beam is split into two arms and recombined in the coupler after traveling in these waveguide arms (Fig. 1).

When the light beams reach the second beam splitter, they get recombined to provide an output beam. The intensity with which the output beam reflects depends upon relative difference in phase acquired by the beam along the two paths; efficiency

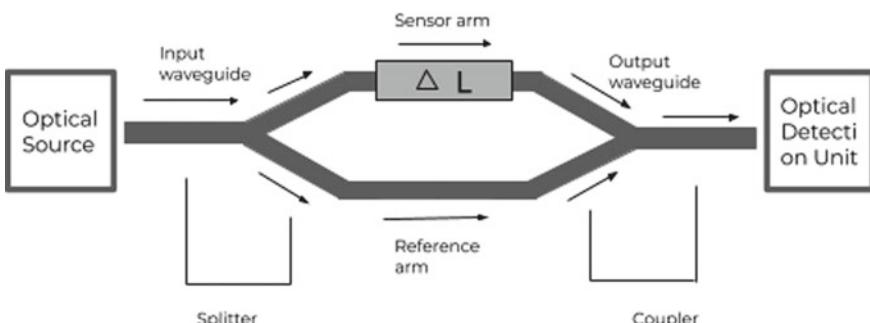


Fig. 1 Schematic diagram of an Unbalanced Mach–Zehnder Interferometer

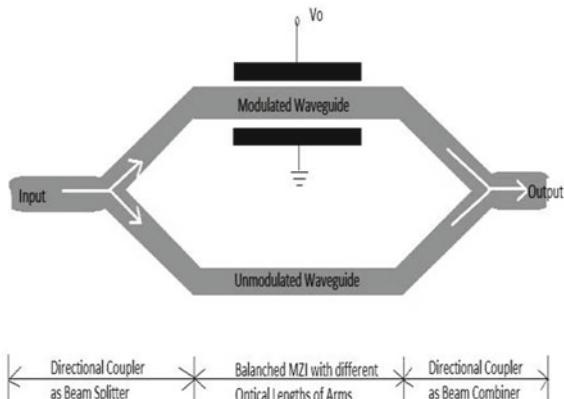
of the light intensity may vary between 0 and 100%, as disinterred by the relative phases. The management of optical outputs at the two terminals depends upon the accurate variation in arm lengths and on wavelength of transmission. By adjusting the path difference for accurate optical wavelength, the optical power can be channelized into one of the channels for well-aligned interferometer, whereas for improperly aligned beams, there will be some variation in pattern obtained at both outputs, and the shapes of the interference pattern vary with varying path length difference [1]. In other words, depending on the phase difference between the two waves received at the combiner, the Mach–Zehnder Interferometer will act either like a switch or a modulator. When the phase difference between the waves is either 0 or 180, the output will prevail through either one of the output arms with the maximum intensity, it possibly can and the other arm will have approximately zero light output. On the other hand, if the phase difference lies between 0 and 180, the MZI will act as a modulator; as a result, light output could be received at both the arms of the output, and the intensity of the two may or may not be equal and is completely dependent on the angle of phase difference.

2 Balanced and Unbalanced MZI

In an Unbalanced Mach–Zehnder Interferometer, in order to get a phase difference at the combiner, a geometrical path difference is encountered in one of the two paths of the interferometer by means of an additional length ΔL . So, one of the arms is of length L , and the other is of length $L + \Delta L$. When the light wave crosses the beam splitter and gets split into the two arms, the time taken by light to reach the combiner end is not the same at both the arms because one of the light waves has to travel a larger distance than the other.

A Balanced Mach–Zehnder Interferometer has both the waveguide arms of the same geometric length unlike the Unbalanced Mach–Zehnder Interferometer. To get the required phase difference between the light waves from the two arms, we make use of the electro-optic property in the materials that are used to fabricate the waveguides. The materials that possess the electro-optic properties happen to demonstrate a change in optical properties, mostly effective refractive index, in influence of electric potential across it. In a Balanced Mach–Zehnder Interferometer, the waveguides are constructed of materials like silicon or lithium Niobate (LiTiNO_3) which displays excellent receptivity to electric potential applied across it in terms of change in its refractive index. Two electric rods are put across one of the waveguide arms such that they form an interleaved pattern of waveguide and the rods. One of the rods is connected to ground, while the other one across the waveguide is connected to some significant voltage as depicted in Fig. 2. When there is an electric potential set across one of the waveguides, its refractive index increases or decreases in reference to the other waveguide arm with no electric potential applied. As a result, the speed of light traveling in the potential-applied waveguide would be different from that of

Fig. 2 Balanced Mach–Zehnder Interferometer



the light in the other waveguide. This would lead to an expected phase difference at the combiner.

3 Modeling and Simulation

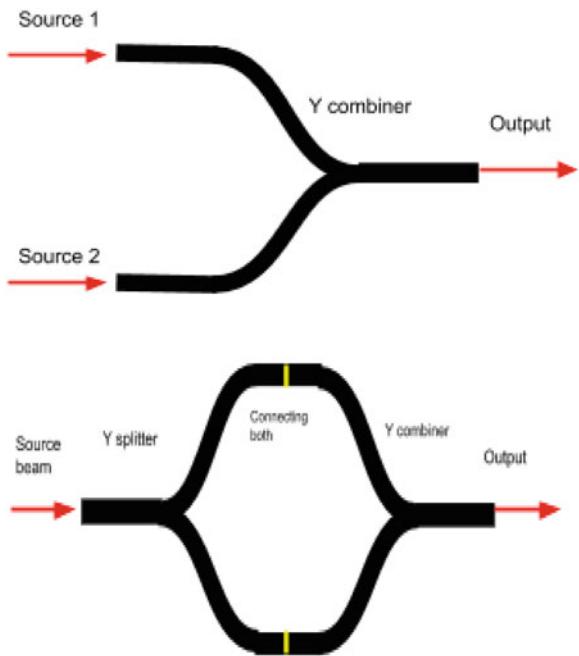
Lumerical mode was used to design and simulate the results for the Mach–Zehnder Interferometer.

3.1 Design Parameters

A MZI was modeled using two Y splitters as shown in figure. A rectangular base of SiO_2 (glass) was placed which was topped by a combination of Y -splitter and Y combiner as shown in Fig. 3. The Y -splitter was formed by a waveguide of $1 \mu\text{m}$ wide and $0.6 \mu\text{m}$ high. The index of the waveguide used was 1.5525 micron. The source used was in forward direction with injection axis X , amplitude 1 and 0 phase. The center wavelength of the source was taken $1.548 \mu\text{m}$. Different modes were observed, and the best was selected using Eigen-mode solver. Simulation was carried out using a finite difference time-domain solver. Further, to observe the constructive and destructive interference in MZI output, two sources were connected as shown in Fig. 3, each to the arm of the Y combiner such that the arms will act as the sensing and reference arm of the interferometer.

To obtain constructive interference, the phase difference for the sources was kept to be 0, and similarly for destructive interference, the phase difference was kept 180.

Fig. 3 **a** Schematic diagram of Y-splitter and combiner for MZI; **b** schematic diagram of Y combiner to observe constructive and destructive interference in MZI



3.2 Simulation Results

Device-level simulation using numerical mode was done, i.e., to simulate the waveguide structure and to extract the modes traveling in the waveguide. The energy density of the first quasi-TE mode is given in Fig. 4. It can be seen that most light is confined inside the waveguide. Since the waveguide is surrounded by silicon-dioxide, it has a high optical confinement and lower cross-talk between other waveguides, but it suffers from higher scattering loss due to sidewall roughness caused during manufacturing.

Constructive and destructive interference was observed with following simulation results showing energy density. As one can see in Figs. 5 and 6, in constructive interference, the intensity is confined in the output waveguide, whereas in destructive interference, the intensity drops in the output waveguide.

4 Application

MZI shows tolerance to electromagnetic interference, and their easy signal detection adds a great deal to its compact size and multiplexing capability. For such reasons, MZI is an extremely attractive option for automotive industries and environmental monitoring. Fiber optic interferometers are used in sensing various parameters like

Fig. 4 Energy intensity in MZI

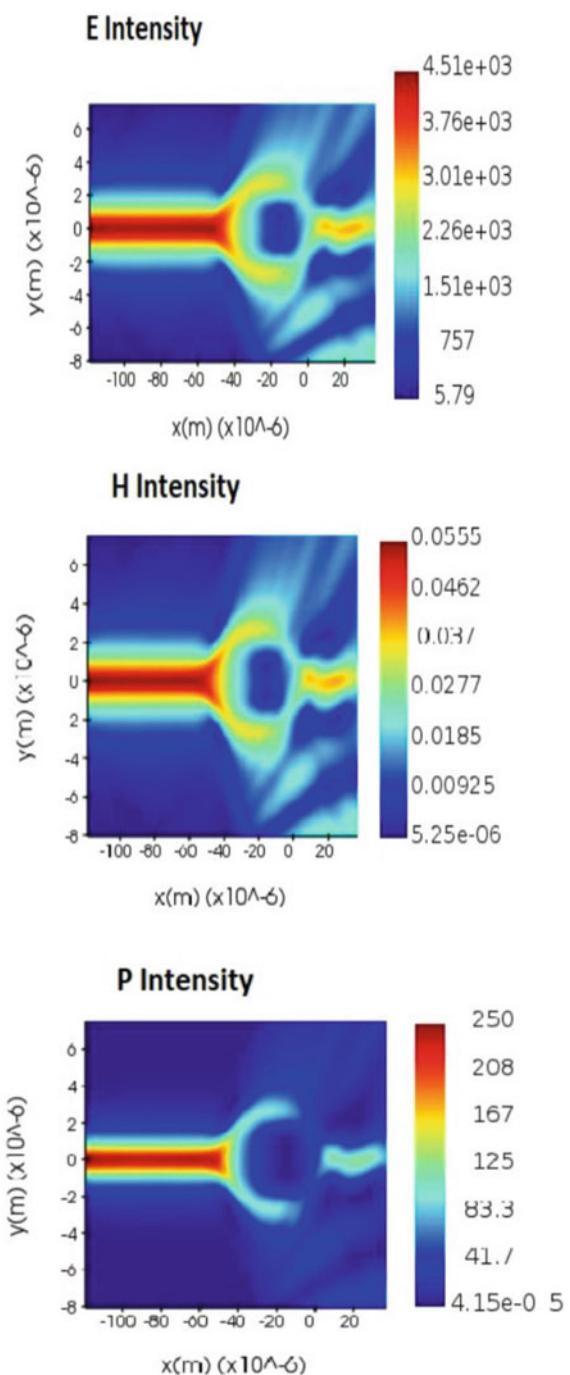
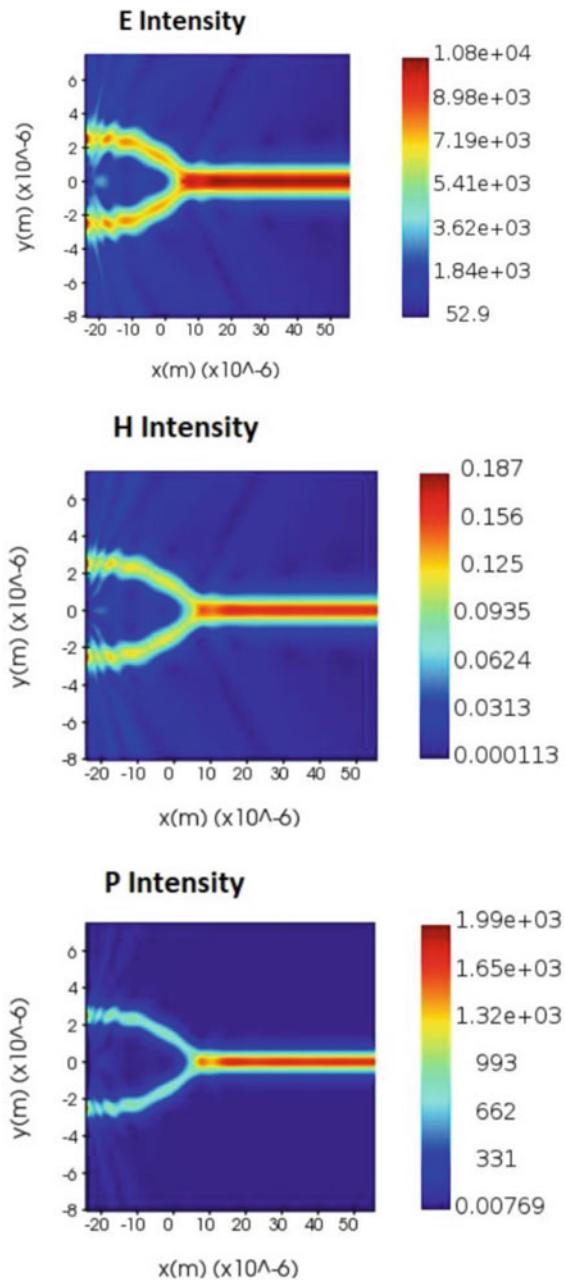


Fig. 5 Constructive interference in MZI



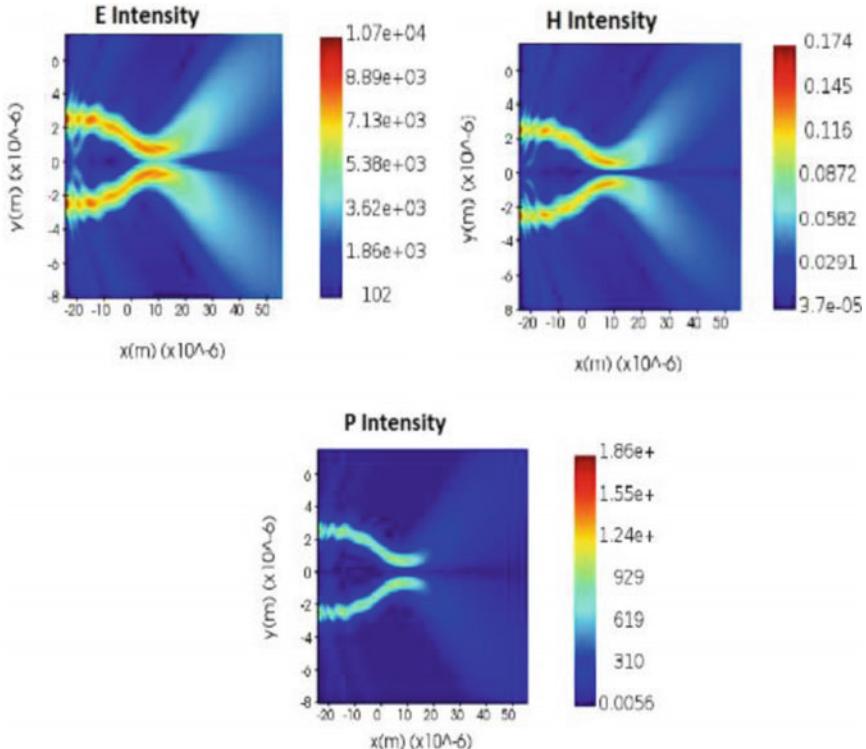


Fig. 6 Destructive interference in MZI

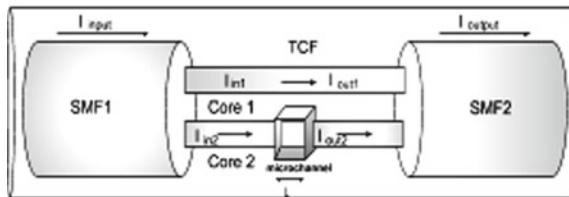
temperature, pressure, refractive index, strain, and others [2]. Among many applications of MZI, two applications have been studied in this paper, which are MZI as pressure sensor and variable ratio power splitter.

4.1 MZI as Pressure Sensor

The coherent light passing through MZI arms recombines at the coupler after traveling through the arms, this recombination leads to interference. For MZI employed as an optical pressure sensor, a membrane is placed under the sensor arm. There is movement of this membrane due to pressure which leads to modification of the optical path through the sensor arm. The optical output intensity is a measure of the phase variation, experienced by the sensor arm when pressure is experienced by the membrane [3]. The phase variation is primarily affected by two factors:

- (a) Rise in optical path length;
- (b) Elasto-optic effect.

Fig. 7 Schematic diagram of TCF-based MZI



These models use pressure sensitive membranes, and hence, these are weak since the thin diaphragm lined up with the fiber end which is easily cracked will limit its performance in a raised pressure. An improvement in the structure with more accuracy is sensors using twin core fiber (TCF). These models use pressure sensitive membranes, and hence, these are ineffective since the thin diaphragm is used at the fiber end. An improvement in the structure with more accuracy was the sensor using twin core fiber (TCF). Figure 7 shows the diagram of TCF-based MZI. A small part of TCF is spliced between single mode fibers (SMF) using femtosecond lasers, and with micromachining, a micro-channel is formed.

When a light beam passes from single mode fiber to twin core fiber, it gets split into two beams of intensity I_{in1} and I_{in2} . On passing through the micro-channel, the light beam I_{in2} goes through phase change; hence, an optical path difference is obtained. When the two beams, i.e., I_{out1} and I_{out2} , recombined in single mode fiber 2, an interference pattern will be observed due to phase change. The output intensity obtained can be expressed by [4]:

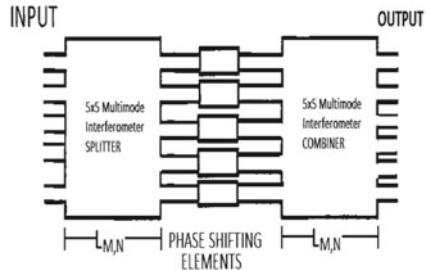
$$I = I_{OUT1} + I_{OUT2} + 2\sqrt{I_{OUT1} I_{OUT2}} \cos((2\pi L \Delta n)/\lambda + \varphi_0) \quad (1)$$

where λ is the wavelength of light, L is the micro-channel length as mentioned in (1), $\Delta n = n_{core} - n_{channel}$ is the refractive index difference between the two interference arms, where n_{core} and $n_{channel}$ are the refractive indices of core and channel, and φ_0 is the phase of interference. The signal of interference reaches the minimum value if (2),

$$((2\pi L \Delta n)/\lambda + \varphi_0) = (2m + 1)\pi \quad (2)$$

where m is an integer, m is the wavelength of m th order interference dip [4]. TCF-based MZI pressure sensors showed a good linear response with high pressure sensitivity.

Fig. 8 Mach-Zehnder Interferometer as variable ratio power splitter



4.2 Variable Ratio Power Splitter

The specification of the output intensity distribution is useful in a variety of optical applications, including the tap function, which extracts only a small portion of the light from a channel for monitoring purposes, ring lasers, where the splitting ratio of the coupler dictates the laser's operation, and WDM devices, where the spectral response can be improved through non-u channels [5].

For the purpose of variable ratio power splitter, a generalized $N \times N$ Mach-Zehnder Interferometer is used which has a similar method of operation to the Balanced Mach-Zehnder Interferometer but can have N number of inputs and outputs. In an interferometer, the number of input and output arms may or may not be equal. An $N \times N$ Multimode Interferometer (MMI) splitter, an active phase shifting region with N phase shifter, and a $N \times N$ MMI combiner make up the fundamental layout of the $N \times N$ GMZI. The MMI splitter creates N self-images of the input that appear at the output ports, resulting in an equal distribution of power to the phase shifting region's all arms. Both the Multimode Interferometers work on the self-imaging principle, which is defined as “a property of multimode waveguides by which an input field profile is duplicated in single or multiple images at periodic intervals along the propagate at periodic intervals.” [5] (Fig. 8).

All the waves at the input are “in-phase” with each other, and then as they move, they get out-of-phase. But after a periodic interval, at a particular distance called “Beat Distance” denoted by $L\pi$, some of these modes again come in-phase (or their phases are integral multiple of π) and lead to the reconstruction of the input field in the form of single or multiple self-images. A characteristic beat length L for a multimode waveguide cm therefore be defined as in Eq. (3),

$$L\pi = \pi/(\beta_0 - \beta_1) = [(4nch)xWe2]/3\lambda \quad (3)$$

Here, nch is the refractive index of the channel or waveguide, and λ is the wavelength of the light through the waveguide.

The field at a distance z along the multimode region given can now be expressed in terms of this beat length

$$LM, N = (M/N) \times 3L\pi \quad (4)$$

Because the self-imaging phenomena is periodic, the parameter M characterizes a multiple of the imaging length that also results in N self-images. The beat length is given by LM , N , where N self-images are reconstructed, each with a length of M units.

After the first MMI, all of the modes exiting from its output were given an equal amount of power. The active phase sifting zones are then used to process the phase modulated light pulses. The active phase shifting elements' job is to change the phases of the self-images so that a certain distribution of relative phases can be produced at the MMI combiner's inputs. The active phase shifters work by using a driving signal (voltage, current, etc.) to alter the phases of the light in the arms using a variety of effects, such as the electro-optic or thermo-optic effect. Finally, phase modulated light enters the final MMI, where the N inputs with similar intensities and phases are combined, resulting in the formation of a specified power distribution at the outputs. The final output images' locations and intensities are determined by the relative phases at the MMI combiner's input. As a result, by altering the p, light from an input port I can be dispersed in a controlled manner to the output ports.

5 Conclusion

Although optical communication has paved the way for innovative and efficient ways of communication, they do endure some impediments as well. The high speed, low power, cheap fabrication, and short switching time are all advantages of the Mach–Zehnder Interferometer [6]. One of the major disadvantages of the traditional Y-junction waveguide is that it loses 50% of its power if it is incident on only one of the two arms of the Y. The Mach–Zehnder Interferometer is sensitive to temperature variations, which causes a RI deviation in the RI base sensor, which is based on an in-line fiber. Interferometer Mach–Zehnder (MZI). Mach–Zehnder Interferometer is operated upon with polarized light only. It affects the performance of the interferometer severely when unpolarized light is used in the process. Many of the biosensors or other sensing-based operations may encounter quite a few problems using Mach–Zehnder Interferometer with a single output wave for sensing. More exploration and development in the field is required.

6 Future Scope

The fact that Mach–Zehnder Interferometer can be used to communicate on nano-scales makes it quite apt for military communication which require immunity to electromagnetic interference and security as the electronic emission cannot be intercepted casually by trespassers. The major categories where Mach–Zehnder Interferometer is employed are sensing, command, control, and power among others [7].

Although, optical data-transmission technologies for use in future optical links at the HL-LHC are currently being investigated, it is one of the promising new technologies being looked at is silicon photonics [8]. Silicon photonics is well suited for a number of applications yet a material called lithium niobate provides better results of electro-optic property. When an external potential is applied across waveguides which are constructed of lithium niobate, it provides a better range of modulation than silicon waveguide.

Owing to the compactness associated with Mach–Zehnder Interferometer, it has potential to serve in the domains of smart dust technology.

References

1. Banerji S, Design of TE and TM mode Mach-Zehnder interferometer based on SOI technology, figshare. J Contrib. Available: <https://doi.org/10.6084/m9.figshare.2069974>
2. Lee BH, Kim YH, Park KS, Eom JB, Kim MJ, Rho BS, Choi HY (2012) Interferometric fiber optic sensors. Sensors (Basel, Switzerland) 12:2467–2486
3. Siarkowski AL, Bulla DAP, Morimoto NI (2001) Mach-Zehnder interferometer simulation results for integrated optical pressure sensor. In: Proceeding of the XVI SBMicro international conference on microelectronics and packaging, pp 233–235
4. Li Z, Liao C, Wang Y, Lei X, Wang D, Dong X, Liu S, Wang Q, Yang K, Zhou J (2015) Highly-sensitive gas pressure sensor using twin-core fiber based in-line Mach-Zehnder interferometer. Opt Express 23:6673–6678
5. Lagali NS, Paiam MR, MacDonald RI (1999) Theory of variable-ratio power splitters using multimode interference couplers. IEEE Photonics Technol Lett 11(6):665–667
6. Kotiyal S, Thapliyal H, Ranganathan N (2012) Mach-Zehnder interferometer based all optical reversible NOR gates. In: 2012 IEEE computer society annual symposium on VLSI. Amherst, MA, pp 207–212
7. National Research Council (2008) Nanophotonics: accessibility and applicability. The National Academies Press, Washington, DC. Available: <https://doi.org/10.17226/11907>
8. Nasr-Storey SSE, Boeuf F (2015) Effect of radiation on a Mach–Zehnder interferometer silicon modulator for HL-LHC data transmission applications. IEEE Trans Nucl Sci 62(1), Feb 2015

IoT-Based Smart Home Security and Automation System



Partha Chakraborty and Sajeda Sultana

Abstract In this age of technology, security is one of the key issues. Also, face recognition is a significant part of the purpose of security and surveillance. Our goal is to explore the implementation of a Raspberry Pi-based face recognition system using conventional face detection and recognition techniques for resolving security issues. For face detection and recognition, the Eigen face method is used, and to implement the steps of this method, the Haar cascade classifier algorithm is also used here. This paper aims at taking face recognition to the next level. So, using this system, the owner/authority can store audio messages and text messages for anyone. After accurate recognition, the system provides an audio message to the recognized person. At the same time, it can also send text messages to the owner/authority's smart phone or any other mobile device that is recognized. With the use of the Raspberry Pi kit, we aim at making the system cost-effective and easy to use with high performance. Tests and proper analysis of the system were done to verify the efficiency of the system.

Keywords Home automation · IoT · Raspberry Pi · Eigen face · Haar cascade classifier

1 Introduction

In this age of technology, the Internet of things (IoT) refers to the network of physical objects (sensors, actuators, etc.) that use sensors to capture real-time data and connected objects to share data over the Internet without any human intervention. From small houses to huge industries, surveillance, and signaling have become essential to meet our security needs [1]. Home automation systems have achieved great popularity in today's world and increase comfort, convenience, quality of life and security for residents. The security of the home is the prime perturbation. To protect

P. Chakraborty (✉) · S. Sultana

Department of Computer Science and Engineering, Comilla University, Cumilla 3506, Bangladesh
e-mail: partha.chak@cou.ac.bd

the home, we need to install an expensive security system. A human attention recognition system has been developed in this work [2]. Wireless security and home automation are the two perspectives of this project. With the present prototype system, the owner/authority can capture images via webcam and store the images of any desired person. At the same time, the owner/authority can store the audio message in the system that he/she wants to give to the desired person or strangers and also store the text message that the system will send to the owner by identifying the person. If someone enters the entrance of the house, the webcam captures his/her face, and the face recognition system helps us to recognize him/her with more accuracy. After successful recognition, it checks if there is any audio message in the system for the recognized person. If there is a message, it will play in the audio box. At the same time, it can tell the authorities who is recognized and can also send text messages to a smart phone or any other mobile device using the Twilio communications platform to find out who entered. This low-cost system with the fewest requirements takes care of both home security and home automation [3]. The general public can easily use the system for security purposes. In the absence of a person, his/her message will be conveyed to the desired person. Adding this system to CCTV will make CCTV more useful. Through the system, it is also possible to avoid liability for many unwanted incidents that happen around us.

This paper will describe an approach for security purposes in which we implement a continuous monitoring system that provides service by playing audio messages and sending alerts by sending text messages to the authority/owner.

The subsequent sections of this paper are listed below. In Sect. 2, we describe the literature review. In Sect. 3, we describe the methodology. Section 4 describes the results and discussion of the system. Finally, Sect. 5 concludes the paper and presents future work plans.

2 Literature Review

Home security and home automation are smart technologies that make human life smarter and more updated. In the existing system, the home automation system focuses only on automating the house or power expenditure range of activities. The presented prototype system focuses on being secure, fast, and easily accessible by ordinary people. It has less cost as compared to the previous system and has extra benefits, such as security. Miss. Pradnya R. et al. introduced a home automation system, namely a door locking system, which takes a code as input from the GSM module to approve entrance to the house [4]. According to yadav2018iot, Prof. Shivarudraiah et al. [5] demonstrate a door locking system controlled by giving voice commands and capturing images with a webcam, as well as sending them to the authorized person with an Android smart phone. Author used the PCA algorithm for face recognition. Firstly, the system collects images and stores them in the database. After successful detection by analyzing the data [6] comparing it with the face database, it can give attendance to the unidentified students [7]. Suraj Pawar et al. introduced

a door lock–unlock system using IoT and face recognition. Their methodology is quite good, but it is not remotely accessible for everyone using a smartphone [8]. Jiakailin Wang et al. proposed a face recognition system for home security service robots. They provide a good explanation of LBP and the working procedure of face recognition. But they did not provide any hardware explanation. Also, their systems are expensive for robotic components [9]. Yugashini et al. paper focuses on the study and development of face recognition for magnetic door access control. The three main subsystems are, namely, face detection, face recognition, and automatic door access control. The face detection and recognition process are implemented by a fast-based principal component analysis (FBPCA) approach in which the face image is detected by a web camera. Then the detected images are compared to the face data. If the image is authenticated, the magnetic door opens automatically. In the case of an unauthenticated image, an SMS is sent to the user using a GSM modem [10]. Various detection and identification processes are carried out in these papers [11, 12]. Mansi Soni et al. demonstrate home automation by speech recognition for the disabled. Mostly, they used machine learning methodologies to determine the attitudes of people. This paper introduced ant colony optimization and a decision tree algorithm. These are useful for creating smart home automation with high accuracy. Using shallow machine learning techniques, the authors implemented an attention recognition system [13, 14].

3 Methodology

Each user who has expertise in existing systems can imagine a system that can assemble more flexibility, for example, sending text messages and playing voice messages. The proposed system supports more efficiency and security. Our proposed system can be shown in Fig. 1

The process can be categorized into two sections: data collection with preparation and model development.

3.1 Data Set Collection with Preparation

New images of users are captured by the webcam. The webcam captures only the particular user's frontal face and the images go through pre-processing. To get region of interest (ROI), images are cropped, and the cropped images are resized to 48 x 64 pixels, which will be used for further processing. The face is detected by an algorithm. The algorithm we use is the Haar cascade classifier. The classifier needs to be trained for face detection before it can be used. An xml file-haarcascade frontalface alt2 is used in the Haar cascade classifier training data. For feature extraction, Haar features will be used. In the face classifier, we are using the detectMultiScale module from

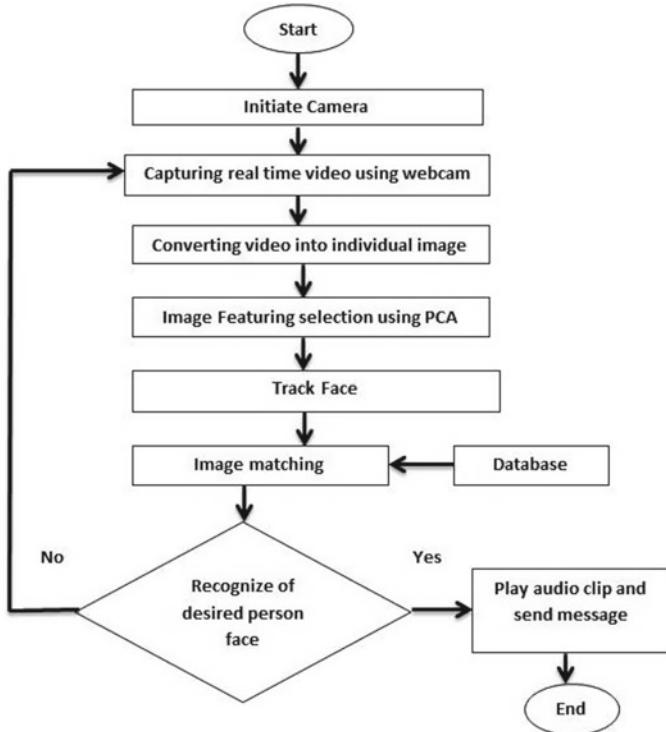


Fig. 1 Flow chart of the system

OpenCV. This module is used to detect eyes and faces from ROI. Then it creates a rectangular box around a face where it detects the face in an image.

Taking all detected face images with a camera and then adding a unique number of the people in the image is saved in a required folder as a data set. We have used 100 samples for a single individual and that gives more accuracy about the person's images. One part of the data set is used as training data for the algorithm, the rest is used for testing. All the training images are resized and then covered with the pre-processed images into a single vector for recognition.

3.2 Model Development

After collecting and preparing data sets, the next step is to recognize the face images. The face is recognized by an algorithm. The algorithm we use is principal component analysis (PCA). PCA is a useful unsupervised statistical technique that is the most widely used in exploratory data analysis, image compression, face recognition, and in machine learning for vertical models. The principal component analysis

(PCA) approach requires full frontal images of a person; otherwise, it shows low performance. The basis of the Eigen face method is a principal component analysis method. It seems to be a suitable method to be used in face recognition because of its speed, plainness, and learning ability. For face recognition, the PCA algorithm is a step-to-step overview:

- i Creating the testing image that would be an individual image.
- ii The test image is pre-processed and then calculated to calculate the feature vector. The test image is transformed into an Eigen face component. Comparing the input images with our mean image and multiplying them by the Eigen vectors.
- iii Calculate the average distance by using the Euclidean distance formula between the test feature vector and all the training features vectors.

In mathematically, the recognition process finds out the minimum Euclidean distance following the given equation:

$$W_{\text{test}} = U_i(T_{\text{test}} - \bar{X})$$

where U_i is the i th Eigen faces and where $i = 1, 2, 3, 4, 5, 6, -K$

$$D_{\text{avg}} = \|\Omega(t) - \Omega(i)\|$$

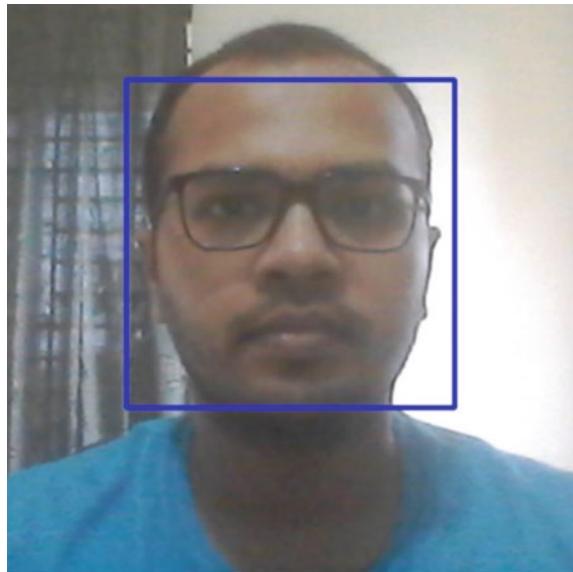
where $i = 1, 2, 3, 4, -K$.

- iv Find the face class with a minimum Euclidean distance that shows equivalence to the test image.

According to steps of the algorithm, several features are calculated from an image to detect whether it is a face or not and to recognize the face too. At the time of face detection, the algorithm applies all features to the image, and if the face is present in the image, it stores the face location. After identifying a face from images, the face recognition system is applied to that face location to match with the pre-trained weights. After matching with the pre-trained weights, it can easily recognize the face of that image. The authority/owner of the system can store audio messages for any dedicated person. So, then the device recognizes that person and plays the audio message for that person if it is available in the system. Using the Twilio package, the system is able to send text messages to a smart phone or any other mobile device.

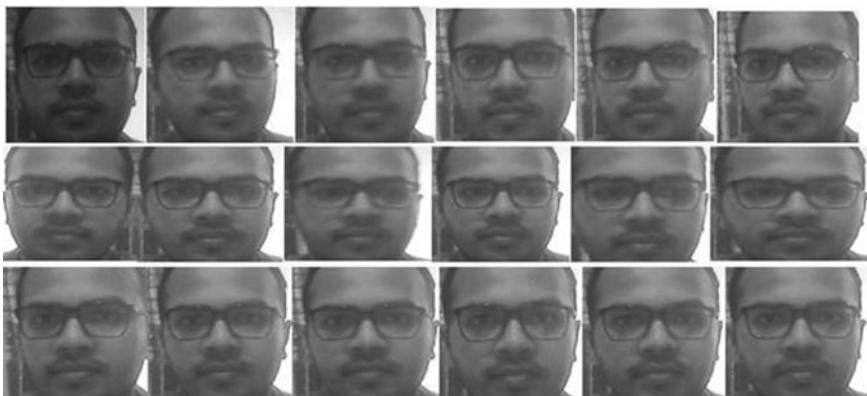
4 Result and Discussion

To interact with the system, we make a system interface layout. The layout has several options, like person information, set audio and text messages, and exit. By clicking the person information button, it is ready to capture the image and asks for an Id. After giving Id, the webcam starts automatically, as shown in Fig. 2, and

Fig. 2 Face detection

starts detecting faces in a blue frame. Then it takes 100 images spontaneously of that person by using a webcam. Then the images will be pre-processed and stored in the image folder as facial data sets in Fig. 3 that are in the recognition process.

By clicking the set audio and text message button, the user sets audio and text messages if required. When any person is in front of a system's webcam, it will automatically detect the face with a blue-colored rectangular box. The detected face is extracted and then compared to the database. After a proper match, it will show the name linked with the face in the rectangular box in Fig. 4.

**Fig. 3** First 18 facial data sets

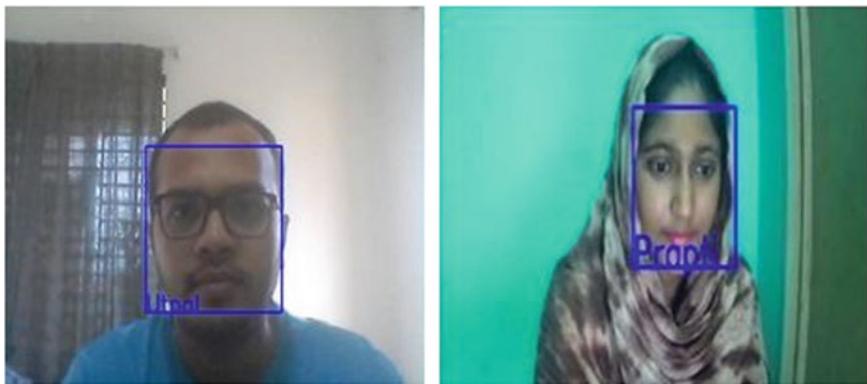
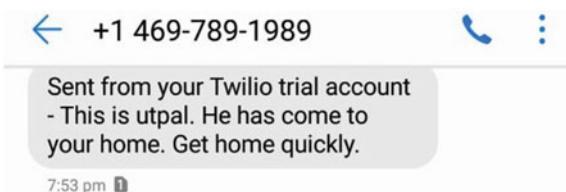


Fig. 4 Face recognition

Fig. 5 Text message from the system



After successful recognition, the system plays a voice message in an audio box if it is available in the system for the recognized person. At the same time, notify the owner/authority by text message using the Twilio platform that is shown in Fig. 5

According to Pawar et al. [8], smart authors have proposed a smart home security system that captures images and recognizes faces. If the person's image matches, then it sends notification via Gmail and sends a message to the smart phone using WiFi or GSM module. But in any case, our system does not require an Internet connection to send the message to the owner/authority. A school attendance system using face recognition by Lukas et al. [15]. The system is very well presented and explains in detail the working attendance system. They developed their system using the discrete wavelet transform (DWT) and discrete cosine transform (DCT). They used five facial training images for each student, and their recognition accuracy was about 82%. We have used 100 samples for each person. That gives more accurate information about a person's images. The accuracy of our system is better than their system. Wati et al. [16] have proposed a smart home security system based on face detection and recognition that captures face images and performs image processing using MyRIO 1900 as the main controller that holds the software for image acquisition, face detection, and recognition. But in our system, we have used the Raspberry Pi 3B as the main processing device. The cost of MYRIO is much higher than the Raspberry Pi (Table 1).

Table 1 Comparison table overview

Authors	Year	Accuracy
Pawar et al. [8]	2018	80% accuracy when it is tested in real time
Lukas et al. [15]	2016	The success rate of recognizing facial images is about 82%
Wati et al. [16]	2017	The accuracy of face recognition is 80%
Sajeda et al.	2021	The average face recognition accuracy is about 83.22%

5 Conclusion and Future Work

Building a secure house is the primary cause of concern in today's world and face recognition helps us to achieve it. The IOT-based proposed automated security system is low power and low cost. The proposed system will be able to recognize known people and provide service by playing audio messages and sending text messages to the owner/authority. Our target is to optimize the system's face recognition performance by increasing the number of images in the training set and then minimizing the computation time that may occur with this enhanced algorithm. If a person uses accessories that cover part of their face, then our system cannot detect it. In the future, we will improve it. In this fast-paced world, quick response time is always a priority. That is why we will reduce the time of message sending. Finally, we have a plan to convert the system in such a way that the system can be trained from any type of image and the owner or the authorized user can control the system and perform all the tasks automatically from an android app.

References

- Patil N, Ambatkar S, Kakde S (2017) IoT based smart surveillance security system using raspberry pi. In: 2017 International Conference on Communication and Signal Processing (ICCP). IEEE, pp 0344–0348
- Chakraborty P, Ahmed S, Yousuf MA, Azad A, Alyami SA, Moni MA (2021) A human-robot interaction system calculating visual focus of human's attention level. IEEE Access 9:93409–93421
- Kodali RK, Jain V, Bose S, Boppana L (2016) IoT based smart security and home automation system. In: 2016 international conference on computing, communication and automation (ICCA). IEEE, pp 1286–1289
- Nehete PR, Chaudhari JP, Pachpande SR, Rane KP (2016) Literature survey on door lock security systems. Int J Comput Appl 153(2):13–18
- Yadav C (2018) IoT based surveillance system and home automation. Int Res J Eng Technol 5(5):2031–2036
- Sarker A, Chakraborty P, Sha SMS, Khatun M, Hasan MR, Banerjee K (2020) Improvised technique for analyzing data and detecting terrorist attack using machine learning approach based on twitter data. J Comput Commun 8(7):50–62
- Chakraborty P, Muzammel CS, Khatun M, Islam SF, Rahman S (2020) Automatic student attendance system using face recognition. Int J Eng Adv Technol (IJEAT) 9:93–99

8. Pawar S, Kithani V, Ahuja S, Sahu S (2016) Smart home security using iot and face recognition. In: 2018 fourth international conference on computing communication control and automation (ICCUBEA), IEEE, pp 1–6
9. Wang J, Zheng J, Zhang S, He J, Liang X, Feng S (2016) A face recognition system based on local binary patterns and support vector machine for home security service robot. In: 2016 9th international symposium on computational intelligence and design (ISCID), vol 2, IEEE, pp 303–307
10. Yugashini I, Vidhyasri S, Gayathri Devi K (2013) Design and implementation of automated door accessing system with face recognition. *Int J Sci Mod Eng (IJISME)* 1(12)
11. Chakraborty P, Roy D, Rahman MZ, Rahman S (2019) Eye gaze controlled virtual keyboard. *Int J Recent Technol Eng* 8(4):3264–3269
12. Sultana M, Ahmed T, Chakraborty P, Khatun M, Hasan MR, Uddin MS (2020) Object detection using template and HOG feature matching. *Int J Adv Comput Sci Appl* 11(7):233–238
13. Chakraborty P, Yousuf MA, Rahman S (2021) Predicting level of visual focus of human's attention using machine learning approaches. In: Proceedings of international conference on trends in computational and cognitive engineering. Springer, Berlin, pp 683–694
14. Sayeed S, Sultana F, Chakraborty P, Yousuf MA (2021) Assessment of Eyeball Movement and Head Movement Detection Based on Reading. In: Bhattacharyya S, Mršić L, Brkljačić M, Varghese Kureethara J, Koeppen M (eds) Recent Trends in Signal and Image Processing. ISSIP 2020. Advances in Intelligent Systems and Computing, vol 1333. Springer, Singapore. https://doi.org/10.1007/978-981-33-6966-5_10
15. Lukas S, Mitra AR, Desanti RI, Krisnadi D (2016) Student attendance system in classroom using face recognition technique. In: 2016 international conference on information and communication technology convergence (ICTC). IEEE, pp 1032–1035
16. Wati DAR, Abadiano D (2017) Design of face detection and recognition system for smart home security application. In: 2017 2nd international conferences on information technology, information systems and electrical engineering (ICITISEE). IEEE, pp 342–347

A Fuzzy Model for Selection of Information and Location-Based Security Attributes in Cloud Environment



Deepika , Rajneesh Kumar , and Dalip

Abstract Security is a prime concern in cloud computing-based environment in which user accesses resources remotely and pays on used bases. Nowadays, the information and location-based security is one of the most important issues to access these cloud resources. This paper proposed a fuzzy model for assigning weights to information and location-based security attributes (FMAW-ILSA). This model is used to assign an experienced-based fuzzy weight to security attacks attributes, information and location-based security attributes. The assigned weight helps in selection of information and location-based security attributes. The criteria for assigning the weights depend on the attribute frequencies. The motive behind this paper is to select the information and location-based security attributes which are less utilized in the existing work such as information security (IS), data location and relocation (DLRL), phishing attack (PA), location tracking (LT) and location and control of data (LCD). The developed fuzzy model helps the researchers for better decision making in order to select information and location-based security attributes for future research to more enhance the cloud security.

Keywords Security attacks (SA) · Security principles · Information security attributes (ISA) · Location security attributes (LSA) · Fuzzy weight

1 Introduction

The cloud users are facilitated with the availability of vast resources. The cloud computing can be categorized into three general types. A private cloud is defined as its utilization is limited to their cloud service users, enterprise owned or leased.

Deepika · R. Kumar

M.M. Engineering College, M.M. (Deemed to be University), Mullana, Ambala, India

Deepika

UIE-CSE, Chandigarh University, Mohali, India

Dalip

MMICTBM, M.M. (Deemed to be University), Mullana, Ambala, India

It is defined as all of its services are accessible over a private internal network to its selected (private/internal users) users and not publically. In public cloud, the resources are available through the Internet to the users, and the users will pay according to their specifications, requirements and utilization. In public cloud, the service providers allow its users scalability and sharing of the resources which is not seen in private cloud. The private and public model together makes a hybrid model. The hybrid model creates a bridge between the cloud used by one user (i.e., is private) and one cloud managed by the public user. The security of data is very critical in public cloud. In this scenario, security becomes an important agenda for cloud users [1]. In this new era where emerging technologies are coming day by day, the users today are searching for different aspects of security like physical and financial security. Hence, there is a need to explore the security in the area of information and location. Data security is one of the important factors which needs lots of concern in cloud computing. The different challenges faced in front of users are insider access, identity management, accessibility, data protection, cloud computing database, data storage, location-based cryptography, location-based access control and location-based identity. In this paper, the two main challenges, information and location-based security, are considered. If unauthorized persons are able to access the information from the cloud that is known as information security attack. The existing studies show that different researchers done their work on cloud security such as information and location. It has been concluded that less secure models for information and location-based security were designed for cloud environments. This will encourage to do this types of analysis for selection of information and location-based security attribute in cloud environment and motivates us to design and implement cloud security models. The aim of this paper is to select information and location-based attributes for making better decision for improving the cloud security by the researchers.

Contribution of the paper

- To improve decision making process for selection of information and location-based security attributes.
- To promote research on less utilized attributes of information and location-based security in cloud.

This paper is discussed as the existing work done by different researchers and finds various gaps in Sect. 2. The Sect. 3 elaborates about information and location-based security attributes. A novel methodology is presented in Sect. 4. Section 5 shows the result and discussion. The conclusion is discussed in Sect. 6.

2 Literature

Ni Zhang et al. in A Research on Cloud Computing Security paper discussed about the fundamental concepts and scope of cloud security. In this paper, different types of permissions which are assigned to cloud users in cloud environment are studied. To

overcome the limitations in cloud, security different mechanisms and solutions are used [1]. The paper explains the importance of secure software and data protection which helps in protecting the cloud services. The different data security issues and their solutions are reviewed in this paper [2]. Dr. P. Dinadayalan et al. explain the important issue in cloud computing called as data security. The study of different data security issues and their methods is described in cloud [3]. Shweta Singh et al. in Security in Cloud Computing explain that the data privacy and accessibility of cloud services are the major problems. The complex architecture is not possible to secure the cloud services at once so continuous efforts are making to enhance the security of cloud. A secure cloud environment develops which consists of hardware, software and data. A secured cloud computing environment is capable to store, manage and access the encrypted data efficiently as well as manage the authenticity of the cloud users [4]. This paper elaborates the different issues in security and its methods. There are various challenges that occur in cloud computing such as loss of private data and data leakage. There are many threats to the users data which is to be stored on the cloud [5]. Mojtaba Alizadeh et al. elaborate on the authentication issues of the cloud users as well as facing problems during accessing resources from the cloud. The main problem in cloud computing is user authentication. This paper explains the effective authentication strategies and their methods [6]. This paper discussed the different security threats and their solutions. A comparative analysis of threats and intrusion detection as well as prevention framework in cloud is analyzed in this paper [7]. The survey in this paper explains the security issues which occurred when local users access data from the remote users. This paper discussed architecture of cloud, services and deployment model, security issues, attacks and its solutions [8]. Muhammad Baqer Mollah et al. explains several security threats raised in cloud. These issues occurred because the communication was held between the mobile devices and the cloud through wireless medium, and the original data was stored on cloud infrastructure. This lead to new security threats in cloud [9]. Tejashri A. Patil et al. discussed about the data integrity in cloud computing. The unauthorized user access is prevented with the use of third-party auditor. An encrypted algorithm was proposed in this paper which performs auditing instead of decrypting data. This proposed algorithm is more secured than the existing one because third-party auditing reduces auditing overhead [10]. Tarek Radwan et al. explain the concept of data integrity, data privacy, data security and issues in cloud. This paper focus on the deployment and delivery models. It also explains the comparative analysis of these computing models [11]. E. Kesavulu Reddy et al. in Information Security in Cloud Computing paper identify various privacy and security issues of cloud users and its providers [12]. Christos Stergiou et al. explain that both the data storage and data execution is done outside the mobile device in the external environment. The main goal of integration between these two technologies is to provide the interaction between the things and objects among wireless networks and act as a single entity. The RSA and AES encryption algorithm used in integrated cloud computing and IoT technologies is used to find the security challenges by using the proposed algorithm model [13]. Bader Alouffi et al. described that data storage, data confidentiality, data availability, data integrity and data privacy are the major area for security concern.

[14]. In the paper, location-based service (LBS) described by Deepanshu Goyal et. al. explains about the authenticity of the user. A user becomes an authorized user if its location is valid within the specific area; otherwise, it will become an unauthorized user. This paper proposed a protected architecture for accessing cloud data using the location-based service (LBS) in mobile cloud computing (MCC) for authorized and registered user within the organization. The user is authentic if its identity is valid within the organization [15]. A novel secure spatial query protocol using fully homomorphic encryption is discussed in this paper which provides data integrity and protection against various attacks [16]. Hamed Tabrizchi et al. define the several security challenges faced by cloud users and the cloud service providers [17]. Hesham Abusaimeh et al. discussed the security and user authentication as a major issues among cloud users during accessing the cloud data remotely [18]. Yan He et al. discussed about the different mechanism of users identity, system authentication and location privacy system in cloud environment. These mechanisms not guarantee the confidentiality of user communication and data transmission [19]. Chetan Jaiswal et al. in Location-Based Security Framework for Cloud Perimeters describe the physical location of the datacenters and its protection mechanism using firewall. This will reduce the effort of cloud service providers to protect the data centers [20]. Fig. 1 shows the different security principles, and Fig. 2 depicts the different security attacks, information and location-based security attributes used by different researchers.

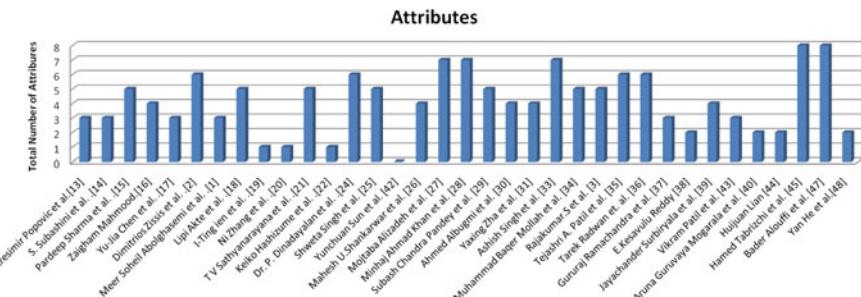


Fig. 1 Security principle attributes used by different researchers

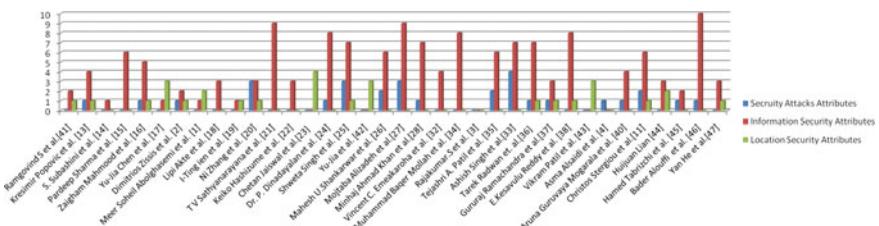


Fig. 2 Security attacks, information security and location-based security attributes used by different researchers

3 Information and Location-Based Security

The contribution of this comprehensive paper as compare with existing studies is depicted in above Fig. 1 which shows the analysis of different security principles, and Fig. 2 shows a year-wise comparison of different security attack attributes, information and location security attributes. Fig. 1 shows a total of eight security principles which are taken for survey such as data confidentiality (DC), data privacy (DP), data availability (DA), data integrity (DI), user privacy (UP), authentication (AU), access control (AC) and client server security (CSS). There are five security attack attributes such as Zombie Attack (DoS/DDoS attack) (ZA), loss of governance (LG) and data loss and leakage (DLL), loss of trust (LT) and data transmission (DT) as shown in Fig. 4. Fig. 6 depicts the sixteen information security attributes such as virtualization (VM), network security (NS), data loss (DL), load balancing (LB), data encryption and key management (DEKM), backup and recovery (BR), data recovery in cloud computing (DRCC), trust management (TM), network security (NS), identity and access management (IAM), data security (DS), data availability (DA) and multi-tenancy (MT), Internet and service (IS), data location and relocation (DLRL) and phishing attack (PA). The seven location security attributes are defined as data location policy (DLP), data location (DLC), geographic location information (GLI), Google latitudes (GL), location tracking (LT), phishing attack (PA) and location and control of data (LCD) as shown in Fig. 5. These researchers Bader Alouffi et al. and Yan He et al. maximum utilized all the security principles in their work, whereas Yunchuan Sun et al. less utilized these security principles. The authors Ashish Singh et al. frequently used different security attacks attributes and Ramgovind S et al. rarely used. The authors Hesham Abusameh et al. and S. Subashani et al. maximum utilized the information security attributes, whereas I-Ting Lien et al. used rarely. On the other hand, the authors Chetan Jaiswal et al. work maximum on location-based security attributes, whereas Bader Alouffi et al. less utilized the location attributes. From the analysis of these several security principles, security attacks attributes, information and location security attributes papers show useful contribution in cloud environment. From this comparative study, it has been quite clear that less work is done on few parameters of information and location security attributes. It motivates the researchers to do their research on information and location-based security attributes to improve the security in cloud environment.

4 A Novel Methodology

This paper proposed a fuzzy model for assigning weights to information and location-based security attributes (FMAW-ILSA). This model assigned weights and threshold values to the information and location-based security attributes as shown in Fig. 3. The proposed model is divided into three phases: First, second and final phase. In first phase, input is taken as all the attributes and check their relevancies. In second

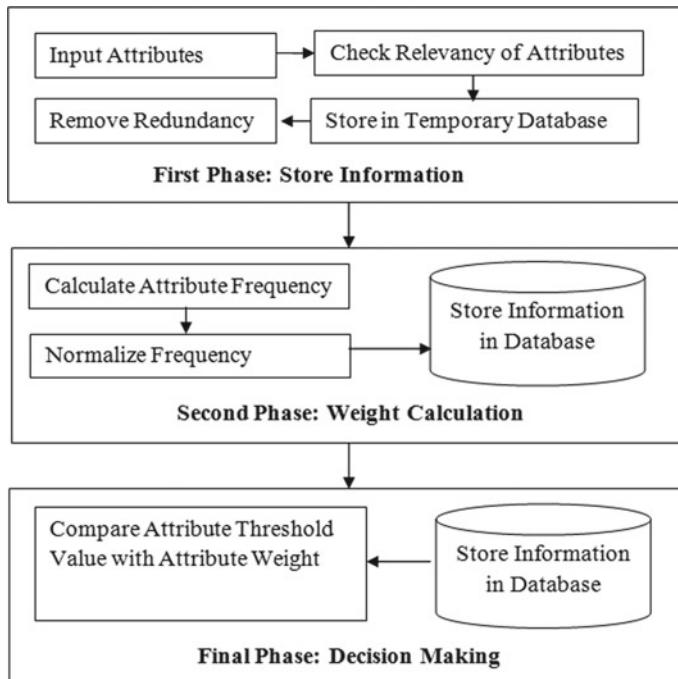
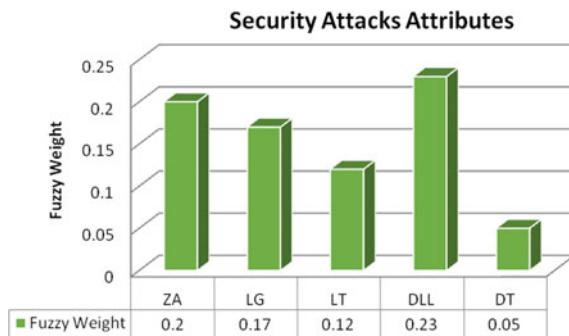
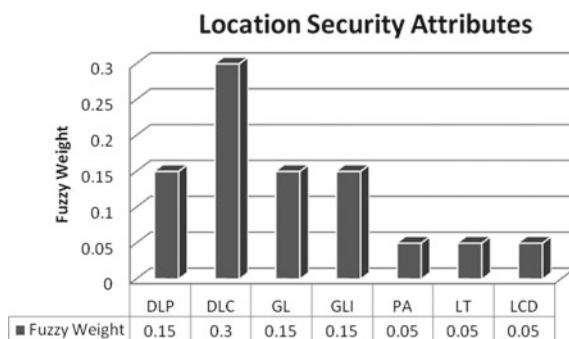
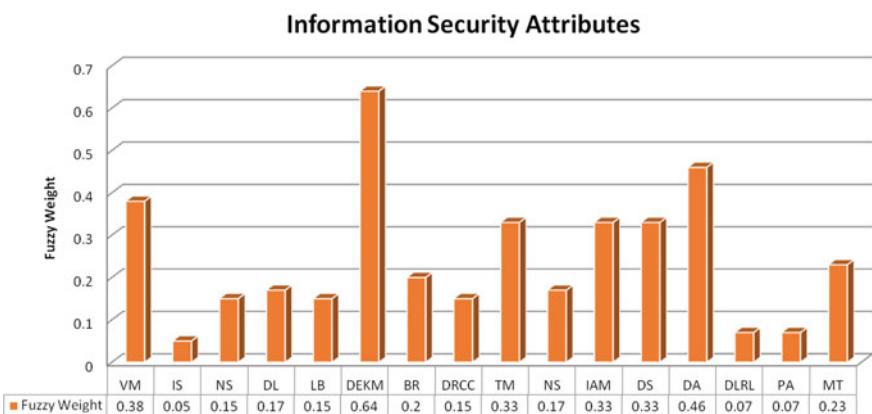


Fig. 3 Fuzzy model for assigning weights to information and location-based security attributes (FMAW-ILSA)

phase, only relevant attributes are considered for weight assignments and stored in the temporary database. The attributes frequency is calculated on the basis of their occurrence and normalized these frequencies. Final phase is decision making phase which compares the threshold values of information and location-based security attributes with their weights. If the attributes weights are high as compared to its threshold value, it indicates that those attributes are less utilized, and if the weights are less to that of threshold value, then those attributes are considered as most utilized by the researchers.

5 Results and Discussion

Figure 3 shows the FMAW-ILSA model. The above model shows the assigned weights and threshold values to the information and location-based security attributes which helps in making the better decision to select the security attributes. This motivates the researchers to do more research on less utilized attributes. The fuzzy weights are assigned to different security attacks attributes, information security attributes and location security attributes as shown in Figs. 4, 5 and 6. Figure 4 shows total

**Fig. 4** Fuzzy weight assigned to security attacks attributes**Fig. 5** Fuzzy weight assigned to location security attributes**Fig. 6** Fuzzy weight assigned to information security attributes

number of five security attacks attributes as DLL attribute weight (0.23) which means this attribute is maximum used and DT attribute weight (0.05) which means it is frequently used. In Figure 5, the DLL attributes are used frequently, whereas less work done on DT attribute.

There are total sixteen information security attributes which are shown in Fig. 6 in which DEKM attributes weight is maximum (0.64) and IS attribute weight is minimum (0.05). It means DEKM attribute is commonly used by the researchers in their studies, and IS attribute is used rarely. The total number of seven location security attributes is used by the different authors, where DLC attribute weight is (0.30) which means this attribute is used maximum number of times, and PA, LT and LCD weights are (0.05) which means these attributes are less used as shown in Fig. 5. These analyses show that DLC attribute is widely used, and PA, LT and LCD are less utilized attributes by the researchers. It has been concluded that the experienced-based weights are assigned that helps in better decision making for selecting the information and location-based security attributes in cloud environment.

6 Conclusion

In this paper, different security principles, security attacks, information security and location-based security attributes have been analyzed. The assigned weights by a developed novel fuzzy model FMAW-ILSA show that most of the parameters are frequently used by the researcher in their work as few parameters are less utilized. Hence, designed model plays a vital role to enhance the security, and less fuzzy weights of attributes shows that less work has been done on location security attributes which encourages the researcher to do their work on location-based security attributes in cloud environment to improve the security. It also motivates the researchers in better decision making to select the security attributes for designing and implements a secure model in cloud environment.

References

1. Zhang N, Liu D, Zhang Y-Y (2013) A research on cloud computing security. In: International conference on information technology and applications, IEEE, pp 370–373
2. Sathyaranayana TV, Mary Immaculate Sheela L (2013) Data security in cloud computing. In: International conference on green computing, communication and conservation of energy (ICGCE), IEEE, pp 822–827
3. Dinadayalan P, Jegadeeswari S, Gnanambigai D (2014) Data security issues in cloud environment and solutions. In: World congress on computing and communication technologies. IEEE, pp 88–91
4. Singh S (2014) Security in cloud computing. *Int J Comput Appl Technol Res* 3(8):488–493
5. Shankarwar MU, Pawar AV (2015) Security and privacy in cloud computing: a survey. In: Proceeding of the 3rd International Conference on Frontiers Of Intelligence Computing (FICTA), vol 2. Springer International publishing Switzerland

6. Alizadeh M, Abolfazli S, Zamani M, Baharun S, Sakurai K (2016) Authentication in mobile cloud computing: a survey. *J Netw Comput Appl Elsevier* 61:59–80
7. Khan MA (2016) A survey of security issues for cloud computing. *J Netw Comput Appl, Elsevier* (71):11–29
8. Singh A, Chatterjee K (2017) Cloud security issues and challenges: a survey. *J Netw Comput Appl Elsevier* 79:88–115
9. Mollah MB, Azad MAK (2017) Athanasios vasilakos: security and privacy challenges in mobile cloud computing: survey and way ahead. *J Netw Comput Appl* (84):38–54
10. Patil TA, Pandey S, Bhole AT (2017) A review on contemporary security issues of cloud computing. *IEEE* 179–184
11. Radwan T, Azer MA, Abdelbaki N (2017) Cloud computing security: challenges and future trends. *Int J Comput Appl Technol* 55(2):158–172, InderScience Enterprises Ltd.
12. Reddy EK (2017) Information security in cloud computing. *Int J Comput Appl Technol Res* 3(8):10–514
13. Stergiou C, Psannis KE, Kim B-G, Gupta B (2018) Secure integration of IoT and cloud computing. *Future Gener Comput Syst* (78):964–975, Elsevier B.V.
14. Alouffi B, Hasnain M, Alharbi A, Alosaimi W, Alyami H, Ayaz M (2021) A systematic literature review on cloud computing security: threats and mitigation strategies. *Syst Lit Rev Cloud Comput Sec* 9:57792–57807. IEEE Access
15. Goyal D, Krishna MB (2015) Secure framework for data access using location based service in mobile cloud computing. *IEEE INDICON*, pp 1–6
16. Lian H, Qiu W, Yan D, Huang Z, Tang P (2019) Privacy-preserving location-based query over encrypted data in outsourced environment. In: *IEEE fourth international conference on data science in cyberspace (DSC)*. IEEE, pp 84–89
17. Tabrizchi H, Rafsanjani MK (2020) A survey on security challenges in cloud computing: issues, threats, and solutions. Springer Science Business Media, LLC, part of Springer Nature
18. Abusameh H (2020) Security attacks in cloud computing and corresponding defending mechanisms. *Int J Adv Trends Comput Sci Eng* (9). ISSN 2278–3091, 4141–4148
19. He Y, Chen J (2021) User location privacy protection mechanism for location-based services. In: *Digital communications and networks 7*, Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. Elsevier, pp 264–276
20. Jaiswal C, Nath M, Kumar V (2014) Location-based security framework for cloud perimeters. In: *IEEE cloud computing published by the IEEE computer society*. IEEE, pp 56–64

Effect of esports Among Students in COVID Era



Ankit Bisht, Hitesh Kumar Sharma, and Tanupriya Choudhury

Abstract Games in any form have always been a part of human life. These play a very important part in socializing and competitive enhancement of human nature. With the invention of computer and various other portal electronic devices, video games came into picture. Advancement in computer technology in last few decades and accessibility of the same has exponentially increased the number of people playing video games. These video games in competitive form are called esports which stands for electronic sports. Nevertheless with the recent outburst because of COVID, the engagement on the same has peaked like never before. Further, these have some positive and some negative outcomes on the life of people.

Keywords Esports · COVID-19 · Cathode-ray tube · Lockdown

1 Introduction

1.1 Early Era

COVID-19 pandemic impacted the outdoor or physical sports in negative way, but this pandemic impacted video games in positive way. The game development companies noticed and exponential growth. It uses the same technology as was there in radar display where an analog device was used to control a vector drawn dot on the screen to stimulate missiles being fired at targets, which were drawing fixed onto the screen [1]. MPL is one of the popular gaming platforms which provide multiple gaming options on single platform. There is a simple tennis game where the players to keep hitting the ball each time it enters their side of the screen with a virtual line controlled with external buttons connected to a knob. Pong was not like any of the fancy graphics video games like today. It uses a simple cathode-ray tube display which used to show a side view of a tennis court represented by just two lines, while one of the line

A. Bisht · H. K. Sharma (✉) · T. Choudhury (✉)

School of Computer Science, University of Petroleum and Energy Studies, Energy Acres, Bidholi, Dehradun, India

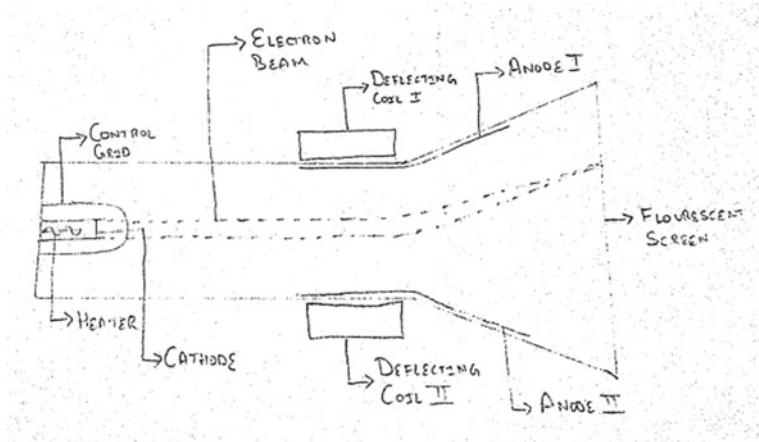


Fig. 1 Cathode-ray tube

represents the ground, and the other one represent the net. The ball was represented by a dot that bounced back and forth on the screen. Players also had to count the score for themselves. (Fig. 1).

The term video game was coined to distinguish the class of electronic games that were usually played on some type of video or screen display. Another term for electronic games was coined as “computer games” which was used to describe these video games as all video games basically require a processor.

Earlier the access to video games was very limited. Only people who can afford the machines were able to play them; thus, the player count was very small. Later people came out with an idea of public video game parlors, where people can play/use these machines on rental basis. This increased the popularity of video games as more people got involved and competitive gaming came into picture. People tried to beat the best record and were motivated to play them more often. Community tournaments also took place, where people from different places came together to compete and become the best community player among all. This somewhere increased the sportsman ship among people and evolved them for a completely different era of sports.

1.2 *Esports*

Esports is short form for electronic sports which is a form of competition using video games. Players compete each other as a part of team or individual. The most common genres of electronic sports are multiplayer online battler arena as termed as MOBA, first-person shooter called FPS, fighting games, card games, battle royal games, and real-time strategy games. The fighting game known as street fighter 2 developed in 1991 started the concept of direct competition among two players. Before that,

competition mainly occurred based on high score of the two players, but in street fighter, players challenged each other directly in a face to face fight to determine who is the best player. In 1996, the first ever Evolution Championship Series (EVO) esports tournament was organized. Another famous esports tournament organized in the 1990s was Nintendo World Championships, where a total of 132 finalists played the final round in San Diego, California.

1.3 Modernization

The modern esports ecosystem includes spectator fans all of the world; offline and online, skilled esports athletes, amateur and professional teams that focus on single video game, event managers and organizers, publishers of that particular video game, global broadcasters, and finally businessmen. Now with the growth in high speed Internet and more number of people with personal computers and handy mobile phones, the growth of esports has seen a large surge of players from around the world. According to the world statistics 2021 [2], the maximum number of gamers are from Asia Pacific region following Europe and other areas.

Today, many online platforms offer these video games on rental or buy basis. Platform like Steam has dedicated store market where video games of different genre can be browsed and purchased according to player's interest. Moreover, these platforms have created charts and user analysis to not only understand likeliness of these games but also popularity over time. Taking it further, they have also maintained dedicated scoreboards for every game. This whole concept has drastically increased the competition among player and thus number of them across the globe.

2 COVID-19 Impact

The pandemic coronavirus COVID-19 is such a dangerous infectious viral infection which enforced almost all countries across the globe to impose lockdown for the each and every citizens of their country. This lockdown was in effect almost 2–3 months. The lockdown was so strict the any movement from their houses was totally restricted. Outdoors games and sports were also totally banned, and the outdoor sports viewers find their alternative options as video games or esports. The video game development industries got a high business due to this sudden conversion of sports lovers. Video games and esports business have shown an exponential growth in their revenue. We have also seen that learning and teaching community found the alternatives online options to continue their efforts in the same way sports influences also find their alternatives on Internet or OTT platform. There are so many gaming platforms available to satisfy the need of this community. As a result of the enormous number of people remaining indoors due to the pandemic, sports fans all over the world

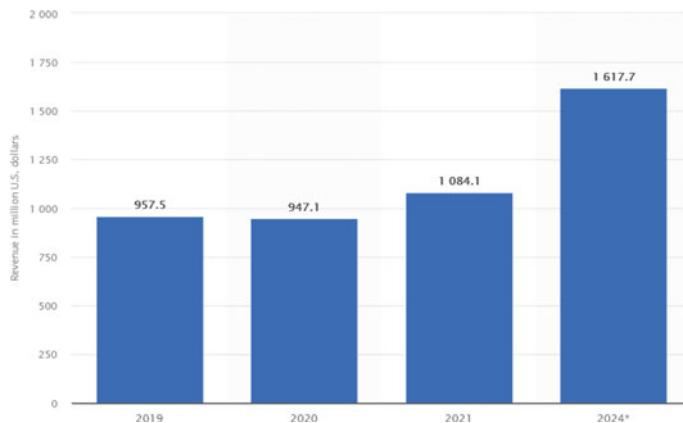


Fig. 2 Esports market revenue worldwide from 2019 to 2024. [Image Source Ref 3]

have turned to video gaming. The esports lockout has undoubtedly increased user engagement.

In 2020, the video game market across the globe is expected to be around \$159 billion, nearly four times the box office (\$43 billion in 2019) and nearly three times the total worth of music industry (\$57 billion in 2019). At a CAGR of 23.82%, the market is estimated to rise to \$2.11 billion in 2023. Furthermore, Asia Pacific region is the largest market in terms of revenue generation, accounting for about half of the total esports market [3]. These statistics shown in Fig. 2 are published by Statista on public url <https://www.statista.com/statistics/490522/global-esports-market-revenue/> and it made available for download publically.

It shows the revenue generated in gaming platform in 2019, 2020, and 2021, and it also shows the prediction of 2024.

2.1 Positive Impact

In another view, esports in schools helps more students succeed. It should go without saying that team sports teach vital abilities. For this reason, many parents have started to encourage their children to join in a sports which involved teams. Beyond the classroom, team sports provide a venue for imparting vital concepts and life skills. Any sport can benefit from the abilities learned and polished as part of a team. Esports are no different. These abilities include the following:

1. Working along with others—Call it collaboration or teamwork. Working well with people, by any name, necessitates the acquisition of numerous additional abilities like assertiveness, communication, and compromise. Conflict resolution, active listening, and respect are all important skills to have.

2. Intersocial skills—Interpersonal interactions are built on the foundation of social skills. They are also necessary for collaborating effectively with teammates and coaches. Social skills can include things like keeping eye contact and utilizing acceptable body language, in addition to the ones mentioned above. I am not going to interrupt you. Knowing when and how to communicate ideas is also important. These abilities do not come readily to many children. Those with little social experience are especially vulnerable to social skill deficiencies. Participating in a esports team from school allows these students to develop and practice these abilities in a safe atmosphere.
3. Strategic thinking and planning—Students must create goals, evaluate the competition, and consider their particular skills and shortcomings when participating in sports. Students learn to build high level tactics, putting together game plans, and alter execution as needed in a team. Any sport necessitates the ability to think swiftly and strategically.
4. Handling success and failure—Harsh lessons like life that is not always fair are taught through team sports. One would not always come out on top, no matter how well one prepare. Kids who learn to manage their emotions when they lose or win become more resilient.
5. Time management—Students must maintain their academic standing in order to participate in school teams. This necessitates the acquisition of and application of executive function abilities such as organization and time management.
6. Pro-social values—Coaches can impart positive values in their players through team sports of any kind. Good sportsmanship, fairness, and respect are examples of principles and norms of behavior. Persistence, honesty, amusement, and good competition are all important qualities. These ideals can help students succeed in life.

2.2 Negative Impact

There are multiple perspective on effect of esports among people. Large number of people think game addiction is a big issue that society must address, and it will have negative psychosocial implications. Building upon the same perspective some scientists is now considering video gaming to be a psychological problem. Adolescents who engage in addictive behaviors may have angry, disagreeable, indifferent, or uptight personalities. Many adolescents' daily lives have become increasingly dominated by sports. The association between computer games and aggression in mentally ill teenagers was investigated, and it was discovered that computer games enhance adolescent aggression while also lowering mental health. Excessive gaming addiction was found to be linked to aggressive behavior, according to similar findings. A study, popularly known as affective aggression model, proved playing violent video games increases aggressive behavior in both the short and long term among

people. Another experimental research concluded that playing violent video games can produce very high level of aggression and physiological arousal than nonviolent video games [4].

3 COVID-19 Lockdown Outcomes

Traditional patterns of education have been disturbed by the COVID-19 pandemic, which continues to provide issues for schooling systems and the pupils they serve. The education officials and teachers will be tasked with making much more difficult decisions long into the 2020–2021 school year and far beyond. Students in the class standards 3–8 performed similarly in reading skilling in the fall of 2020 to students in the same grade in the fall of 2019, but around 5–10 percentile points worse in math [5]. Since the start of COVID-19 interruptions, most students have improved their reading and math scores, and math growth was lower than in any other normal year.

There are a few remote learning platforms available online these days, notably BYJU'S, which is a Bangalore-based online educational institution and tutoring corporation started in the year 2011 [6], are now world's one of the highly valued educational companies, also offering free access of their services with respect to overwhelming demand. In COVID scene, BYJU has seen a 200% rise in the count of new students using its famous Think and Learn app. According to UNESCO, a number of children effected due to COVID by March 23, 2020, were around 1.38 billion. A real-time analysis technique is presented by authors in their publication [7].

It has been published in many research work that during the pandemic, the efforts placed in online learning and the resources used in online learning proved that the learning outcome is 30–40% more compare to offline learning. The learning and its retention rate is higher for online classes compare to classroom learning. It is because of the freedom of learning. Each student can learn the things as per his/her grasping power. The drawback of traditional learning is that it cannot be customized for individual student. Every student need to go with a predefined pace and as per predefined strategy which is not suitable for every student. Online learning makes it happen to change this traditional strategy. Now in this pandemic, students have freedom to work as per their capability. Some drawback can also noticed in this online system, which includes reduction in physical movement. An automated system is proposed based on data analytical approach for data analytics [8, 9].

A solution is provided by an author for remote sensor-based healthcare solution [10]. According to faculty, online education was manageable, helped assure distant learning, and allowed scholars to easily contact lecturers and teachers. In research work [11], author has also proposed a solution for e-healthcare in pandemic. It also

cut down on the usage of travel resources and other costs. It has also simplified admin responsibilities such as lecture recordings and attendance tracking of students. In the course of lockdown, most students and teachers agreed that online learning model has facilitated student focus. Students now have evolved into self-directed learners who can learn at any time during the day.

4 Conclusion

The study proves that, despite certain problems, students are able to adapt to a new distant learning system and proves that distance learning is somewhat more efficient than average classroom learning. Although the sudden closure of institutions around the world due to the COVID-19 pandemic is promoting esports to a large extent, it also serves as a catalyst for a cultural revolution in the educational sector. As more advanced technologies are taking a part in higher education, it is important to understand the positive outcomes for the same, which are combining the best aspects of classroom and online learning. And the same can be utilised in e-Sports for improving the overall learning experience with e-learning concepts.

References

1. Siwek SE (2017) Video games in the 21st century (PDF) (Report). Entertainment Softw Assoc. Retrieved 22 Jan 2020
2. Lemmens JS, Bushman BJ, Konijn EA (2006) The appeal of violent video games to lower educated aggressive adolescent boys from two countries. *Cyberpsychol Behav* 9:638–641
3. Gough C (2021) eSports market revenue worldwide from 2019 to 2024 Statista, March 2021. [Online]. Available: <https://www.statista.com/statistics/490522/global-esports-market-revenue/>. Accessed on: July 11 2021
4. Tarasawa B, Samuel A (2021) Learning during COVID-19: initial findings and 4 considerations for policymakers, 13 Jan 2021 [Online]. Available: <https://ednote.ecs.org/learning-during-covid-19-initial-findings-and-4-considerations-for-policymakers/>. Accessed on: 11 July 2021
5. Amir LR, Tanti I, Maharani DA et al (2020) Student perspective of classroom and distance learning in COVID-19 pandemic in the undergraduate dental study program Universitas Indonesia. *BMC Med Educ* 20:392. <https://doi.org/10.1186/s12909-020-02312-0>
6. Mukhtar K, Javed K, Arooj M, Sethi A (2020) Advantages, limitations and recommendations for online learning during covid-19 pandemic Era. *Pak J Med Sci* 36(COVID19-S4):S27–S31. <https://doi.org/10.12669/pjms.36.covid19-s4.2785>
7. Khanchi I et al (2019) Automated framework for real-time sentiment analysis. In: International conference on next generation computing technologies (NGCT-2019)
8. Kshitiz K et al (2017) Detecting hate speech and insults on social commentary using nlp and machine learning. *Int J Eng Technol Sci Res* 4(12):279–285
9. Li C (2020) The COVID-19 pandemic has changed education forever. This is how 29 Apr 2020 [Online]. Available: <https://www.wefrum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/>. Accessed on: 11 July 2021
10. Ahlawat P et al (2020) Sensors based smart healthcare framework using internet of things (IoT). *Int J Sci Technol Res* 9(2):1228–1234

11. Taneja S et al (2019) I-doctor: An IoT based self patient's health monitoring system. In: 2019 international conference on innovative sustainable computational technologies, CISCT 2019

Handwritten Digit Recognition with Neural Network



Satyana^rayana Malla V, Hitesh Kumar Sharma, and Tanupriya Choudhury

Abstract Object detection is the task of classifying and finding certain objects including people, cars, handwritten text, and more. This is a growing field with application in several fields, identifying and locating cars, pedestrians and other objects for self-driving cars, monitoring objects such as crops, or the ball during sports, facial detection for protection and many more. Handwritten text recognition comprises character recognition and digit recognition. An effective CNN based model with high accuracy can help an OCR (Optical Character Recognition) system for accurate conversion of written text into digital text. In this work, we have explained a CNN based approach for recognizing hand written digits and predict the written numbers.

Keywords eSports · COVID-19 · Cathode ray tube · Lockdown

1 Introduction

Machine learning has become one of the most impactful technologies, and it focuses on analyzing data, interpreting patterns to train machines, make predictions and can even simulate intellectual tasks performed by humans. ML has applications in various fields, one of the most important is image recognition. Image recognition refers to identifying images and classifying them into some predefined classes using various algorithms. Neural networks have multiple layers and perform multiple information extraction processes to keep learning from new input data, and the output is produced based on knowledge gained from previous states. Convolutional neural network is another such architecture where inputs are images. Images are treated as matrices having dimensions width, height and depth (RGB color), CNN can be used for training models to classify images, it transforms the input 3D value through its various layers to generate output vectors.

One of the first works on this topic was by David Marr in his book *Vision* published in 1982. He proposed that vision precedes from a 2D visual array to a 3D description

S. Malla V · H. K. Sharma (✉) · T. Choudhury (✉)

School of Computer Science, University of Petroleum and Energy Studies, Energy Acres, Bidholi, Dehradun, India

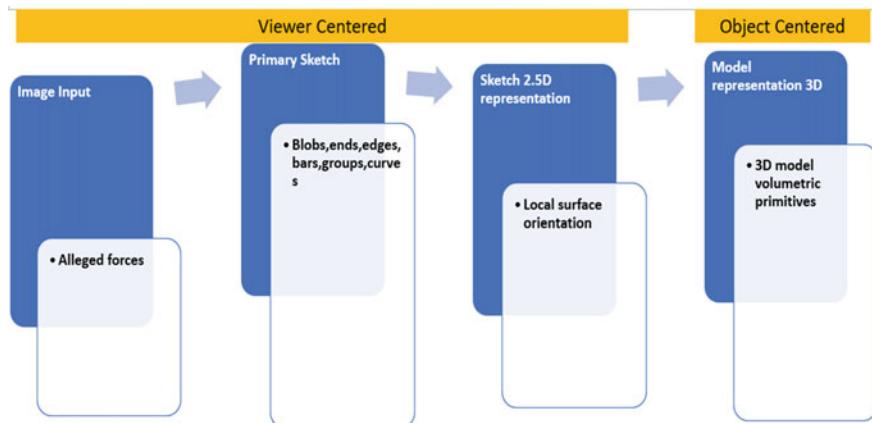


Fig. 1 David marr methodology

of the world as an output. His methodology is best explained in figure (Fig. 1). Primal Sketch is where mostly the edges, the bars, the ends, the virtual lines, the curves and the boundaries are represented, and this is very much inspired by what neuroscientists have seen. The early stage of object detection had a lot to do with edges. After the edges and curves is “two and half sketch” this is where we start to piece together the surfaces, the depth information, the layers, or the discontinuities of the visual scene, eventually we put everything together and have a 3d model.

2 Literature Review

Image classification is what a computer can identify images on its own and distinguish the “class” the picture belongs to. A class is basically a name, for example, “vehicle”, “creature”, “building, etc. For instance, we input a picture of a frog. Image classification is done in a way so that the computer will tell us whether a picture is of a frog or not.

As far as we might be concerned, characterizing images is not a big issue. However, it is a tedious task for a machine to identify images. Hence, different algorithms had been made for image classification. In 2020, authors in their work of image classification build a deep convolutional neural network for image classification (cat and dog images). An accuracy of 89.52% was obtained in their algorithm [1]. In ref [2], author achieved the best classification accuracy of the MNIST dataset with deep neural network [2]. Also, authors developed a convolutional neural network for MNIST image classification and they achieved an accuracy of 63%. One of the author have also used CNN network model for sentiment analysis [3]. In ref [4], author has used neural network model for speech analysis and provided a good accuracy. In ref [5], author has used deep learning is used for COVID-19 detection. In ref [6], author

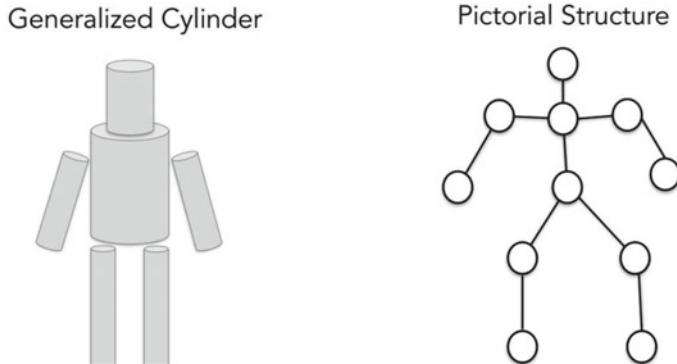


Fig. 2 Generalized cylinder and pictorial structure

used some concept of IoT and machine learning in smart healthcare. In ref [7], a neural network model is proposed for handwritten digit identification.

3 Generalized Cylinder and Pictorial Structure

The pictorial structure was introduced by Fischler and Elschlager in 1973, while the generalized cylinder model was introduced by Brooks and Binford in 1979. Both of these models were a step away from the simple block world and they started recognizing or representing real world objects. The basic idea is that every object is composed of simple geometric primitives. A person for instance could be pieced together by generalized cylindrical shapes or a person could be pieced together by a critical path in part in their elastic distance between these parts. This was a very simple model; it could possibly detect a person, but it wouldn't be able to differentiate between people based on face or expression. Below is an image of the generalized cylinder and pictorial structure side by side (Fig. 2).

4 Normalized Cuts and Image Segmentation by Shi and Malik

Instead of just focusing on the local features and their consistencies in image data they extracted the global impressions of an image. They treated this as a graph partitioning problem, where the goal is to minimize the capacity of edges between separate components. The normalized cut is used for segmenting the image, and this measures the dissimilarity between the different groups as well as total similarity within groups.

4.1 Data Set

The data used for this research work is taken from MNIST dataset which is publically available. It contains 60,000 training images and 10,000 testing images of 0–9 digits. These images are in gray scale and bitmap images. The size of these images are 28×28 . These images can be passed to the CNN model, and the model will be trained from unknown images. The images of some digits which are taken randomly are shown in Figs. 3 and 4. In Fig. 4, we have shown the number of images of different digits in graphical representation.

5 Face Detection by Viola and Jones

This was one of the first algorithms which was able to perform facial detection in real time. In a given box, this algorithm looks for features, specifically a face with

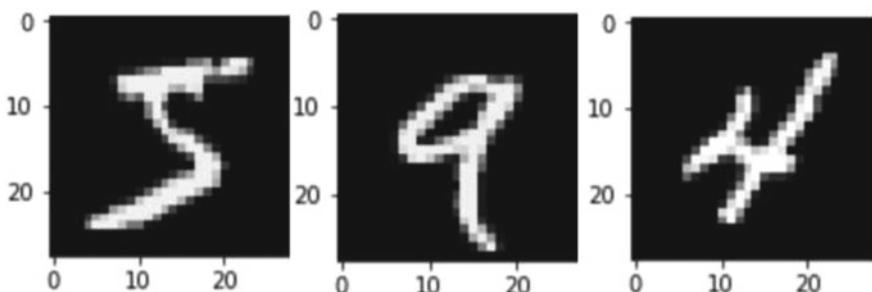


Fig. 3 The numbers from the MNIST dataset

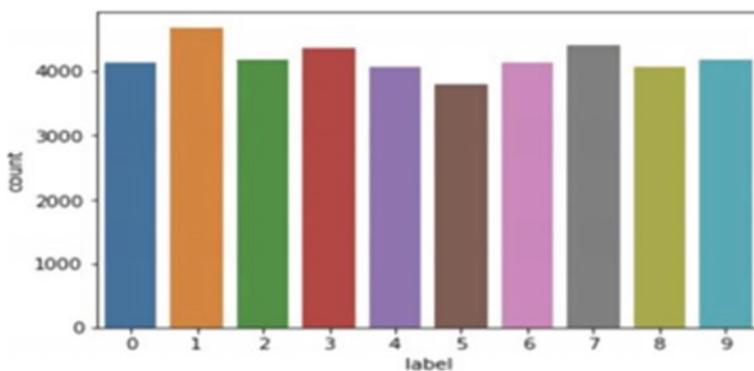


Fig. 4 The similar counts for the digits

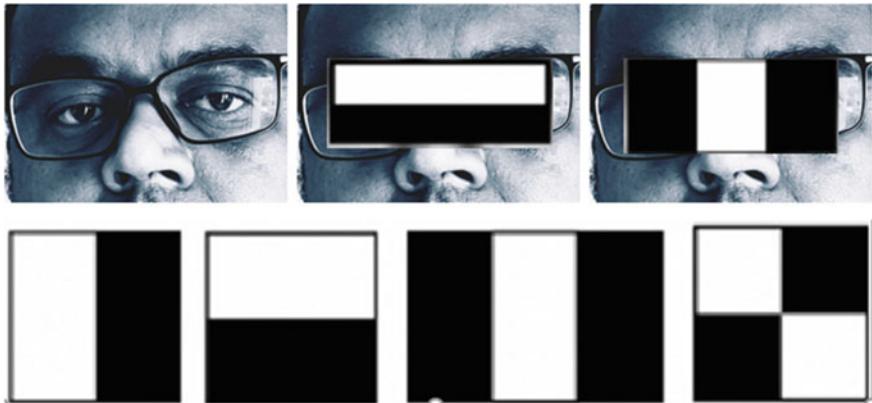


Fig. 5 Haar features used for Viola Jones face detection method

HAAR like features. HAAR features show a box with light and dark sides. There are three different types, edge features, line features and four rectangle features (Fig. 5).

The above image shows HAAR features being detected on human faces.

5.1 SIFT Object Recognition, David Lowe 1999

SIFT, short for Scale Invariant Feature Transform is a method for object detection. This method would first store all key features of an object in a database. Then, when an object needs to be detected each of the features in the objects are compared with each of the features in the database. The object where the difference between features was the least was chosen. While this method did show good results, it was very slow, since a lot of comparisons had to be made.

5.2 Spatial Pyramid Matching in 2006

This method is mainly used to classify images on a higher level for example sceneries, indoor versus outdoor and more. For such tasks looking at each and every small feature is not as necessary hence, we approach it with a holistic approach. Three different levels of histograms are made and compared.

6 Methodology

Step 1: Create a plot of the first nine images by loading the dataset from the Keras API in the training dataset (Fig. 6).

Step 2. Split the training set into a train and validation dataset to estimate the performance of a model.

Step 3. Evaluate neural network model.

Step 4. Develop a baseline model.

Step 5. Load dataset.

Step 6. Preparation of dataset (Fig. 7).

Step 7. Define model.

Step 8. Present the outcome.

Above figure (Fig. 8) shows the confusion matrix for the given neural network. It shows the cross matrix between actual and predicted values. The diagonal entries shows the similarity between actual and predicted value. And other cells show the deviated results.

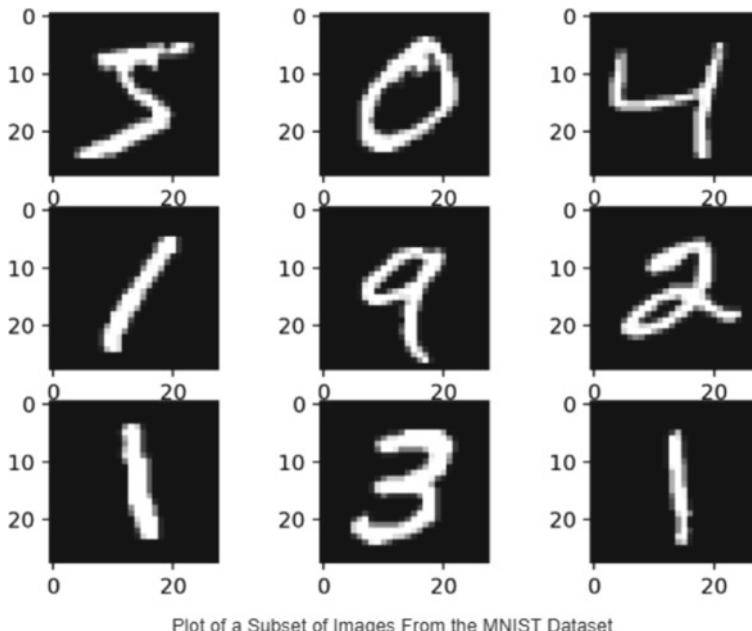


Fig. 6 MNIST dataset images

NEURAL NETWORK OUTPUT FORMAT

Rows	Handwritten digits									
	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	0	1

Fig. 7 Neural Network Output

	0	527	0	0	1	1	0	0	2	0	0
T	1	0	435	0	0	1	0	0	1	0	1
R	2	0	0	470	0	0	1	1	0	0	1
U	3	0	0	0	540	0	0	0	1	1	1
E	4	1	0	1	0	533	0	2	0	0	0
V	5	1	2	0	0	0	412	0	1	1	1
A	6	0	0	2	0	0	0	456	0	0	0
L	7	1	0	0	1	1	0	0	329	0	0
U	8	1	0	2	0	2	0	0	0	533	0
E	9	1	1	1	0	0	0	1	0	0	489
	0	1	2	3	4	5	6	7	8	9	
PREDICTED VALUES											

Fig. 8 Confusion Matrix

7 Conclusion

Image classification is an important task to identify images and is used in various fields for different purposes. From the above performed experiment results, it can be concluded that convolutional neural networks can be used building models for image classification. Above are notable methods which still see usage, but a lot of these methods have been replaced by convolution neural networks which are a type of artificial neural networks. These networks have shown promising and exceptional results in video and image detection. These networks normally have alternating convolution and activation layers with pooling layers which are all usually connected to a fully connected layer. Convolution neural networks (**CNNs**) have been successful in tasks such as facial recognition, pose recognition and image description.

References

1. Amir LR, Tanti I, Maharani DA et al (2020) Student perspective of classroom and distance learning in COVID-19 pandemic in the undergraduate dental study program Universitas Indonesia. *BMC Med Educ* 20:392. <https://doi.org/10.1186/s12909-020-02312-0>
2. Mukhtar K, Javed K, Arooj M, Sethi A (2020) Advantages, limitations and recommendations for online learning during Covid-19 pandemic era. *Pak J Med Sci* 36(COVID19-S4):S27–S31. <https://doi.org/10.12669/pjms.36.covid19-s4.2785>
3. Khanchi I et al (2019) Automated framework for real-time sentiment analysis. In: International conference on next generation computing technologies (NGCT-2019)
4. Kshitiz K et al (2017) Detecting hate speech and insults on social commentary using nlp and machine learning. *Int J Eng Technol Sci Res* 4(12):279–285
5. Li C (2020) The COVID-19 pandemic has changed education forever. This is how. Apr 29, 2020 [Online]. Available: <https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/>. Accessed on: July 11 2021
6. Ahlawat P et al (2020) Sensors based smart healthcare framework using internet of things (IoT). *Int J Sci Technol Res* 9(2):1228–1234
7. Hochuli A, Oliveira LAB Jr, Sabourin R (2018) Handwritten digit segmentation: is it still necessary? *Pattern Recogn* 78:1–11. <https://doi.org/10.1016/j.patcog.2018.01.00486>

Metastability Mitigation and Error Masking of High-Speed Flip-Flop



Reshma Kumari, Sneha Pandey, Swarnima, and Surya Deo Choudhary

Abstract In digital circuits, metastability occasions are common, and synchronizers are crucial for defending from its catastrophic consequences. When taking part in partner asynchronous enter, synchronizers have been first indispensable (that is, one synchronous with the clock enter so might also amend exactly as soon as the sample). Everything will change, but there will be consistency. One can obtain stability by way of switching its statistics enter at the identical time as the clock's sampling fringe. The relative period of each cycle for the two warning signs modifications a little, in the end ensuing instability, which is shut sufficient to all exceptional switches? On occasion, this conglomeration of consistency and favored exhibit gadgets arises. Recent semi-conductive steel chemical compound progress (CMOS) to boot influences in digital proper judgement systems at until now unseen degrees of integration. In digital circuits, failure arises due to path lengthening and temporal association clock maintain time configuration issues. To be aware of temporal association faults, the proposed flip-flops take gain of the thinking of either delayed data or a pulse-based approach. By transferring direct statistics rather than master latch output to slave latch, the temporal association violations are hidden. When in contrast to ultra-modern steadiness immune flip-flops, simulation results display that the projected flip-flops minimize error masking latency by using up to a 0.33 in conventional approach corners, so enlarging the temporal affiliation error looking window. The proposed flip-flops additionally are employed in purposes involving dynamic voltage and frequency.

Keywords Metastability · SAFF · CMOS · Flip-flop · Synchronizing · PDFF · Latch

R. Kumari · S. Pandey (✉) · Swarnima · S. D. Choudhary

Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida, India

Swarnima

e-mail: Swarnima.ece@niet.co.in

1 Introduction

The motive of enforcing the setup and hold time situations on combinational paths is to constrain the entire of every turn-flop: to make sure that it is held solid for at least (t_{su}) seconds before the clock part and that it remains strong for no much less than seconds afterwards. By doing so, flip-flop outputs are guaranteed to behave in a predetermined manner: the transition to the logic stage of the input monotonically, with a nominal transition time and within a nominal clock-to-q postpone. These homes are vital for the design of deterministic synchronous structures [1–3]. The scaling of CMOS devices is the primary reason for improved performance of digital integrated circuits [2]. Process, voltage, and temperature (PVT) variations in scaled technology nodes cause significant performance uncertainty in the digital designs. Timing or supply voltage guard bands are added to maximum operating frequency (MOF) or minimum supply voltage to cope with PVT variations [1].

2 Proposed Work

1. To indicate on the other hand the planned flip-flop is employed, a pc stage implementation of the planned flip-flop provided a series dominating latch and a clock gating controller place unit covered within the laptop stage implementation for error recovery. The implementations of a series dominating latch and clock gating controller at the circuit diploma vary from the modern methodologies inside the literature.
2. In a separate element, the advised flip-flop is checked for statistical route gadget flaws and metastability.
3. In ISCAS'89 benchmarks, the counseled flip-flop is used to examine. In the worse designs, large electrical energy consumption by using variable voltage and strength scaling (DVFS) features. The benchmark circuits are implemented in 138 nm CMOS science through an industrial scale, with a nominal grant voltage of 1.2 V.
4. Four of the simulations follow the procedure to get the diagram's minimal walking voltage or most walking frequency through combining the recommended FF with the DVFS.

3 Description of Test Circuit

When the most enter signal ends with a violation of installation time and maintain time requirements, a latch has the chance to enter a metastable condition. This is the case when the D enter of flip-flop changes is used at each clock facet. Any difference in the frequency of the sign on the D entry in contrast to the clock frequency reduces the chance that the latch under investigation will enter a metastable condition. When

the frequency (f_{in}) on the D enters is exactly half of the clock frequency (f_{clk}), the worst case state of affairs occurs.

Go with the flow in terms of design. We used digital schematic editor (DSCH) to create the frontend layout, which is primarily based on a virtual circuit and makes use of a non-public in-built ground simulator. Users can additionally construct an analog circuit and convert it to digital.

4 Functional Simulation

Microwind assists all front-end builders in lowering back-end documents and utilizing third-party simulators such as Microwind or pspiceSPICE. DSCH can convert digital circuits in Verilog archives to be synthesized for FPGA/CPLD gadgets from any manufacturer. Setting Microwind can also be transformed by the usage of the identical Verilog record. The built-in combination signal simulator, as well as analyses for the DRC, delays, location 2D, 3D view, and other CMOS provisions, can be used to confirm CMOS provisions.

5 Simulation of the D Flip-Flop

Because our instance is a D flip-flop that makes use of splendid side touchy thing elements, the Q output follows the D enter when the clock adjustments from low to excessive. There is no doubt that sound judgement is bestowed at the time of the clock because the date is consistent each earlier than and after the clock. However, it rapidly became a simple illustration. We are actually worried with a small element of the clock, ranging from low to high. It is not possible to overlook a terrible aspect. Peering Q and D at the clock above is especially smooth. As an example, look at Figure 1.

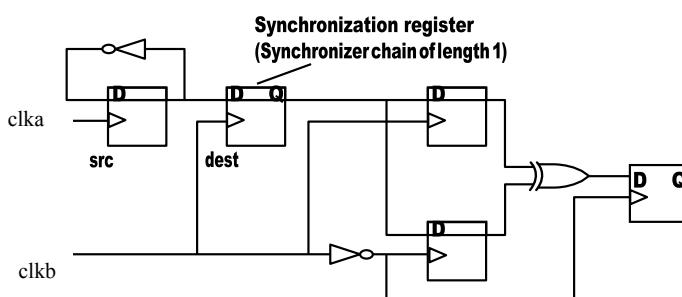


Fig. 1 Test circuit structure for metastability characterization

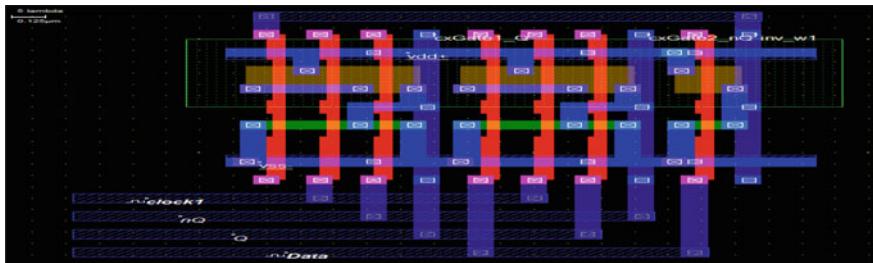


Fig. 2 CMOS design basic D flip-flop

The estimated metastability window () values of the turn-flops and their quicker and slower output propagation delays are compared for quantitative evaluation (Figs. 2, 3, 4, and 5).

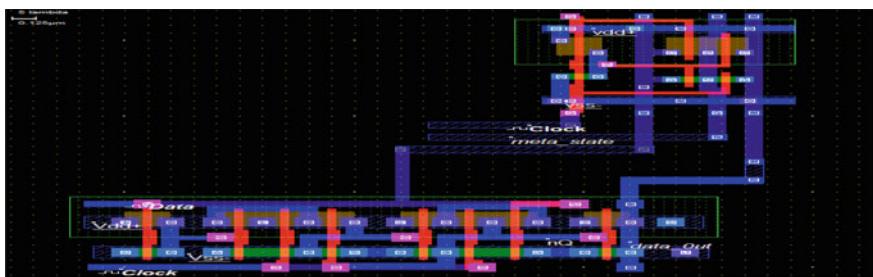


Fig. 3 CMOS design metastability of basic flip-flop

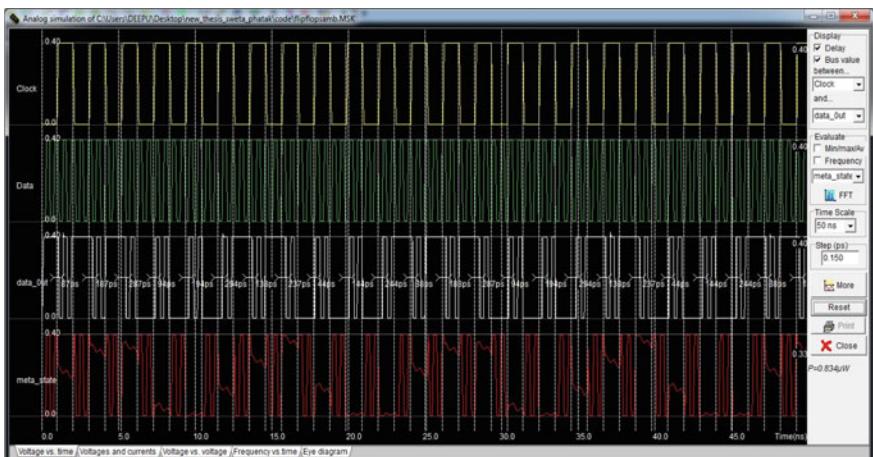


Fig. 4 Metastability of basic flip-flop voltage versus time

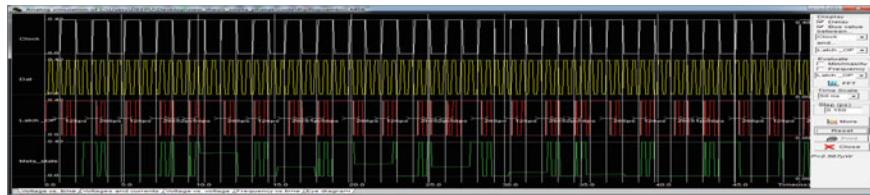


Fig. 5 CMOS design metastability of proposed flip-flop voltage

Comparison of Proposed Design with Conventional Designs:

Designs	Power dissipation (μW)	Clock frequency (MHz)	Number of transistor	Metastable error
[1]	372.4	500	50	11.0
[2]	394.9	416	45	11.2
[3]	476	250	67	13.2
Proposed design	2.577	1700	22	10.0

The artwork which has been proven in Fig. 6 suggests the metastable errors 10 compared with the artwork carried out inside the previously given artwork which heaps greater better overall performance as it offers us the mistake that is 10; however, previously it changed into eleven to forty-nine. The electrical power loss and switching lengthen are two definitely different things. The metastability extend is induced with the aid of electricity dissipation, which influences the circuit's common performance.

The flip-flop circuit must be designed to make use of a transmission gate that decreases stray capacitances and permits for a vast range of transistors. The reduction in stray capacitances increases the circuit's put off and energy dissipation. A metastability event making an attempt out circuit is constructed and the usual overall performance of metastability robustness in forty five nm turn-turn is tested. Using faster flip-flops reduces the turn-setup flop's and maintains conditions, which reduces the time window during which the flip-flop is inclined to metastability. As the enter frequency lowers, the odds of the enter altering at some stage in the setup and hold time cut down as well.

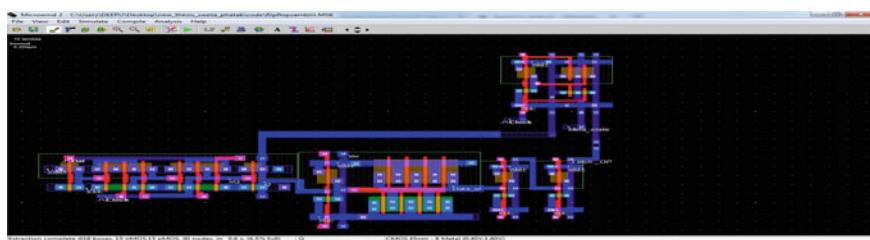


Fig. 6 Metastability of proposed flip-flop voltage versus time

6 Conclusion

The circuit performance degrades as a result of the strength dissipation, which is reflected in the metastability ejected. The turn-flop circuit must be designed with the use of a transmission gate to reduce stray capacitances and transistor quantity. The put off and strength dissipation in circuit are improved by reducing stray capacitances. Using faster flip-flops reduces the turn-setup flop's and protection times, decreasing the time window at some point of which the turn-flop is prone to metastability while the input frequency lowers. The odds of editing the input at some factor throughout the setup and saving time are additionally reduced. By synchronizing one or more subsequent synchronizing flip-flops, our circuit recalls metastable actions and tolerates them. It is presented how to decrease the worst-case timing guard bands by way of the usage of an errors masking flip-flop. In this technique, an error indicator is acknowledged due to timing breaches in the event of turn-flop mishaps. The clock gating controller utilizes this error signal to shift the glorious fringe of the clock one cycle to compensate for timing breaches. Because it is proof towards records route metastability, the advocated flip-flop eliminates the want for a metastable detector to prevent timing violations. To validate the counseled timing error protection technique, sizable simulations were achieved in every of the turn-flop and block levels. Compared to the trendy technique in the literature, the proposed flip-flop minimizes the extent of mistake covering via sixteen percent.

7 Future Scope

There are quite a few methods to lengthen our work to accommodate emerging concerns in circuit reliability. First, we recommend proceeding to tackle the central trouble of improving the scalability of precise reliability computations and the use of edge-valued choice diagrams. Next, we suggest enhancing the SER of sequential circuits by way of taking advantage of the expanded resynthesis opportunities available flip-flops are an additional goal for errors in a sequential circuit. Protecting flip-flops from metastable mistakes is integral due to the fact that latched tender errors cannot be eliminated by way of electrical or timing masking. Several methods have been proposed to harden latches or add more common sense to defend latches. However, these methods tend to incur high vicinity overhead. We advocate using retiming to enhance logic overlaying barring high location overhead. Retiming refers to the procedure of relocating flip-flops, usually to limit the place or the clock duration in a given circuit. The minimum-area and minimum-period retiming problems can both be formulated as linear packages (LPs) and solved by the usage of the simplex method. Therefore, specific optima are possible using retiming. The thought of SER-aware retiming is to quit error propagation from latches to any predominant outputs through relocating latches to regions of low observability. Recall that sequential

observability is calculated by transforming latches into buffers and expanding the circuit by countless time frames.

Acknowledgements In the absence of a mother, the beginning of a baby is no longer possible and in the absence of a trainer, the right direction of know-how is impossible. This mission is by ways the most extensive accomplishment in our existence and it would be impossible barring human beings who supported us and believed us. We would like to lengthen my gratitude and my honest thanks to my honorable, esteemed guide.

References

1. Sannena G (2018) Low overhead warning flip-flop based on charge sharing for timing slack monitoring. In: IEEE transactions on very large scale integration (VLSI) systems vol 99. pp 1–10 Apr 2018
2. Wang S (2017) A metastability-immune error-resilient flip-flop for near-threshold variation-tolerant designs IEICE Electron Expr 14(11), May 2017
3. Sannena's G (2016) A metastability immune timing error masking flip-flop for dynamic variation tolerance. In: Conference: the 26th edn, May 2016 <https://doi.org/10.1145/2902961.2902976>
4. Jahanuzzaman SM, Islam R (2010) TSPC-DICE: a single phase clock high performance SEU hardened flip-flop. In: Proceeding of the IEEE international midwest symposium on circuits and systems, pp 73–76
5. Jahanuzzaman SM, Rennie DJ, Sachdev M (2009) Soft error robust impulse and TSPC flip-flops in 90nm CMOS. In: Proceeding of the microsystems and nanoelectronics research conferences. Ottawa, ON, pp 45–48
6. Krueger D, Francom E, Langsdorf J (2008) Circuit design for voltage scaling and SER immunity on a quad-core Itanium® processor. ISSCC Dig Tech Papers 94–95
7. Jahanuzzaman SM (2008) Modeling and mitigation of soft errors in nanoscale SRAMs. Ph.D. thesis, University of Waterloo, Ontario, Canada
8. Oliveira R, Jagirdar A, Chakraborty TJ (2007) A TMR scheme for SEU mitigation in scan flip-flops. In: Proceeding of the internatioanlsymposium quality electronic design, pp 905–910
9. Baumann RC (2005) Radiation-induced soft errors in advanced semiconductor technologies. IEEE Trans Device Mat Rel 5(3):305–316
10. Karnik T, Hazucha P, Patel J (2004) Characterization of soft errors caused by single event upsets in CMOS processes. IEEE Trans Dependable Secure Comput 1(2):128–143, Apr-Jun 2004
11. Wang W, Gong H (2004) Edge triggered pulse latch design with delayed latching edge for radiation hardened application. IEEE Trans Nucl Sci 51(6):3626–3630
12. Dodd PE, Massengill LW (2003) Basic mechanisms and modelling of single-event upset in digital microelectronics. IEEE Trans Device Mat Rel 50(3):583–602, Jun 2003
13. Mavis DG, Eaton PH (2002) Soft error rate mitigation techniques for modern microcircuits. In: Proceeding international reliability physics symposium. Dallas, TX, pp 216–225, Apr 2002
14. Baumann RC (2001) Soft errors in advanced semiconductor devices—part I: the three radiation sources. IEEE Trans Nucl Sci 1(1):17–22
15. Baumann RC (1991) Investigation of the effectiveness of polyimide films for the stopping of alpha particles in megabit memory devices. TexInstrum Tech Rep

Student Performance Prediction Using Technology of Machine Learning



Kaushal Kishor , Rahul Sharma , and Manish Chhabra

Abstract In the given paper, the main focus of this report is education. Student performance prediction is our main target. Various factors have been taken into account to create a model used for student performance prediction. This helps to analyze the student's study environment so that his success rate increases in the field of studies. Our project makes use of various effective machine learning algorithms for creating the predictive model. Mainly, it is based on linear regression, decision trees, Naïve Bayes classification, K-nearest neighbors (KNN), and some improvements carried out through feature engineering that modifies the data to make it easier in understanding for ML. Data sets containing students' information are arranged in a tabular format. The row represents the name of the student, while each column contains different details about the student such as his background of the family, sex, any information about medical reports, and age. An additional column contains the variable of success rate that the algorithm is trying to predict. The final report is evaluated through these algorithms in which a function outputs whether the student can be successful or not. "Feat Hunch–Student Performance Predictor in ML" aims at connecting all the students and teachers in an institute.

Keywords Student performance prediction · Naïve Bayes · ANN: Logistic regression · Decision tree

K. Kishor

Department of IT, ABES Institute of Technology, Ghaziabad, U.P, India
e-mail: kaushal.kishor@abesit.in

R. Sharma

ABES Institute of Technology, Ghaziabad, U.P, India

M. Chhabra

Department of CSE, Koneru Lakshmaiah Education Foundation, Hyderabad, India
e-mail: ch.manish@klh.edu.in

1 Introduction

The student's scholastic performance focuses on different aspects, creating analysis little bit difficult. In upcoming years, there has been a rise within the percentage in rate of interest and concern over individuals within the use of data mining for analyzing academic qualities [1]. Data processing depicts growing and upcoming areas of researches in education, and it has separate discrete needs that some fields lack. During this project, the performance analyzes of scholar's are mentioned. The goal at of this project is providing students' performance using given strategies through different algorithms [2]. A lot of studies in this field are that investigate the ways for applying techniques associated with machine learning in educational fields. It focuses on identifying high-risk students and also student performance [3]. In ML, a system predicts using knowledge. While ML's capabilities change, its purpose remains the same [4]. The machine searches huge data sets for definitions and information [5]. These rules and definitions are calculus in nature and may be used to define and process by computer. This study evaluates the efficiency of ML algorithms and techniques [6]. This article uses linear regressions, Bayes classification, and decision trees as well as other related techniques for building predictive models. The technique of feature engineering is assessed and then recommends data that may be changed and built in such a manner that ML can comprehend it. In this method (machine learning), categorization rate is often a deciding feature. That a classifier should only utilize reasonable time and memory for training and application was specified by Gaga in 1996. Data processing is one of the emerging areas of study in education, and it has distinct requirements from other fields. Scholars' performance is evaluated in this article review [7].

The rest of this paper is structured as follows. Section 2 goes through background information and associated studies for certain current mechanisms. Section 3 describes the proposed method and architecture. Section 4 describes our simulation environment as well as the test findings. Section 5 discusses comparative analysis. Section 6 discusses our conclusions.

2 Literature Review

The study conducted by Kotsiantis [8] for prediction of dropouts in distance learning using machine learning is the first studies to examine the learning technology. The main powerful donation through this study was a primary and rounded the track for many such researches. While this ML technique has already been imposed on other settings, it was first time being implied on academic environment. Hardwar and Pal conducted a survey across India in which students were shown to be the most influential factor which was most important. This performance took place in Faizabad. In the life of a student, he used the Bayesian classification to study it [9]. In the paper, "Data Mining Approach for predicting performance of a student"

published by Hasan et al. [10] they implemented 3 machine learning algorithms with supervision for prediction of success in a course. It was found out that Naïve Bayes' classified performed way better in decision tree prediction and other methods of neural network. Snehal Kekane published "Automatic predicting performance of a student and Automatic Student Performance Analysis and Monitoring" that proposed a system to display results of the students' performance through the user and helps in releasing staff pressure [11]. In the paper titled Student's Performance Analysis Using Machine Learning Tools disseminated by Atul Prakash Prajapati, et al. employed some characteristics in the upcoming given framework during its phase of implementation. Further performance in user interface predicted. There were some of them who ensured that the goal is achieved [12]. In the Jain and S. Solanki (2019), published in this paper, he analyzed the student perform testimonial and data mining point of view there he exhibited a survey cum experimental method used to create a database using the primary as well as secondary data source. It has been proved here that multilayer perturbation is considered the best performing qualifier for predicting student performance. The research conduct by Err [13] upon was based on S. Kotsiantis along with other stuff studies. They conclude that the Naïve Bayes base algorithm actually performs better than any other machine learning algorithm. However, this research helped to evaluate time learning characteristics that prevent the process of learning of ML, and therefore, it was recommended to keep out of the research completely. It was then evaluated that demography of students was lower in characteristics than the previous attending rate and homework rate of students to forecast rates in earlier stages. A common issue among ML and data mining researchers is the classification technique employed. It predicts attribute values based on other attribute values. Some techniques are utilized [14, 15]. With the help of WEKA, Kaunang and Rotikan [16] developed a novel prediction model for students' academic performance. The Decision Tree outperforms the Random Forest. The findings show variables influencing the student's learning behavior. These findings may be used to evaluate student learning behavior. To improve performance, add additional characteristics to the dataset; utilize different categorization methods for accurate findings and comparison analysis.

2.1 Naïve Bayes Algorithm

A common issue in ML and data mining is classification. Axiom predicts attribute values based on methods are utilized. Bayes categorization uses the Bayes law [17]. The Bayes theorem may be used to determine a feature's probability. Nowadays, several types of deep neural networks are deployed - each with advantages and disadvantages depending on the application. It is unlike other algorithms since it models the human nervous system. They can spot and use non-linear data patterns. I want more. Pupils' performance is improved, while at-risk students are identified. Ability rankings may be extremely helpful. Effective student training. True. Prospective speakers may classify themselves in naive Bayesian. A simple parameter selection

is what makes naive Bayes models ideal for big data. Most of the time, the Naive Bayes classifier is superior. The Bayes theorem [18] can compute $P(\text{clx})$. Suppose no prediction affects other future speakers. Conditional independence of classes [9]

Bayes' Theorem-

The Bayes theorem is a mathematical equation that predicts the likelihood of an event occurring.

$$P(A/B) = [P(B/A)P(A)]/[P(B)]$$

$P(B) = 0$ when A and B are occurrences.

We attempted to find a chance of occurrence A if B is true. Case B is also evidence.

$P[A]$ is A 's priori (the prior prospect, i.e., Prospect of event before authentication is found). The authentication is a random figure of an unknown sample, like case B .

$P(A|B)$ is the a posteriori probability of B occurring after authentication two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

2.2 Neural Network

The artificial neural network is mostly called just a neural network. One is the model that is inspired by biological neural networks. A network of systems consists of a purse of man-made construction, and it modifies the information using a connection-oriented approach in a communication way. Various inputs are transferred through the network, and their specific output is taken out after passing the hidden layer [19, 20].

2.3 Existing Methodology

Bhardwaj and Pal [21] studied 300 students from five colleges who applied to the Dr. R. M. L. Awadh University, Faizabad, India. This study found that kids' success in school is linked to their parents' credentials, their family income, and their social standing. In the present research, values with credibility values more than 0.70 were considered as understandings and overwhelming influencing values. This feature was used for projection of sample construction. The selection of variables and projection construction of model, the publisher has used MATLAB [6]. It was concluded the next high possibility for student good performance is their home place, and the third is teaching. Like in UP native language is Hindi. Hence, students are more comfortable in Hindi and less in English.

2.4 Existing Methodology Singularity

Erkan Er managed a study that [7] confirmed important for the uniqueness in projected application. His/her work finish that all exiting machine learning applications in academies were on prediction of dropping out rates for distance learning courses such that no applications that attempts prediction of student performance.

3 Proposed Methodology/Architecture

The diagram shows in Fig. 1 the important processes as steps, and their sources involved in the proposed ML system.

Step-1: Firstly, the data sources provide the data for collection. The other data sources include surveys and the results of students.

Step-2: In the second phase, data is processed for getting the normal dataset and getting the rows labeled.

Step-3: The next step involves containing the previous step results and the training dataset to the machine learning algorithm.

Step-4: Now the ML Algorithm is ready, it designs a model based on the training data to verify the model through the given data used in test.

Step-5: Finally, the ML algorithm has developed a trained model or a classifier that takes data as data row in input and label is predicted.

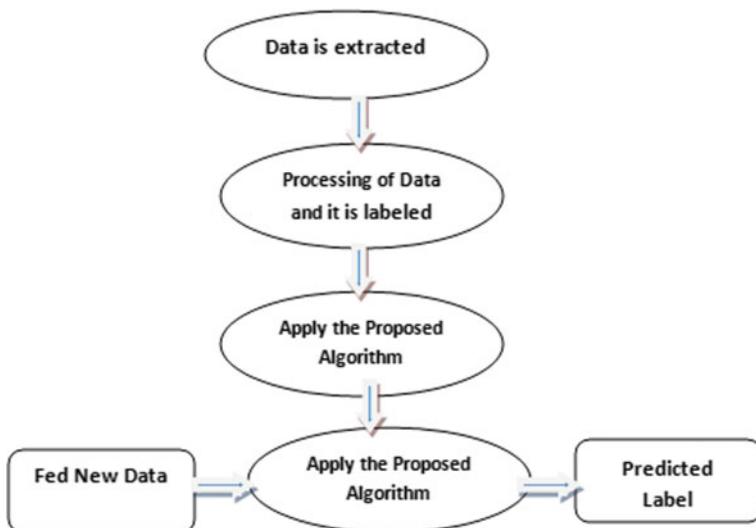


Fig. 1 Steps and components of the proposed system

The performance is calculated based on student's outcomes of learning method. Different assessments form various other outcomes giving valuable info about the student method of learning, and how much the teaching method is accurate. Still grading is much more important. It is the standard method of predicting level of a performance. Grade is divided into 2 types like Grade point Average and cumulative Grade point Average. GPA is evaluated by how much student is earned grade points in given time period. Student education is another recognized participative factor related to progress. Here, some basic things like gender, how many people are in the family, age, parents' cohabitation, how far the parents have been educated, and other things have been mentioned in this inquiry paper.

3.1 Steps Implemented of Proposed Model

This dataset consists of age, gender, grade point, father's job, and home status, lack of study, romantic activities, alcohol consumption, rotation, extracurricular activities, number of students, and three different semesters.

Step-1: The next stage is pre-processing in which any data with zero or zero values are released.

Step-2: In the third stage, after pre-processing the data is selected according to the requirement of the particular object.

Step-3: Then in this phase *K*-Nearest Neighbor (KNN) and Naive Bayesian classification Algorithm are used to training and testing the model, respectively, implemented in Python.

Step-4: In the last step, *R* language is use to graphically represents the GPA attached by a student and one after another among the above elements.

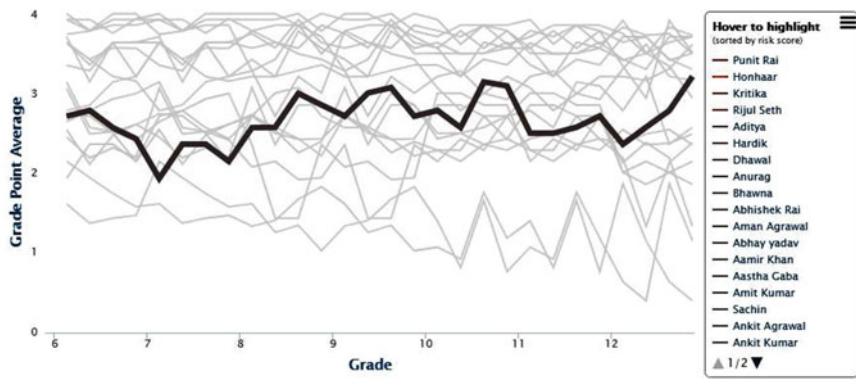
3.2 Data Selection

Only those fields were chosen in these stages that were necessary for data mining. A small number of derived variables were chosen. While some of the data for the variables was taken from the database, some were not.

4 Result

The student performance report card in various areas is shown in the graph 4, which is based on the input data set used in the proposed algorithm.

The overall study of Bhawna's (as per record of data set) report card reveals that her report card does not provide favorable results owing to a high-risk score, which is caused by a poor GPA and high absence rates, as shown in Fig. 2.



Report Card for Bhawna

**Fig. 2** Graph analysis of student performance using given algorithms

The graph shown in Fig. 3 can be suggested that shows the performance of Ankit Agarwal based on his report card. Factors like GPA; GPA is evaluated after the number of points is collected of a student, and its average is calculated for a given time period. Attendance rate, penalties, and migration are all dependent factors in computing the student's ultimate risk score. Overall analysis for Ankit Agarwal shows that her report card does fetch good result because of low (0) risk score which are due to high GPA and low absence rates. The final result can be visualized in a graph. This report provides the information about the student's performance. The factors considered here are student's GPA, Risk Score, Absence rate, Suspensions, and Mobility. Absence rates are the percentage of how a student is regularly attending a class. Suspension is also calculated for analyzing the discipline of a student and his behavior in the class. Mobility is a dependent factor for absence rate as students have to move from other cities has to move the institute during holidays, and it can cause a spike in absence rates.

Mobility is a dependent factor for absence rate as students have to move from other cities has to move the institute during holidays, and it can cause a spike in absence rates. The result of Ankit Agarwal is therefore calculated and final report is achieved as shown below.

Risk Score: 0

GPA: 3.85

Absence rate: 0.09%

Suspensions: 00

Mobility: 00.

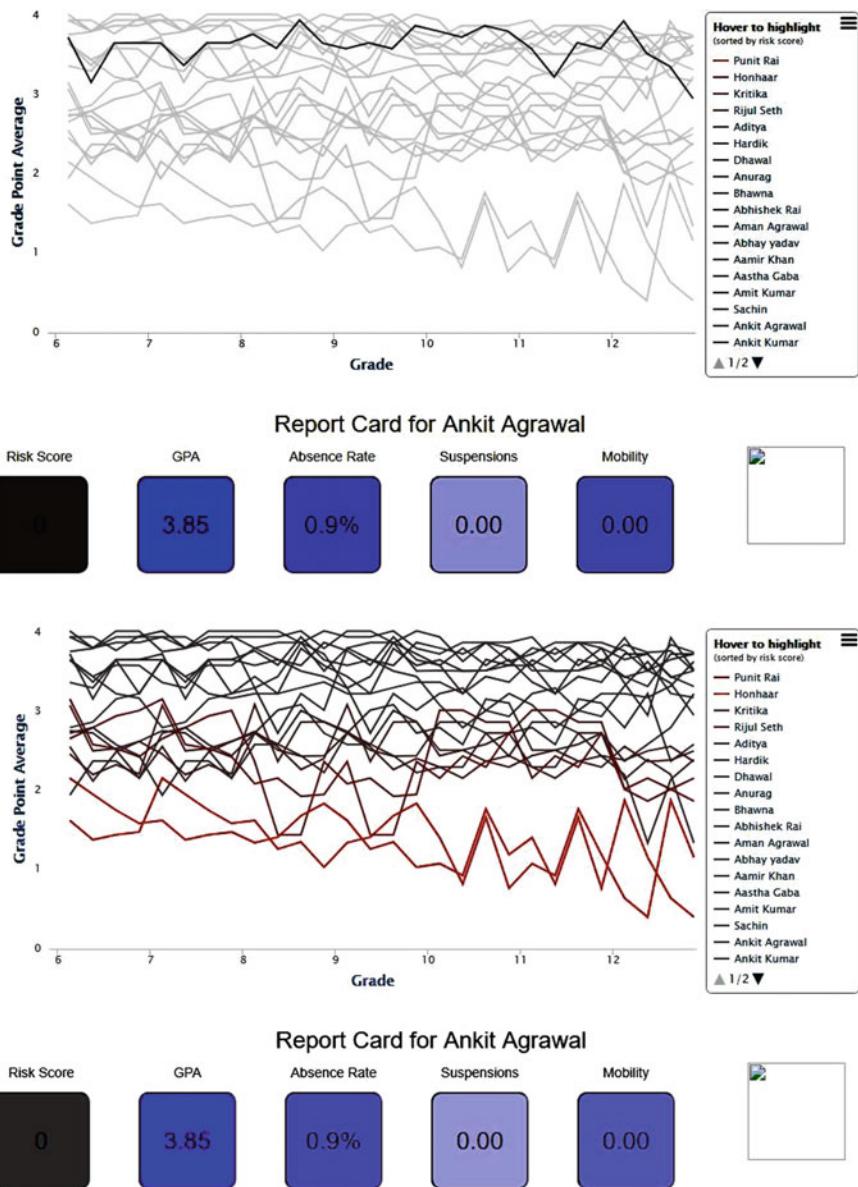


Fig. 3 Graph analysis of student performance using given algorithms

5 Comparative Analysis

In a previous study, Bhardwaj and Pal selected 300 students from five different degree colleges who were enrolled in the Dr. R. M. L. Awadh University's BCA (Bachelor of Computer Application) program in Faizabad, India. It was discovered that factors such as students' grade in senior secondary exams, living location, medium of instruction, mother's qualification, students' other habit, family annual income, and student's family status were highly correlated with student academic performance using the Bayesian classification method on 17 attributes [11].

Attribute	Proposed model	Existing model
Dependency of element	Our suggested approach has a significant effect on students' performance. Our approach will enhance insights over current approaches	A prior research showed that students' academic success is not necessarily self-driven. Others
Analysis or prediction	It is essential to notice the provided statistics since they will help us predict the student's future success. The risk score helps a student assess his overall capacity to acquire and apply new skills. It all relies on his GPA, absence rate, and other things. Suspensions and mobility problems may substantially impact a student's performance and are kept in mind	Not mentioned GPA, absence rate, and mobility problems
Highly influential factors with high probability	The present study took into account highly important variables with probability values greater than 0.70. These traits were used to create a prediction model. I utilized MATLAB to pick variables and build prediction models	But in the current model, variables with probabilities smaller than 0.70 were investigated [9]
Prediction provided using large data	Institutions may utilize this method to track student progress over time. This may significantly assist institutions achieve performance consistency over a large number of students	Existing approach has flaws for complex data sets
Factor consideration	GPA, Risk Score, Absence Rate, Suspensions, and Mobility are incorporated in suggested model student	Not included in current model

6 Conclusion

For the benefit of the student and his/her performance, it is important to take note of the given stats, and hence, it will give us an idea of his/her future performance. The risk score is beneficial for a student to understand his overall capability to learn and implement new things. It majorly depends on the basic idea that how his GPA scores, Absence rate and other factors perform. Other factors like suspensions and mobility issues can significantly degrade a student's performance so they are kept in analysis. Overall analysis for Kritika shows that her report card doesn't fetch good result because of high-risk score which are due to low GPA and high absence rates. Institutions can make use of this system to analyze the performance of their students over a long period. This can significantly help the institutes to attain stability across a large number of students in their performance.

References

1. Oyelade OJ, Oladipupo OO, Obagbuwa IC (2010) Application of k means clustering algorithm for prediction of students academic performance. ArXiv preprint ArXiv: 1002.2425
2. Roy C, Pandey M, Rautaray SS (2019) A proposal for optimization of horizontal scaling in big data environment. In: Kolhe M, Trivedi M, Tiwari S, Singh V (eds) Advances in data and information sciences. lecture notes in networks and systems, vol 38. Springer, Singapore
3. Sikder MF, Uddin MJ, Halder S (2016) Predicting students yearly performance using neural network: a case study of BSMRSTU. In: 2016 5th international conference on informatics, electronics and vision (ICIEV). Dhaka, Bangladesh, pp 524–529
4. Jain A, Sarma Y, Kishor K (2020) Financial supervision and management system using MI algorithm. Solid State Technol 63(6):18974–18984, Publication Year: 2020
5. Alshabandar R, Hussain A, Keight R, Khan W (2020) Students performance prediction in online courses using machine learning algorithms. In: 2020 international joint conference on neural networks (IJCNN). Glasgow, UK, pp 1–7
6. Tyagi D, Sharma D, Singh R, Kishor K (2020) Real time ‘driver drowsiness’ & monitoring & detection techniques. Int J Innovative Technol Exploring Eng 9(8):280–284, June 2020. ISSN 2278–3075
7. Nonis SA, Hudson GI (2006) Academic performance of college students: influence of time spent studying and working. J Educ Bus 81(3):151–159
8. Shaukat K, Nawaz I, Aslam S, Zaheer S, Shaukat U (2016) Student’s performance in the context of data mining. In: 2016 19th international multi- topic conference (INMIC). IEEE, pp 1–8
9. Fok WW, Chen H, Yi J, Li S, Young HA, Ying W, Fang L (2014) Data mining application of decision trees for student profiling at the Open University of China. In 2014 IEEE 13th international conference on trust, security and privacy in computing and communications. IEEE, pp 732–738
10. Hasan R, Palaniappan S, Mahmood S, Abbas A, Sarker KU, Sattar MU (2020) Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Appl Sci 10(11)
11. Prajapati AP, Sharma SK, Sharma MK (2017) Student’s performance analysis using machine learning tools. Int J Sci Eng Res 8(10):487–491
12. Jain, Solanki S (2019) An efficient approach for multiclass student performance prediction based upon machine learning. In: International conference on communication and electronics systems (ICCES). Coimbatore, India, pp 1457–1462

13. Err E, Roy C, Pandey M, Rautaray SS (2018) A proposal for optimization of horizontal scaling in big data environment. In: Advances in data and information sciences. Springer, Singapore, pp 223–230
14. Ramesh VAMANAN, Parkavi P, Ramar K (2013) Predicting student performance: a statistical and data mining approach. *Int J Comput Appl* 63:975–8887. <https://doi.org/10.5120/10489-5242>
15. Kotsiantis S, Pierrakeas C, Pintelas P (2003) Preventing student dropout in distance learning systems using machine learning techniques. In: AI techniques in web-based educational systems at seventh international conference on knowledge-based intelligent information and engineering systems, pp 3–5
16. Mayilvaganan M, Kalpanadevi D (2014) Comparison of classification techniques for predicting the performance of student's academic environment. In: 2014 international conference on communication and network technologies (ICCNT). IEEE, pp 113–118
17. Gerritsen L, Conijn R (2017) Predicting student performance with neural networks. dissertation, Dept. Humanities, Tilburg University, The Netherlands
18. Vamshidharreddy VS (2020) Student's academic performance prediction using machine learning approach. *IJAST* 29(9s):6731–6737
19. Chemers MM, Hu LT, Garcia BF (2001) Academic self-efficacy and first year college student performance and adjustment. *J Educ Psychol* 93(1):55
20. Arsal PM, Buniyamin N, Manan JLA (2013) A neural network students' performance prediction model (NNSPPM). In: 2013 IEEE international conference smart instrumentation, measurement and application, ICSIMA. no July 2006, pp 26–27
21. Bharadwaj BK, Pal S (2011) Data mining: a prediction for performance improvement using classification. *Int J Comput Sci Inf Secur (IJCSIS)* 9(4):136–140

Low-Power IoT Architecture, Challenges, and Future Aspects



Saurabh Sambhav and Shilpi Singh

Abstract This study looks into the aspects of Internet of Things (IoT) architecture, low-power IoT architecture, with their functionality and challenges with opportunities in future aspects for implementation of LP IoT. Low-power wide area network (LPWAN) technology is essential for IoT applications deployment. The Internet of Things is an emerging paradigm that paves the way for smart and sustainable applications, environments, infrastructures, and services. IoT is happening at a rapid pace, and it has already brought in changes to some basic human needs such as healthcare, supply chain management, sustainable energy management, and logistics. Low-power IoT allows any object with an on-board microchip to connect directly to the Web. At the same time, the architectures should also take care of number of challenges in terms of range scalability, limitation, security, computing, battery lifetime, etc. In this paper, low-power Internet of Things (IoT) architecture along with the opportunities, challenges, and future aspects has been identified.

Keywords Internet of Things (IoT) · Low-power IoT architecture · Networking · LP IoT challenges · Mobile computing

1 Introduction

Internet of Things (IoT) is an advancement in technology that aims to connect all the physical objects into Internet through manageable, smart environments. IoT can be considered as the next phase of Internet evolution after the Internet era has started since 1980s. IoT usually inferred as a technology domain which deals with connecting everyday non-computer devices like lights, washing machines, human body, etc., to the Internet by embedding them with sensing device and means for wireless communication. The benefit of this approach is that it will facilitate easy management and controlling of these devices using PCs and mobile phones even when users are away from their homes or offices. This Internet of Things trend has become popular

S. Sambhav (✉) · S. Singh

Amity School of Engineering and Technology, Amity University Campus, Patna 801503, India

recently after being introduced by Kevin Ashton at the TED conference in 2006 [1]. Internet-enabled devices already exist in our everyday lives, for instance traffic sensors, industrial equipment, and home appliances. Internet-connected objects are creating new opportunities for business because of the availability and accessibility of data generated from these smart objects. It is expected that IoT market would grow to about \$7 is billion within next 10 years which shows its great potential [2]. There are several ways to connect an object into a network such as using wired or wireless technologies. Wired approach implies connecting things either directly or indirectly through controllers or hubs having wired Internet access. Wireless technologies on the other hand are more suitable in scenarios where wired connectivity is not possible or too costly [3]. The Internet of Things encompasses several fields such as sensor networks, mobile ad hoc networks, wireless sensor networks, and wireless mesh networks. These research areas complement each other and can be combined to create IoT systems having various applications, for example, security system with motion detection using video streaming from cameras at home connected through Internet to a cloud storage service accessible via smartphone [4]. Internet of Things (IoT) is emerging as an infrastructure that enables connection among billions of physical objects embedded using any number of wireless technologies such as Wi-Fi, Bluetooth low energy (BLE), ZigBee, or LoRa across both time varying or static environments for communication among device and cloud services. Internet of Things (IoT) is evolving as a network of interconnected devices that exchange data over the Internet to do useful work. IoT consists of physical objects, embedded with electronics, software, sensors, and actuators connected through Internet or local networks. The Internet of Things (IoT) architecture needs to be designed according to certain key requirements such as low-power consumption Internet connectivity, secure authentication and authorization, reliable networking between end-users and Internet users, communication across multiple wireless technologies, and orchestration services for complex operations among heterogeneous devices spread over large geographic areas spanning both time varying or static environments to keep effective control on these deployed systems. These architectural issues pose many challenges in incorporating new features required by emerging applications. Also, Internet connectivity is required at extremely low energy consumption because Internet of Things (IoT) is expected to support large number of distributed devices with a power budget far below the level currently being implemented]. (i) Low-power Internet connectivity requirements for IoT. Energy consumption has always been one out of major interest points in designing various Internet technologies such as Internet protocols, in-network processing techniques, and communication algorithms. With this background, IoT poses several new challenges that need to be addressed by new design choices. At very first over last few years, several papers appeared describing low-power Internet connectivity architectures [6–8]. These works focus on general Internet of Things application scenarios. As a result, the Internet connectivity requirements are defined as “low-power connectivity” or LPC. On one hand, this seems acceptable since Internet routing does not need certain features commonly described as mobility (e.g., location update, handoff, handover) [9]. On the other hand, Internet routing does need some mobility features (e.g., IP address prefix and structure changes).

Moreover, Internet of Things application scenarios cover a wide range of applications which may require different sets of Internet protocol stack functionality. Internet connectivity requirements are related to the Internet protocols used by IoT device(s) in general [10]; therefore, they must consider all Internet protocol components and their interactions. This article considers low-power Internet connectivity requirements for Internet of Things devices. It also proposes possible solutions to be implemented in low-power IoT application specific integrated circuits (ASICs). The paper is organized as follows: Section 2 reviews main characteristics of IoT, LPIoT, and LPWANs. Section 3 provides an overview of different types of LPWANs. Section 4 explores the challenges, and Sect. 5 presents future scope and possible IoT solutions.

2 Architectural Overview

2.1 *Internet of Things (IoT) Architecture*

Internet of Things (IoT) is a term used to define networked devices embedded with electronics, software, and sensors and connect these physical objects with Internet enabling them to communicate with each other. According to an estimate by market research firm due to its widespread use in various applications like manufacturing, transportation, and healthcare, every second thing will be connected with Internet. Internet of Things (IoT) has various definitions, and it is categorized into two subsets, i.e., fog computing which refers to IoT systems having small computational power but large communication ability and cloud computing which includes higher computation power but smaller communication ability. It can interface any device or machine that has an Internet connection to the Internet. Internet of Things (IoT) is a secure network of devices connected to each other via various communication protocols. Internet of Things enables communication between objects which may be very small and almost invisible or even non-existent. Low-power Internet of Things (LPWAN) architecture involves low-power connectivity among Internet-connected devices using low-power wide area networks. This architecture consists of LPWAN gateways which can comprise all kind of Internet technologies such as Internet of Things (IoT) service platforms like IBM Watson IoT platform, Rtiwrf for developing LPWAN gateways (Fig. 1).

2.2 *Low-Power IoT Architectures*

The low-power Internet of Things Internet (LPWAN) architecture consists of a network layer on top of which apps can be deployed [5]. LPWANs are closely related to other types of cellular wide area networks such as cellular broadcast networks like SigFox or enhanced general packet radio service (EGPRS)/enhanced data rates for

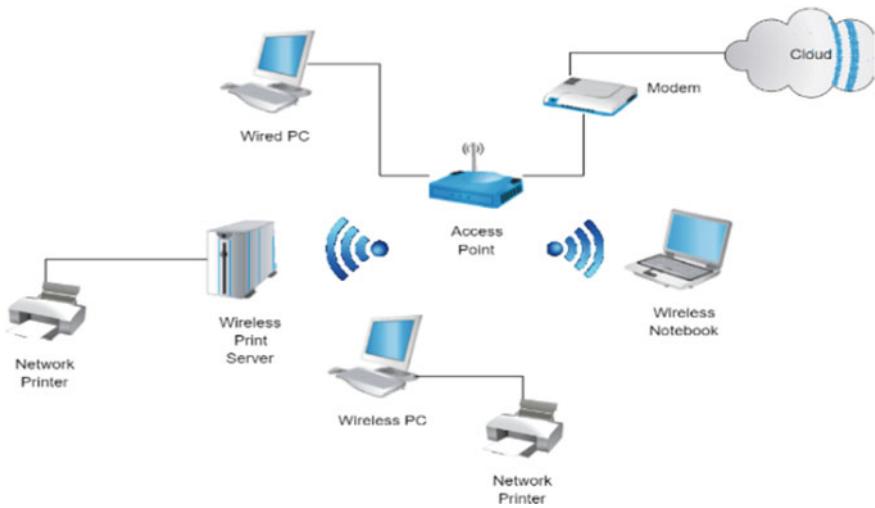


Fig. 1 IoT network architecture inside any small setup like office, school, etc.

GSM evolution (EDGE) networks. Low-power Internet of Things (IoT) architecture has a network based on Internet protocols, and energy harvesting or low-energy wireless technologies such as ZigBee, low frequency narrowband (NFB). The main reason that makes LPWAN more preferable at this stage than cellular Internet in IoT applications is the lower cost since no network operator would be involved and hence no cost associated with Internet connectivity [8]. The other main benefit of LPWAN is its low-power consumption, which makes it ideal for battery-powered systems. The low-power Internet of Things (LPIoT) architecture provides a new networking paradigm for Internet of Things (IoT), whereby devices can become connected to the Internet without an intermediate base station or gateway infrastructure.

2.3 Low-Power IoT Wide Area Network (LPWAN)

The devices will be able to communicate by using a low-power and long range wireless connection with the Internet. The Internet gateway includes hardware and software components that connect LPWAN devices to the Internet. The connection between two LPWAN devices is similar to traditional Wi-Fi networks, but it uses low data rate protocols like narrowband-IoT (NB-IoT), IEEE 802.15.4 or GSM/GPRS. In case of battery-operated devices, the network lifetime for Internet connection is limited and depends on the power source, e.g., batteries or solar-powered devices. The Inter Internet of Things (IoT) gateways will decide what data packets are forwarded to the Internet where they can be further processed by higher level services like Big Data analytics or Internet of Things (IoT) application platforms. The Internet of Things (IoT) is a new era in which physical objects are able to communicate with Internet

users and share sensor data. These devices are not supposed to be directly connected to the Internet but rather should act as gateways for smaller IoT devices. From our perspective, this gateway function can best be implemented using low-power wide area networks (LPWANs).

3 Different Types of LPWANs

3.1 NB-IoT (*Narrowband-IoT*)

The Internet of Things devices connected to an NB-IoT network are generally not battery operated. They can be powered by the energy grid, building power connections, or solar panels. The main advantage that NB-IoT offers is lower cost and higher data rates than LoRa or Sigfox networks. As Internet of Things applications grow in demand and IoT will cover a larger part of our daily lives, Internet access will also need to become more ubiquitous that is why several LPWAN standards have been created that enable Internet connectivity for low-power devices all over the world with only very small costs. The Internet of Things requires reliable Internet transport at large scale which translates into a high number of Internet-connected devices.

3.2 EC-GSM IoT (*Extended Coverage—GSM—IoT*)

It is an Internet of Things (IoT) technology which takes advantage of existing GSM networks as backhaul. The Internet traffic for connected devices is transported by M2M Mediators and relayed to the Internet routers through the cellular network infrastructure. A low-power IoT device such as a sensor or actuator may obtain its connectivity from a nearby Internet router through an intermediary M2M Mediator device that in turn connects to a remote Internet router over a GSM/GPRS network. Efficient use of cellular bandwidths with minimal cost makes EC-GSM-IoT a viable option for Internet access wherever cellular networks are available. It should be noted that EC-GSM-IoT can be implemented in a variety of low-power Internet connectivity solutions.

3.3 Long Range (*LoRa*)

LoRa is a wireless networking technology developed by Internet of Things (IoT) and Internet of Everything (IoE) experts Semtech. LoRa is the leading LPWAN, with over 1000 deployments around the world to date. It was designed specifically for low-power communications and uses a spread-spectrum modulation method, which

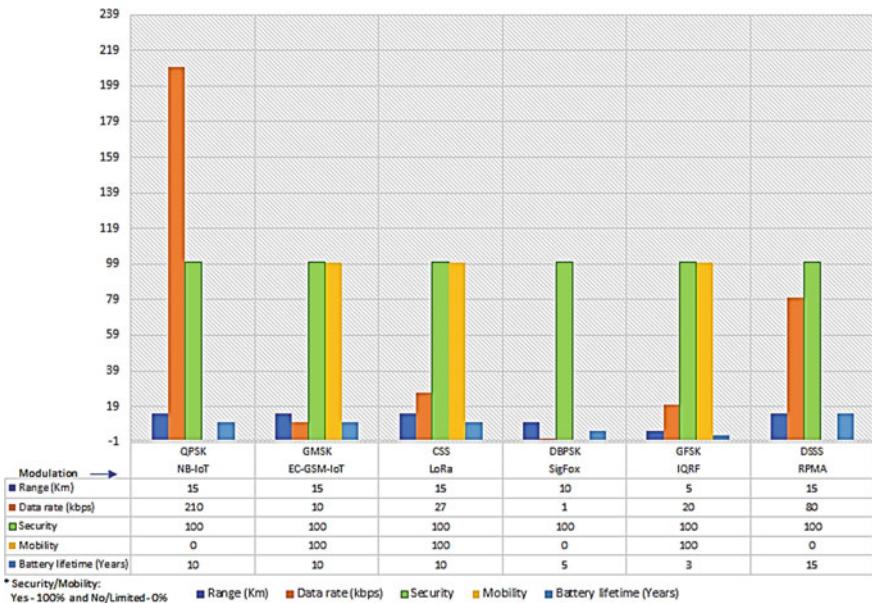


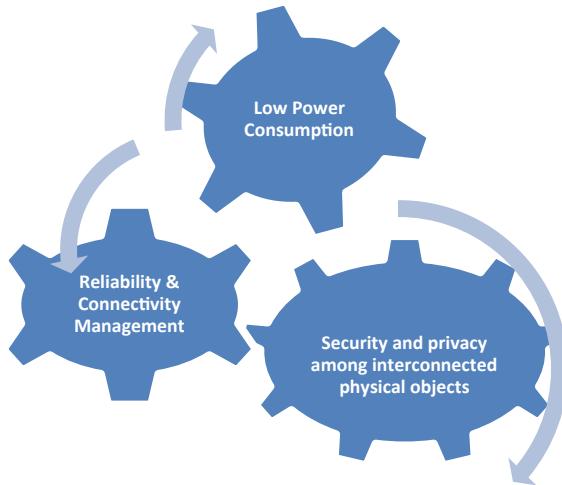
Fig. 2 Comparison of different types of LPWANs

helps remove interference from other networks. LoRa offers high security through its integrated encryption capability that lets devices communicate securely with each other without Internet access or backhauls. LPWAN technologies are expected to be used in various applications such as smart cities sensors network, Internet of vehicles, traffic management system, healthcare monitoring solutions, and various remote Internet of Things applications. While there are several low-power wide area network technologies available in the market, LoRa is one of the most popular ones due to its high level of security and relatively long range (1–10 km) (Fig. 2).

4 Challenges in Low-Power Internet of Things Architecture

The challenges in the Internet of Things architecture include power management, Internet of Things sensors configuration, Internet connectivity, Internet cloud, Internet security, Internet applications, IoT devices data analytics, Internet of Things (IoT) architecture, Internet of Things cloud management, Internet-embedded devices Web server, and Internet Web applications. The wireless sensor networks have been identified as a fundamental technology for enabling the low-power IoT architecture. Interconnected physical objects, people, and software via Internet protocol (IP):

Fig. 3 Highlighting the challenges in low-power Internet of Things wide area networks



Challenges in architecture Internet of Things (IoT) include interconnected physical objects, people, and software via Internet protocol (IP). There are some other challenges: Following are some challenges IoT needs to address: (a) low-power consumption, (b) quality or reliability issues due to Internet connectivity, and (c) security and privacy issues (Fig. 3).

5 Future Aspect and Opportunities

In the meantime, while you are integrating your chips with the IoT cloud platform, other companies can be developing their own devices based on your chip/my hardware design that has been successfully integrated with my software/your software solution in a single development process according to my design. Revenue model: you earn from the software installs on my hardware, and I earn from other companies who have your chips in their devices that they sell to consumers or businesses, or governments. APIs for Mobile Apps Developers? The Internet of Things cloud platforms like IBM's Bluemix Internet of Things platform offer Internet of Things service APIs for its IoT application developers, but are there Internet of Things connectivity APIs for mobile (Android and iOS) app developers? Innovations in the Internet of Things Cloud Architecture? NoSQL Database MongoDB can be used in the Internet of Things cloud platform to enable application development via Node-RED or other programming languages like Apache Groovy, JavaScript, and Python for building scalable and resilient low-power IoT architectures.

6 Conclusion

This paper focuses on the study of Internet of Things architectures low-power Internet of Things and wide area networks. The aspects of LP IoT and implementation have been covered in the paper. The different types of LPWANs have been discussed in detail, and the key differences have been identified. The challenges and future opportunities have been discussed in the latter part of the paper. There is many scope, but at the same time, we should also take care of many encounters in terms of range scalability, limitation, security, computing, battery lifetime, etc. In this paper, low-power Internet of Things (IoT) architecture along with the opportunities, challenges, and future aspects have been identified.

References

1. Ferran A, Vilajosana X, Tuset-Peiro P, Martinez B, Melia-Segui J, Watteyne T (2017) Understanding the limits of LoRaWAN. *IEEE Commun Mag* (Jan 2017)
2. Augustin A, Yi J, Clausen T, Townsley WM (2016) A study of LoRa: long range and amp; low power networks for the internet of things. *Sensors* 16:9
3. Bardyn JP, Melly T, Seller O, Sornin N (2016) IoT: the era of LPWAN is starting now. In: ESSCIRC conference 16: 42nd European solid-state circuits conference
4. Bor MC, Roedig U, Voigt T, Alonso JM (2016) Do LoRa low-power wide-area networks scale? In: Proceedings of the 19th ACM international conferences on modeling, analysis and simulation of wireless and mobile System, pp 59–67
5. Cattani M, Boano CA, Romer K (2017) An experimental evaluation of the reliability of LoRa long-range low-power wireless communication. *J Sens Actuator Netw* 6:2
6. Xu LD, He W, Li S (2014) Internet of things in industries: a survey. *IEEE Trans Ind Inf* 10(4):2233–2243
7. De Poorter E, Hoebeke J, Strobbe M, Moerman I, Latré S, Weyn M, Lannoo B, Famaey J (2017) Sub-GHz LPWAN network coexistence, management and virtualization: an overview and open research challenges. *Wireless Pers Commun* 95(1):187–213
8. De Poorter E, Hoebeke J, Strobbe M, Moerman I, Latré S, Weyn M, Lannoo B, Famaey J (2017) Sub-GHz LPWAN network coexistence, management and virtualization: an overview and open research challenges. *Wirel Pers Commun* 95(1):187–213, (July 2017)
9. Ekiz N, Salih T, Kucukoner S, Fidanboylu K (2007) An overview of handoff techniques in cellular networks. *Int J Inf Technol* 2 (2007)
10. Floreano D, Wood RJ (2015) Science, technology and the future of small autonomous drones. *521:460–466*

ANURL: An Innovative Management Scheme for Web Uniform Resource Locators



Ashish Karn, Suyash Thakur, and Pankaj Badoni

Abstract The developments in the Internet, social media platforms, and cloud technologies have necessitated the need for sharing online contents of various kinds. Despite a slew of popular shorteners for uniform resource locators (URL) that are available such as Bitly and TinyURL, we delineate the need for a more comprehensive URL management scheme. The current paper reports the development of an innovative and comprehensive management scheme for Web uniform resource locators (URL) named ANURL. The paper presents a detailed account of its features and its several use cases, which makes it stand out over the conventional ones. A brief description of the ANURL architecture and its working is provided. Finally, some preliminary data on the tool evaluation and user testing is presented to substantiate its growing user base with geography and time.

Keywords ANURL · Uniform resource locators · URL · Short URL · Bitly · TinyURL

1 Introduction

In recent times, the developments in Internet and cloud technologies and the burgeoning growth of social media platforms have facilitated an easy sharing of online content such as images, videos, e-books, newsfeed, product specifications, shopping lists, commercial websites, educational quizzes and assignments, and personal

A. Karn

MultiPhase Flows Laboratory, Department of Mechanical Engineering, School of Engineering, University of Petroleum and Energy Studies, Energy Acres, Bidholi, Dehradun, Uttarakhand, India
e-mail: akarn@ddn.upes.ac.in

S. Thakur · P. Badoni (✉)

School of Computer Science, University of Petroleum and Energy Studies, Energy Acres, Bidholi, Dehradun, Uttarakhand, India
e-mail: pbadoni@ddn.upes.ac.in

S. Thakur

e-mail: 500067151@stu.upes.ac.in

and professional profiles of individuals. Usually, such online content is shared via social networks, news media, blog posts, online communities and postings, instant messaging, and other Web services by individuals and organizations [1]. In all such interactions, a *Uniform Resource Locator*, or URL, is of paramount importance, and it signifies a specific location of a resource on the Web. Put simply, a URL is equivalent to a Web address, and it has two important components: the first part typically being a *protocol identifier* that signifies the format of data transmission being used, and the second one is the *resource name* that is specific to the IP address of the resource. Further, the resource name may consist of a domain name and file path. For instance, a typical Web link that hosts some digital storytelling resources on the webpage of an innovative educator in thermal-fluid sciences looks like: <https://www.drkarnteaching.com/digital-storytelling>. In this particular case, “https:” is the protocol, and the rest of the portion is the resource name with “drkarnteaching.com” being the domain name and “digital-storytelling” being the file path. Similarly, Internet is replete with a variety of content and Web pages, each of which needs its own unique URL. In order to generate such URLs in a manageable manner, some organization and structured approach is required. Usually, this is achieved by applying slashes, dates, keywords, names of individuals or random strings, etc. [2]. However, this also poses a constraint. With the introduction of these additional strings and characters, the URL becomes longer, difficult to memorize, reproduce while typing, or distribute to others through non-digital modes, and copying and pasting it is probably the only reliable means to work with it.

To overcome these constraints, URL shortening services have been used widely for sharing and publishing online content by providing a short equivalent URL that is redirected to the corresponding long URL by the service provider through an “HTTP 301 Moved Permanently” response [3]. Initially in 2001, when the URL shorteners first surfaced, the underlying concept behind them was to prevent the breaking of long URLs while copying text and to prevent email clients from rendering the URLs unclickable by the insertion of line breaks between them [1]. Since then, its acceptance has been slow before these became prevalent in online social networks. Now, with the proliferation of social media platforms as well as its accessibility through mobile devices, URL shorteners have almost become a requirement, partly owing to character limitations in some social media such as Twitter. There are many available URL shortening services available for public use, tinyurl.com and bit.ly, being the most popular. However, there are many open-source packages and web modules that can be alternatively used in a web application within a domain, such as YOURLS, Polr, and Shlink. In addition, some other packages are available to Python-based Web application such as djanurl and microurl [1].

However, many of these common services are limited in the features that they provide to individuals without a premium license. For instance, tinyurl.com provides individuals to create short URL, but does not provide an option to edit the landing URL, or an option for a particular user to verify the entire list of URLs he/she has created. Similarly, many URL shorteners do not provide analytics information for all the short URLs created. To the authors’ best knowledge, there is no URL shortener that provides an option for the expiry of a short URL at a chosen date/timestamp. This

may be useful particularly if one needs to create a short URL for sharing over a short duration of time. But, in the current scenario, a keyword selected by a user cannot be later used by any other user, and this may limit the choice of keywords for many. Similarly, most of these URL shorteners do not permit the keyword to be a collection of strings and slashes which can be used to present a feel of the file directory path on a particular website and facilitate easy recollection. Another important concern stems from security issues. Short URLs have often been used recently by spammers and for malicious content phishing and malware attacks. Since hiding their original content URLs, they are often used for sharing malicious content such as spam or phishing [4–6]. In view of these points enumerated above, there is a clear need for the design and development of a novel URL shortener scheme that can tackle these challenges with ease and provide a simpler, more secure, and free alternative, not just for businesses but also for individuals. The current paper thus reports the development of an indigenous and innovative URL shortening scheme in the Indian academic setting. The new scheme for uniform resource locators, named ANURL has been designed, developed, and later deployed.

2 ANURL: A Description of the Innovation

Semantically speaking, ANURL is a compound word formed by the collocation of the Sanskrit root “*anu*” meaning small, and URL, which is an acronym for uniform resource locators. So, ANURL does refer to a short uniform resource locator on the Web, and can be accessed over the Web at <https://www.anurl.in/>. Apart from the instructions for use and contact information, ANURL sidebar presents choices for login and signup. Further, the login presents choices for login with Google account or a “Guest Login” that does not require any information to be fed. However, considering the security of the created short links, the validity of the short links created using the guest login is limited to 48 h. The homepage lists different features of ANURL, which can also be described as follows.

After the login, the user lands up at a place where three pieces of information have to be entered: the long URL, a date/timestamp, and a short string or a keyword which is referred to as “hash” by Bitly or more generically, backhalf. The date/timestamp is typically an instant of time attaining, which the ANURL assignment being created by the user deems to be invalid. This may be useful in cases where a user may want the short link to be active only for the time duration that he/she wants it to be and wishes to make it defunct later. A possible reason behind this may be the time-specific nature of the link being shortened, or simply to ensure that keywords or strings are freed up for others to choose from. This ensures that unlike Bitly, the same backhalf could be assigned to different individuals at different instances in time, if it is available later. However, this choice is optional only. If a user does not provide an expiration timestamp, his short link will always remain active and will continue to direct him/her to the chosen URL and no one else can be granted the same backhalf. Unlike tinyurl, ANURL does not generate a randomized backhalf. However, ANURL does provide

The screenshot shows the ANURL interface. At the top, there's a header 'Create Link' and a form with fields: 'www.anurl.in/' (Short URL), 'Original URL', 'Expiration date/timestamp', and a red 'Create' button. Below this is a section titled 'Links' with a search bar and a 'Sort by' dropdown. A table lists three created links:

Short Link	URL	Clicks	Date Created	Expiration date/timestamp	Action
www.anurl.in/screw	https://www.dropbox.com/sh/gq7ixdembv1269h4/AADoQ91kU4AR5b3nREcVEsGwa?dl=0	2	28-07-2021	NA	Edit Delete
www.anurl.in/ddnews	https://www.linkedin.com/posts/mplf-upes-771a57211_toycat-hon2021-health-upes-activity-68140829126866809089-Ux5H	1	29-06-2021	NA	Action
www.anurl.in/pm	https://www.youtube.com/watch?v=ubwohKmkqSs	2	29-06-2021	NA	Action

Fig. 1 A Snapshot of the ANURL interface upon logging in, showing some sample created ANURLs, along with the details

a very unique feature of editing the short links and its connection to the parent Web link. So, a user can redirect the same short link to different locations at a later date. Conversely, a user can change the backhalf to refer to the same long URL. A user also has the option to delete his short URL and its assignment. The user can also track the number of clicks on his short link, and can copy the link or directly click it as well, to check its functionality. Further, as Fig. 1 shows, one can also search a particular backhalf in the list of assignments one may have made in the past. This facilitates easy management of all short links in one place: search, check, edit, relink, or disable.

3 ANURL Architecture and Working

The entire ANURL coding is done in JavaScript, and the technology stack used to create the application is described as follows. For the server side programming, nodeJS was employed with Express framework. As far as database is concerned, mongoDB, which is a nosql database was utilized. For frontend, Ejs, a templating language was used. The frontend was server side rendered and sent to the client side as payload to the API requests. The technical stack of ANURL is shown in Fig. 2 and can be understood as follows. A user interacts with the front end, which then sends a HTTP request to the Express server with all the necessary payload, followed by the

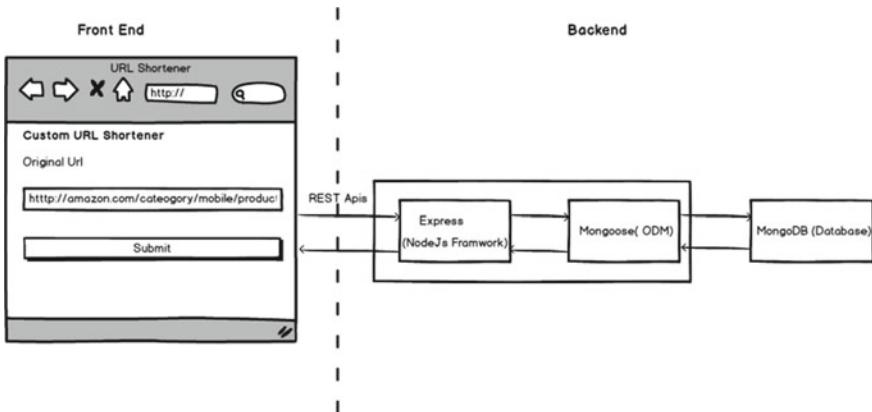


Fig. 2 Technical stack of ANURL, clearly showing the front end and the back end

Express server making a query to the database server via Mongoose. Database holds all the data for the application.

Next, we diagrammatically present the four processes of user signup, user login, creation of a short URL, and the process of navigating original URL. In the user signup process, a user enters email and password, and the password is hashed at the server with a secret value, and finally, the hashed value along with other user details is stored in the database. The database returns a user JSON object to the server, which then shows up a success message for the user.

As shown in Fig. 3, while authentication (i.e., login), the user enters email and password, and server checks if the user with a given email exists in the database. Then, the server hashes the password entered and checks if the hash value matches with the password in the database. The database communicates to the server with a user JSON object, and upon successful login, an authentication token is generated.

Figure 4 illustrates the procedure of the creation of a short URL. Typically, a user enters all the required fields to create a short URL, whereafter the front end sends the HTTP request to the server with the required information as payload in the request body along with the authentication token. Then, the server middleware verifies the token, and the server creates a new entry in the URL collection of the database which contains reference to the user collection.

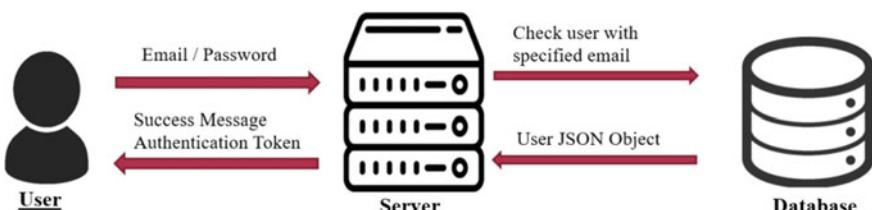


Fig. 3 The user login procedure

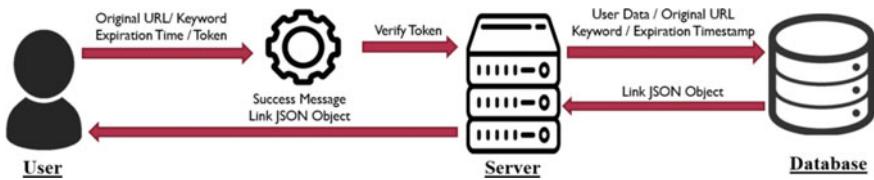


Fig. 4 The process of creating a short URL

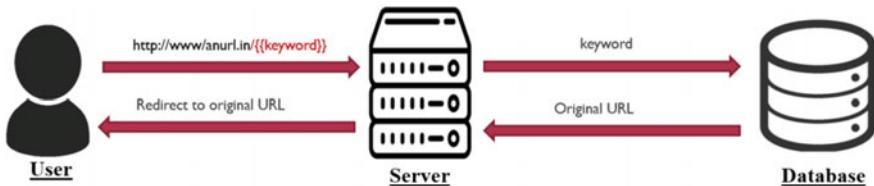


Fig. 5 The process of navigating a short URL

Next, Fig. 5 demonstrates the set of steps involved in navigating the original URL. When opening a short URL, the keyword is extracted from the URL as params. In the case of a directory-like route (for instance, “anurl.in/keyword1/keyword2”) all the keywords are extracted and concatenated to create a single searchable string. Then, the server sends a query to the database to extract the requested URL mapped to the given keyword and redirects to the original URL.

Figure 6 presents a schema of the database in the current study. Although nosql database does not have a strict schema, a visual representation helps to understand the type of data stored in the collection. There are broadly two collections: user and link. Every unique user has an entry in the user collection and every link created has an entry in the link collection. To map link collection to user collection, there exists a field in the link collection called author which is of type ObjectId and has reference to user collection. The author field contains the Id of the user that has created the link (similar to a foreign key in SQL database). The TTL field in the link collection is used to track the expiration time of the links.

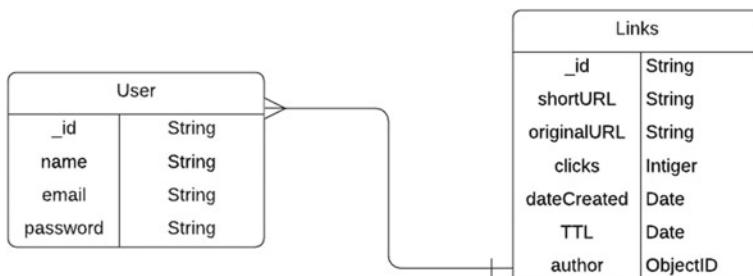


Fig. 6 The schema of the database

4 ANURL Use Cases

ANURL provides individuals with a way to shorten links to get a manageable and readable URL to share content [3, 5, 6]. These links are not just personalized but easy to recollect and distribute to others at any opportune moment. An obvious use-case is that of microbloggers, who use short URLs to share content in their microblogs that have a specified character length limit, say, a limit of 140 characters as in Twitter [3]. In addition, users who want to get rid of semantics from an original URL may use ANURL for content sharing. In general, users can choose to create a customized backhalf that are easy to recollect, as long as the keyword (or backhalf) is not already taken. For instance, “anurl.in/rajcv” is a customized backhalf for a resume of a person Raj. An instructor, when asked for his calendar may direct his students to “anurl.in/calendar,” or when asked orally for a meeting link, may simply text “anurl.in/zoom.”

There is another interesting feature as far as customization of short URLs is concerned, what we call as “URL directories,” that may make dealing with the URLs much simpler, particularly for individuals and small organizations owing to the absence of ready availability of such tools. Apart from characters a-z, A-Z, and digits 0–9, ANURL allows insertion of “/” symbols in the short links, allowing to form strings that appear as directories and filepath. For instance, an individual by the name of “Hari” can store his pictures in this way—“anurl.in/hari/codes/year1” (alluding to Hari’s codes from year 1), “anurl.in/hari/research/year2” (directs to a cloud folder that houses all research done by him in the second year) etc. This can provide great ease of tracking in organizations such as the list of research publications in different departments of a university: “anurl.in/mech/publications” (Publications of Mechanical Engineering department), “anurl.in/physics/publications” (a reference to publications of physics department). Or, for that matter, in storing the repositories of a conference such as “anurl.in/icmete/2021/day1/ppts” may be used to store all the presentations from day-1 of the conference ICMETE-2021. Similarly, “anurl.in/icmete/2020/day3/papers” may be used to refer to all the papers related to the research presented at third day of ICMETE-2020.

The URL editing feature of ANURL can serve a number of useful applications in scenarios when the same link can be used for a variety of chronological operations. For instance, in the submission, online review, and result declaration for a proposed research call, the same link “anurl.in/project21” may point one to the webpage for submission of the proposal before the submission deadline, to a link for a Zoom meeting for the online evaluation, and later to the document showing the final results. This feature also ensures that, if required, the links can be shared with others even before the final link of content on the Web is known (for instance, a YouTube video). This can prove to be very useful in situations of strict deadlines where content has to be submitted. In such scenarios when the evaluation of the submitted content is clearly a later event, ANURL provides a sound buffer time. Similarly, ANURL can provide a unique link for all meetings of a department, even if the meeting can be

scheduled by different people at different times, provided the back-end correction is made.

The analytics feature of ANURL can help one to check the number of hits on a short URL and may reveal the accessibility and extent of use of particular online content. This can be profitably used by an instructor, for instance, in gauging the student's response to a particular theme or aspect of the course. Similarly, the ANURL allows one to set an expiration date/time stamp for every URL one creates and deactivate a URL when required. This may again prove very useful in educational settings, for instance, to auto-set deadlines for a task. By setting up an expiration timestamp, an instructor can automatically ensure the closure of the submission of assignments since the short links may not remain functional later.

However, an objection may be raised regarding the security concerns in the usage of short links such as ANURL, since short URLs are convenient tools in the hands of spammers and attackers who hide original URLs in this way [5]. However, multiple steps have been taken to ensure the security of ANURL short links. First, the Google sign-in makes sure that user authentication is done and minimizes the chances of misuse to some extent. However, in order to minimize the misuse by guest login, the validity of the short URLs is limited to a maximum of 48 h. In addition, a rate limit has been applied to the API to protect it from spam, with the current limit being 100 requests per five minutes. This ensures that if any short link receives more than a hundred hits in five minutes, the short URL is automatically blocked at that particular instant. These features greatly curtail the scope of misuse of ANURL by spammers.

5 ANURL Deployment and Testing: Survey and Statistical Analysis

Finally, we ran behavioral testing to ensure excellent user experience, seek feedback on the novelty, working, and usefulness of the tool, and also collect inputs that could be incorporated into it. The developed tool underwent rigorous testing and evaluation by computer science students and experts through a Google Form, and sixty responses were received and analyzed. The survey queries consisted of both qualitative and quantitative responses, and the received responses are then examined through a MATLAB script. The response of the users toward three aspects of ANURL: its conceptual clarity on the website, the usefulness of the ANURL scheme, and its ease of use were collected. Figure 7a shows the user responses toward these in stacked bar format. In all three cases, about 40–60% of users thought that ANURL was a “great” tool, whereas approximately 30–50% of users in all the cases claimed it to be a “good” tool. Overall, around 90% of the users attest to the utility of ANURL with the top two responses. Upon digging deeper as to what fascinated the users most about ANURL (as per their understanding), the users had varying opinions. Figure 7b illustrates that 63% of the users liked the “custom short URL” feature of

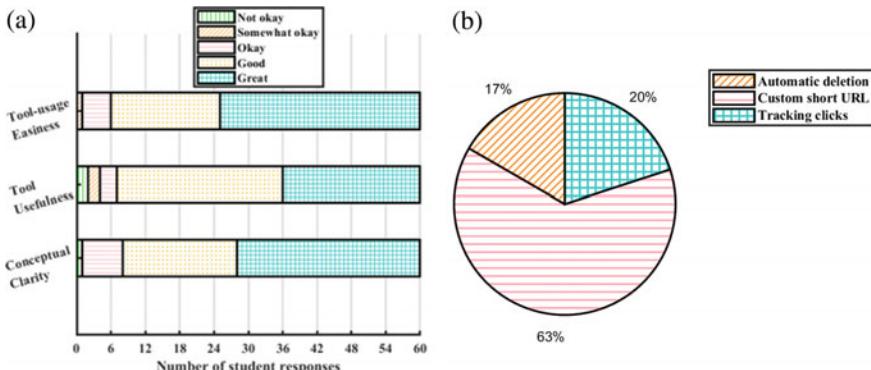


Fig. 7 **a** General acceptance and comments on ANURL and **b** The best feature of ANURL as per the survey results

the tool most (with its edit feature), 17% vouched for automatic deletion using a chosen timestamp and the remaining 20% approved the analytics feature.

Apart from the quantitative feedback, users also provided much qualitative feedback as well. Most of the users responded in affirmative lauding this new initiative, with pithy comments such as “Good idea,” “Great work,” “Really useful website for sure,” “I found this project really cool,” “I’ve seen these features for the first time,” “Amazing UI, smooth and functional! Loved it,” and “Excellent functionality.” Needless to say, since the entire sample set for the data collection consisted of computer science fraternity, they were quite familiar with the conventional URL shorteners. Some users provided constructive feedback of integrating the feature of guest login, a page outlining the instructions for users etc. Users also requested to provide a direct link to redirect instantly, a feature to copy links, enhanced security features, etc., and these features were eventually integrated into the Web tool. Users also commended the logo and the simple user interface of the tool.

Finally, we present some data on the usage of ANURL with respect to geography and time. The data was collected during two months labeled as month-1 (18 April 2021–18 May 2021) and month-2 (18 June 2021–18 July 2021). Figure 8a shows that during these two periods, India, the USA, and Germany constituted the most of the users followed by the Russian federation, Belgium, the UK, and Norway. In addition, during month-3 (18 July 2021–18 August 2021), 184 users from China and 242 users from Hong Kong also registered for the tool. Further, Fig. 8b shows that during the two periods, total requests and the number of unique visitors remain largely the same. It can then be safely concluded that the number of visitors from the USA has increased steadily from month-1 to month-2. This assertion can further be validated from Fig. 9, which shows the expanding user base of ANURL over all the three month periods in 2021, for three leading countries: India, the USA and Germany.

It is expected that owing to its unique features, the user base of ANURL will expand in the future, particularly for individual segments. Clearly, this humble Web

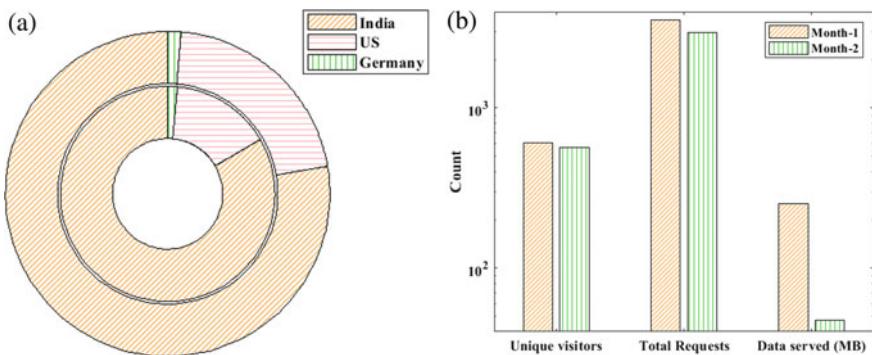
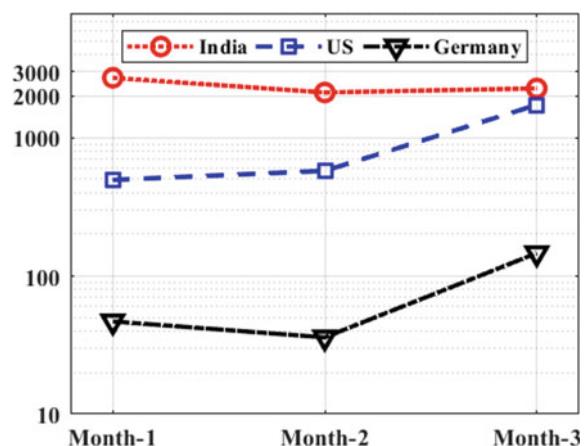


Fig. 8 **a** Geographical (the inner bar denotes Month-1) and **b** Variation in the usage of ANURL

Fig. 9 The increase in user base of ANURL over three months in 2021



tool emerges from an academic setting and does not target to contend with the conventional tools that are being marketed currently. Yet, the current development should be seen only as an innovative tool that was targeted to simplify things for individuals, particularly educators, who are encumbered by the conundrum of Web links in today's digital age.

References

1. Prasetyadi G, Hantoro UT, Mutiara AB (2018) Singkat: a keyword-based URL shortener and click tracker package for Django web application. *Int J Adv Comput Sci Appl* 9(9):118–122
2. Azar E, Alebicto ME (2016) Swift data structure and algorithms. Packt Publishing, Birmingham, p 236
3. Antoniades D, Polakis I, Athanasopoulos E, Ioannidis S, Markatos EP, Karagiannis T (2011) Web: The web of short urls. In: Proceedings of the 20th international World Wide Web conference (WWW 2011)

4. Chhabra S, Aggarwal A, Benevenuto F, Kumaraguru P (2011) Phi.sh/ocial: The phishing landscape through short urls. In: Proceedings of the 8th annual collaboration, electronic messaging, anti-abuse and spam conference
5. Gupta N, Aggarwal A, Kumaraguru P (2014) bit.ly/malicious: Deep dive into short URL based e-crime detection. CoRR abs/1406.3687
6. Nikiforakis N, Maggi F, Stringhini G, Rafique MZ, Joosen W, Kruegel C, Piessens F, Vigna G, Zanero S (2014) Stranger danger: exploring the ecosystem of ad-based url shortening services. In: Proceedings of the 23rd international conference on World Wide Web conference (WWW 2014)

A Novel Voice Recognition System with Artificial Intelligence



Sabhav Gupta, Adarsh Pandey, Shivam Naruka, and Keshav Gupta

Abstract An AI personal assistant is a software that takes verbal commands and execute task assigned by the user. It is done using Python using libraries like pytsxs3, sapi5 and many more. The proposed voice assistant can also open YouTube, Gmail, Google Chrome and system applications. It can also show the current time, take a picture, check Wikipedia for data gathering, forecast the weather of multiple cities, get the daily news headline, search for the songs to play and much more. Aim is to create an assistant which answer voice and save user's time.

Keywords Pytsxs3 · Datetime · Speech recognition

1 Introduction

Speech recognition technology is that we have seen in movies. Writers' imagination has given birth to the future technology From Gideon in Flash to Ironman's Jarvis and Friday, sci-fi writers has set some standards that how this technology will look in future. However, achieving the accuracy will be a challenge. It will give new direction to human life; every sector can have advantage of using this tech. Not only this, when techs like big data, deep learning, machine learning, and AI will come together, then this tech will be on the heights no one can imagine. An AI personal assistant is a software that takes verbal or written commands and completes task assigned by the user. Facebook, Amazon, Microsoft, Google and Apple are providing these techs like amazon Alexa echo.

Speech recognition technology is a concept which is because it comes with various advantages from organizations to individuals, and it brings revolution to the traditional ways of working to the modern ways. The transcription ability it gives is one of the best benefits of speech recognition technology. The transcription ability it gives

S. Gupta (✉) · A. Pandey · S. Naruka · K. Gupta
Galgotias University, Greater Noida, India

K. Gupta
e-mail: Keshav.gupta@galgotiasuniversity.edu.in

is one of the best benefits of speech recognition technology. The transcription ability it gives is one of the best benefits of speech recognition technology. With the use of this technology, users can monitor devices from any location and create documents by speaking. This also helps in speeding up the process, i.e., the work is done when required easily just by the speech, which is sometimes much faster than someone can type. Speech solutions are not only used by individuals but also by other institutions and companies. This technology also make contributions in growth of these institutions. Institutions that offer customer services enjoy the benefits of the systems to make self-service in a way that delivers the best customer experience and lower cost of service. With the assistance of speech technology, without interfering with a real person, customers can access personal details, and other information. Instead of making callers idle, they can use this system to enable their clients involved. That is why, speech recognition technology helps in reducing the cost. However, the technology we are using is not sufficient as these also face many problems like lack of accuracy and misinterpretation. Voice recognition won't always give you the same output as expected because these systems are not able to understand the reference of language the way humans can which lead to output errors. Time cost and productivity: It is not necessary that the AI will save your time; sometimes it will take hell of an amount of time to take your commands. Background noise interfaces: Sometime, background noises will affect your input and you will get the desired input.

We have created a speech recognition system that will contribute toward efficiency. Using Python inbuilt packages like pytsxs3, Sapi 5 and speech recognition modules, we will create full working assistant which will take commands faster and provide much accurate results our proposed system will provide online as well as offline support and will perform quicker action. Our work will consist of Related work, Detailed working of proposed system, comparative study, future Scope and conclusion.

2 Related Work

Each developer uses unique methods to build this technology. One assistant can work accurately, some can work faster and some with some explanations. Therefore, there is no such assistant who works perfectly and does all tasks accurately because at the end, it is a man-made machine. Characteristics of the assistant depend upon the interest of developer where he gave his most attention. This paper gives knowledge about how a machine will take your voice as a command and supply a desired output. The following virtual personal assistant and adjustments that can be made to connect with the assistants are presented in this paper. Sumit Kumar Sarda, VPA: Virtual Personal Assistant Published in 2017 [1]. This paper provides a blueprint for creating a private assistant that only gives the voice command to restrict the use of any output device like the printer, keyboard or mouse. This paper gives idea about using personal assistant in our daily life instead of doing it manually.

P. Milhorat, Building the subsequent generation of non-public digital Assistants [2], Published in 2014. Voice-enabled digital assistant like Siri is one of the best known yet, but still as they promised, tech is not working accordingly so it could not be called a true personal assistant. In this paper, we highlight a number of problems arise while building speech recognition technology and propose several research and development directions we have undertaken to solve them. M. Al- focused in, Design of an Intelligent Home Assistant, published in 2006 [3]. An intelligent system for home which will control home appliances on a voice command. The only motive is to reduce human labor, effort, time and errors thanks to human negligence. The goal of this project is to style a voice control and remote-based intelligent system.

3 Architecture Diagram

See Fig. 1.

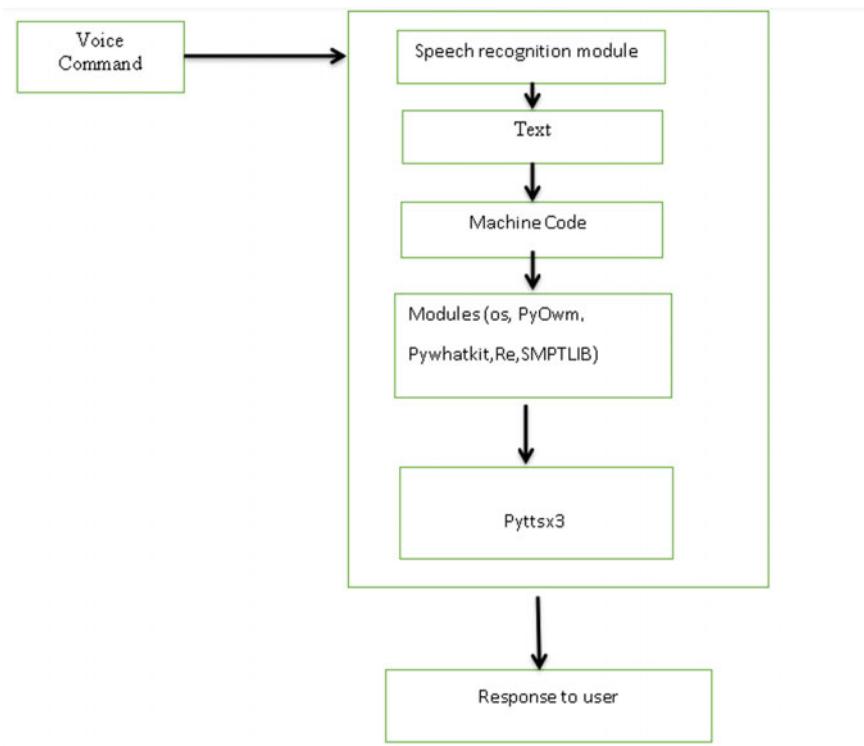


Fig. 1 Architecture diagram

4 Implementation

As the system got voice command, it will pass to a module called speech recognition. For converting the “Speech input” to “Text input,” it uses the online speech recognition system provided by Google. The microphone records and temporarily stores in the unit which can be used to access texts from special institutions that are arranged on the computer network system at the information center and then transferred for speech recognition to Google Cloud. The corresponding text is then received by the central processor and supplied to it. Now, our voice command is converted into text. In the next step, text will be converted into machine language, i.e., machine understandable code. Then, it goes to various libraries; according to voice command, some libraries are:

webbrowser: This module presents a greater platform that allows users to view Internet documents. You may use the Web browser module to enable a platform-independent browser. Os: This Python module provides communication functions with the operating system. This module is included in the Python simple utility modules. This module allows users to use functionality in a variety of ways depending on their operating system. Smtplib Python comes with the simple mail transfer protocol’s integrated smtplib module for sending emails. For SMTP, smtplib utilizes the RFC 821 protocol.

Beautiful Soup: It is a Python library for HTML, XML and other markup languages to retrieve data. For instance, we have identified some websites that show information relevant for our study, such as day or address information, but that does not provide a way to quickly download the data. There we can use it to retrieve the required information. PyOWM: It is a Python wrapper client library for Web APIs from OpenWeatherMap. It enables OWM data from Python applications to be accessed efficiently and simply via a basic application domain and in a human-friendly manner.

Re: A regular expression describes a string set that matches it. The functions of this module allow us to check if a particular string suit a specific regular expression (if a given regular expression matches any unique string). PyWhatKit: It is a Python library that is used for sending WhatsApp messages. It also provides many other functionalities such as playing music on YouTube using voice command. After going through these libraries according to user command, it retrieves data from Internet or the system you are working in. Now, result is ready to serve for giving output; it uses another library called Pyttsx3. In Python, it is a conversion library that converts the text to speech. It also runs offline, unlike alternative libraries, and works with both Python 2 and Python 3. By using this library, machine speaks the output in a very human way which saves time and generate interest while working, and the information we get is very much accurate resulting excellent results.

4.1 Working with an Example

See Fig. 2.

User gave voice command to retrieve recent updates about COVID-19. It passed the voice data to speech recognition module which converted voice data to text and further gave it to Webbrowser module which make possible for the machine to search over Internet as Internet is vast. Information is everywhere, but we trust Wikipedia. So Wikipedia module allowed webbrowser module to search over Wikipedia and retrieve information. Now, another module pytsxs3 came into action. It converted all the data which is in text form to voice data and finally end user got his desired output within no time.

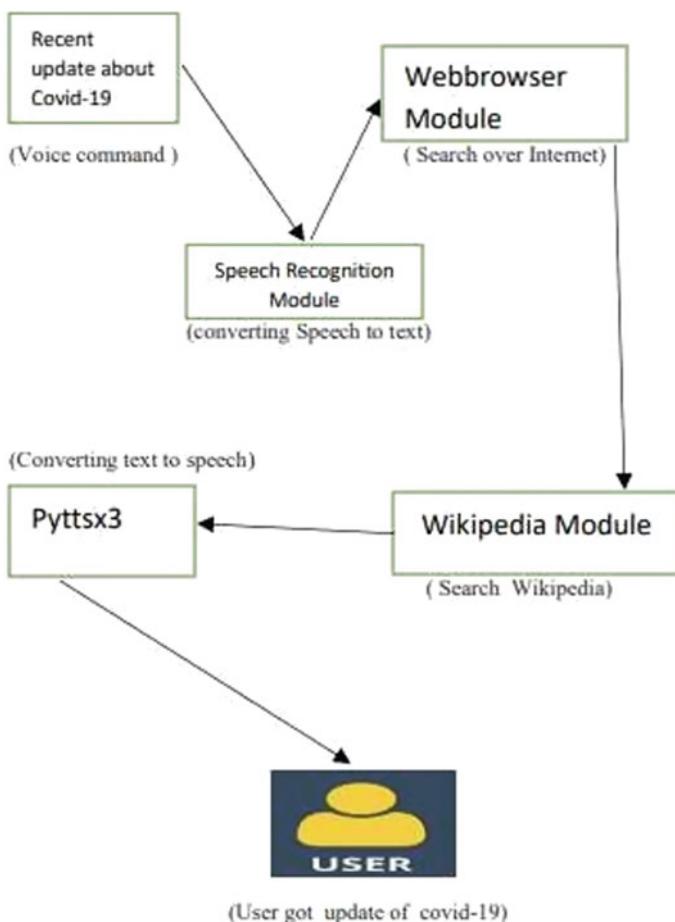


Fig. 2 Working with an example

Table 1 Comparative study

Features	Siri	Echo	Zee
Accuracy	✓	✓	✓
Speed	✓	✓	✓
Recognition	✗	✗	✓
Name adaptation	✗	✗	✓
Platform-independent	✗	✓	✓

5 Comparative Study

See Table 1.

6 Conclusion and Future Scope

We addressed the concept and implementation of a digital assistant in this paper. This project is developed using VS code, an open-source software module that can handle any changes in the near future. The modular structure of this project makes it more appealing and easier to add new features without modifying current system functionality. This not only responds to user commands, but also provides answers based on the user's question like opening any file or doing any other task. It embraces the user in a way that makes the user feel more relaxed and free to communicate with the voice assistant. Any unnecessary manual work involved in the life of the user to perform each and every task should also be excluded by the application. Instead of the textual one, the whole device operates on verbal feedback.

References

1. Sarda SK (2017) VPA: Virtual Personal Assistant Published in 2017
2. Milhorat P (2014) Building the subsequent generation of non-public digital Assistants^[2], Published in 2014. Voice-based digital Assistants like Apple's Siri and Google's Noware currentlybooming
3. Al-Amin M (2006) Design of an intelligent home assistant, Published in 2006^[3] An intelligent system for home automation could be a desired technology within the 21st century
4. Bohouta G, Kępuska VZ (2017) Comparing speech recognition systems (Microsoft API Google API and CMU Sphinx). Int J Eng Res Appl 2017
5. Artificial intelligence (AI), sometimes called machine intelligence. https://en.wikipedia.org/wiki/Artificial_intelligence
6. Thakur N, Hiwrale A, Selote S, Shinde A, Mahakalkar N Artificially intelligent chatbot
7. Mohasi L, Mashao D (2006) Text-to-speech technology in human-computer interaction. In 5th conference on human computer interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI), pp 79–84
8. Fryer LK, Carpenter R (2006) Bots as language learning tools. Lang Learn Technol

9. O'Shaughnessy D (2003) "Interacting with computers by voice: automatic speech recognition and synthesis" proceedings of the IEEE, vol 91, No 9, Sept 2003, Senior Member, IEEE
10. Hashimoto1 K, Yamagishi J, Byrne W, King S, Tokuda K (2011) An analysis of machine translation and speech synthesis in speech- to-speech translation system" proceedings of 5108978-1-4577-0539- 7/11/\$26.00 ©2011 IEEE
11. Goksel-Canbek N, Mutlu ME (2016) On the track of artificial intelligence: learning with intelligent personal assistant. *Int J Hum Sci* 13(1):592–601. <https://doi.org/10.14687/ijhs.v13i1.3549>
12. Feature: Von IBM Shoebox bis Siri: 50 Jahre Spracherkennung—WELT. [From IBM Shoebox to Siri: 50 years of speech recognition]. Die Welt (in German). Welt.de. 20 Apr 2012. Retrieved 10 Dec 2017
13. Chung H, Iorga M, Voas J, Lee S (2017) Alexa, Can I Trust You? Computer 50(9):104. ISSN 0018-9162. PMC 5714311. PMID 29213147. <https://doi.org/10.1109/MC.20173571053>

Investigation on Malware Detection Using Deep Learning Methods for Sustainable Development



M. Anusha and M. Karthika

Abstract The world of today is interconnected with computer systems, which is more liable to cyber-attacks. Cyber security was basically designed to help developers in understanding the modern information, system protection technology and methods using machine learning techniques is not a novelty to solve computer security challenges, and the progressive technology of deep learning has a significant thirst for knowledge in the community of cyber security for green environment. Deep learning was basically designed to detect and deal with cybersecurity issues like intrusion detection and prevention system, dealing with malware, spam and social implementation detection, network traffic, and user behavior analytics. This survey is a complete audit of different deep learning techniques for detecting malware. The aftereffect of this paper is an answer to an insufficient and complete beginning to malware recognition by utilizing deep learning by investigating most recent exploration papers in the field of deep learning to malware detection, the impact of datasets, algorithms, merits, and demerits.

Keywords Cyber security · Cyber-attack · Deep learning (DL) · Malware detection

1 Introduction

Cyber security is a prime issue in computer systems and network security [1]. It incorporates strategies, methods, innovations, and interaction, that are consolidated to save the confidentiality, integrity, and availability of registering system resources, organizations, programming projects information from attacks in Fig. 1. People and associations are exposed to cyber threats on instant access to the global network [2]. Digital safeguard systems exist at the application, organization, host, and information level by methods for different techniques as firewalls and antivirus programming have been being utilized to secure both client's protection and delicate information.

M. Anusha · M. Karthika (✉)

PG & Research Department of Computer Science, National College (Autonomous), Affiliated To Bharathidasan University, Trichirappali, Tamilnadu, India

Fig. 1 Resources of cyber security



Cyber security influences to incur more cost to ensure security all over the world. Even the US Government estimates a huge amount in malicious cyber activities protection. Malware is a great challenge to cyber security. This urges to detect malware by advanced techniques and great concern in computer security [3]. A huge numbers of researchers have analyzed different techniques for detecting and classifying malware. A traditional predictive tool depends on signature-based method [4]. Due to its limitation, many malware can easily run away from signature-based detection. Hence, machine learning for malware detection has been used as a tool to displace the signature-based anti-malware systems [5]. Malware detection approaches are classified as static analysis, dynamic analysis, and hybrid analysis approach. A static approach takes up source code and verifies the code without implementation to detect malware. Dynamic type monitors the interactions of the executed code to address the malware, whereas hybrid type takes the advantages of both static and dynamic. Traditional machine learning (ML) algorithms like SVM, Bayesian networks, logistic regression, and MLP have been used for malware detection and categorization [6–9]. Machine learning algorithms need lot of domain expertise, human intervention, and its shallow architectures make them not to scale well with large datasets. The fileless malware is harmful to any network due to its persistence and power to get away from any filtering solutions [10]. Hence, deep learning (DL) models are a fortunate way to overcome these limitations to malware detection and analysis mainly based on PC and Android for both file and file-less malware [11–14]. This survey aims to provide a significant contribution based on cyber-attack and detection of malware using deep learning and delighted to precede this research work. This survey is followed as Sect. 2 provides a brief history of different cyber security malware attacks and the various deep learning methods for cyber security is focused. Next Sects. 3 and 4 are the literature review on malware detection and an investigation study encountered with the platform used, datasets, and learning algorithms for cyber security using deep learning. Finally, provided with results and discussion and a conclusion based on deep learning.

2 Cyber Security Malware Attacks and Deep Learning Approaches

In malicious software hackers injects the malicious code to infect and obtain the access permission to the computer without the consent of users by means of file or fileless software is termed as malware attack in cyber security. The malware is of different types like spyware, ransomware, viruses, adware, worms, Trojan horses, etc. There are many reasons for the hackers to set going with injecting malware on any target system. The attackers explore the vulnerability of the already installed software or OS on the machine as a factor to inject the malicious code. Malware attacks are of different kinds, namely Trojan horse, Virus, Adware, Bot, Ransomware, Rootkit, and Spyware. Adware is an attacking software used for advertising takes the loop through popups on websites and ads. In Botnets attack, hackers assume responsibility for numerous frameworks and convey pernicious exercises like malicious activity. Targeted computers are used for malicious activities like distributed denial of service attacks, identifying theft, phishing, spoofing, and spamming [15]. Ransomware is malicious software for criminal money making by infecting the targeted computer and displays messages demanding a payment to continue work, unlock the encrypted file, and improve the performance via the links through a mail, popup message or website. A rootkit is a malignant programming, permits an unapproved admittance to the PC and to its confined zones in the product. Spyware is a malicious program expected to get to douse the figuring device by taking the Web usage data and fragile information or damage your PC, as often as possible without your understanding. A Trojan horse or Trojan is a kind of malware goes about as a genuine programming and gain secondary passage admittance to your framework to empower digital lawbreaker's exercises. A virus is a replicating software to infect the target machine by injecting its code in another software coding. Deep learning allows machines to solve cyber security problems. Since, it is a part of machine learning and artificial intelligence. Convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory network (LSTM), deep belief network (DBF), auto encoder (AE) are some popular kinds of deep learning models which are utilized in cyber security for identifying malware [16, 17]. CNN is a deep neural network, with different layers and are chiefly utilized for image recognition and object detection. CNN can process and extract features from data with the elements of multiple layers [2] like convolutional, rectified linear (ReLU), polling, and fully connected layers. RNN is a feedforward neural network, which has a significant obstacle gradient exploding or vanishing for training the data [1] in Fig. 2. RNN has been extended with numerous memory units as variants, including LSTM and gated recurrent unit to overcome exploding gradients and vanishing [8] (see Fig. 3). LSTM is a type of RNN that can learn and memorize long-term dependencies to resolve the vanishing gradient problem [5] Fig. 4. The Boltzmann machine is a probabilistic and unsupervised neural network, consists of binary values as hidden layer to decide about the activation but brings about passive learning. RBM is a stochastic network model to solve the complexity issue of Boltzmann machine by using two visible and hidden units. A deep belief

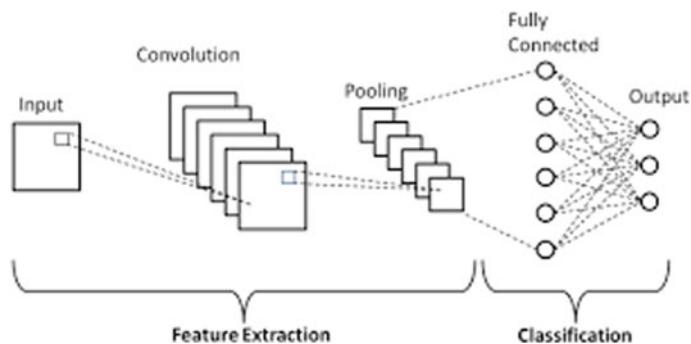


Fig. 2 Convolutional neural network

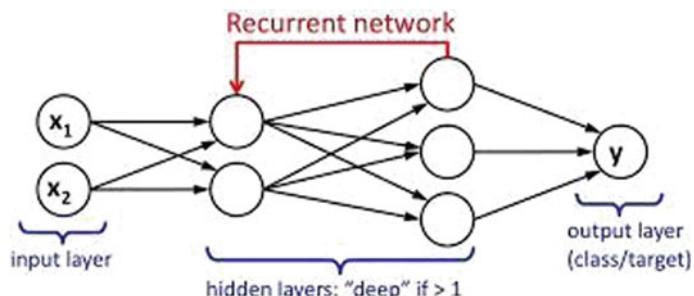


Fig. 3 Recurrent neural network

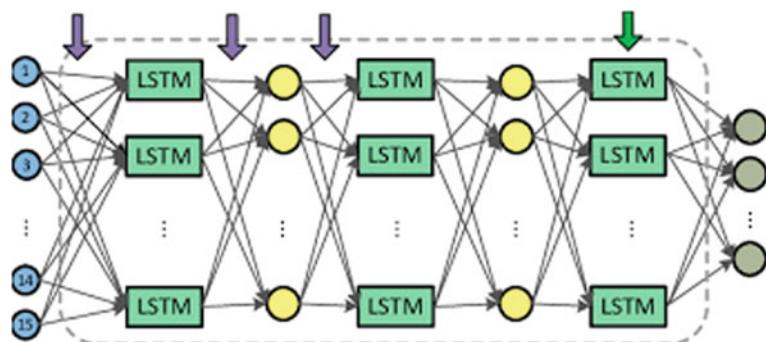
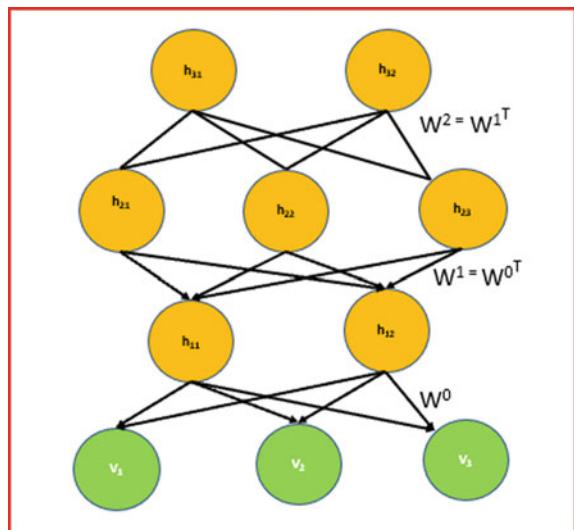
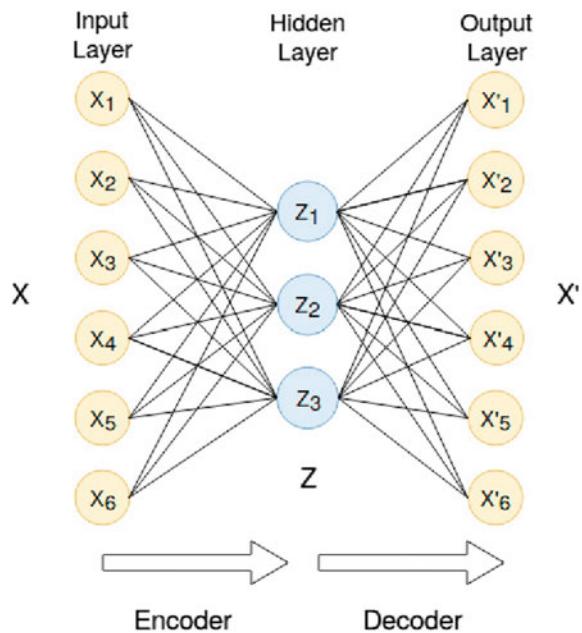


Fig. 4 Long short-term memory

network (DBN) is made out of stacked RBMs, which are prepared in an insatiable layered style to communicate with one another [8]. DBN overcomes the problem of not optimal solutions for dimensionality reduction [1] in Fig. 5. An auto encoder can be designed for dimensionality reduction and to have a better data representation

Fig. 5 Deep belief network

including layers and feature vectors by consolidating an encoder and decoder to train them together with back propagation to extract and learn the raw features and reconstruct the original features [8] in Fig. 6.

Fig. 6 Auto Encoder

3 Literature Review on Deep Learning-Based Malware Detection

Zhiqiang et al. [18] proposed an adaptable system dependent on DNN by applying distinctive feature reduction techniques, activation functions to get the best outcome by testing the model in ISCX 2012 and CICIDS 2017 datasets to detect zero-day attacks. It is observed with a precision value of 99.79 and 99.78%, recall value of 99.97% and 99.36%, FAR value as 0.21 and 0.22, F1-score as 99.79% and 99.78%, AUC as 99.12% and 99.91%, and the overall accuracy of 99.76% and 99.3% using the two datasets. Ahmed et al. [19] suggested a model to detect zero-day botnet attacks using deep learning, and the model is trained and tested with various neural network and hidden layers with CTU-13 dataset. The botnet detection accuracy 99.6% is reached in this model. Srinivasan et al. [20] had arranged his recognition to catch a few kinds of data in URL to identify variations of malicious URLs and named it as DeepURL Detect approach. It is computationally modest when contrasted with different important cutting edge deep learning-based character level embedding techniques and produces a precision of 97.4%, recall of 96.4%, F1-score of 96.9%, TPR value of 87.6%, FPR of 0.116, AUC as 0.9964, and the detection accuracy as 97.2%. Pei et al. [21] build a novel framework (AMalNet) using reliable deep learning methods like graph convolutional networks (GCNs) for displaying undeniable-level graphical semantics, which naturally recognizes and learns the semantic and consecutive examples, and utilizations different benchmark datasets with independently recurrent neural network (IndRNN) to decode the deep semantic data, utilizing far off ward data between hubs to freely extricate highlights for Android malware identification and to distinguish its family attribution. The proposed model results with precision of 99.52%, recall of 99.4%, F1-score of 99.46% and accuracy of 99.47%. Cui et al. [22] proposed an attack distinguishing proof model LeNet-4 convolutional neural network by disposing of the first pooling layer and the last completely association layer of the first LeNet-5 network. LeNet-4 decreases the computational load and network complexity in the model, and the network self-learning capacity was improved by assessing CICIDS2017 dataset by its design as two convolutional layers and one pooling layer. The recursive feature elimination algorithm in LeNet-4 has enhanced the accuracy rate of 97.8% and 98.5% on multiclass and binary attack identification while diminishing the time cost. Zhang et al. [14] projected a new feature engineering model to detect malware by using hashing and gated approach and extracted a test performance report as recall of 95.33%, AUC of 98.71%, and an accuracy of 95.33%. Darabian et al. [15] used LSTN and CNN strategies to advance the analysis of cryptomining malware by analyzing it in cuckoo sandbox and achieves an accuracy rate of 95% and 99% in static and dynamic analysis. Amin et al. [23] investigated the use of deep learning with fully connected neural network (NN), CNN, AE, DBN, and RNN to detect static malware in Android-based smartphones by interrogating the nomenclature of a large-scale byte-code dataset and prove that it beats the state-of-the-art. The contributed model has achieved an F1-score of 0.999, AUC of 0.988, and accuracy of 0.999. Imtiaz et al. [24] manages another scope of

difficulties for the vendors of the devices, software developers, cybersecurity professionals, and comes out with novel approach DeepAMD and is assessed by utilizing the current CICAndMal2019 dataset for distinguishing and recognizing Android malware on the static as well as dynamic layer. DeepAMD model results with a precision of 93.5%, recall of 93.4%, and achieves an accuracy rate of 93.4% in binary classification. Gibert et al. [25] have introduced a multimodal deep learning novel malware characterization system by joining both hand-engineered features and end-to-end components in a modular architecture to detect and classify malware by using CNN, DeepConv, and Malconv models. HYDRA model has a macro F1-score of 99.54% and accuracy of 99.75%. Jha et al. [26] proposed an efficient method for malware detection using RNN and Kaggle dataset. This RNN model has been measured with various word embedding feature vectors and brings out with an AUC value of 91.69, F1-score of 91.76, and accuracy of 91.91%. Tian et al. [27] dealt with overfitting, low classification accuracy, and high false positive rate issues in the existing intrusion detection systems by using improved deep belief network (DBN) and propagation tests are performed on the various public dataset. Lu et al. [28] recommended an innovation with new static features with solid anti-obfuscation capacities are added, and the dynamic features of the application programming at runtime are separated to enhance the Android malware include set by building a cross breed deep learning model dependent on DBN, RNN, and gate recurrent unit (GRU) by utilizing different dataset to identify Android malware. This hybrid model was performed with both benign and malware samples and attains a precision of 95.79%, recall of 97.62%, and accuracy of 96.82% for malware samples. Sun et al. [29] proposed a deep learning-based intrusion detection system based interruption recognition framework by utilizing the crossover network like CNN and LSTM to separate the spatial features and temporal features of network traffic to recognize the malware and interruption. A dataset which includes all the common, updated intrusions, and cyber-attacks is CICIDS2017 and tested for DL-IDS to explore an accurate detection rate of 98.67% and F1-score of 93.32%. Kim et al. [30] analyzed the universal use of smartphones and the exponential increase of number of malware. To dispense this issue, a novel Android malware detection framework is based on MNN approach and evaluated the model with Virus Share, Google Play App store datasets and produces a precision of 0.98, recall of 0.99, F-measure of 0.99, and accuracy of detection as 98%.

4 Investigation Study

A relative investigation is led to audit the various attacks experienced based on model, datasets in the area of cyber security using deep learning are listed in Table 1.

Table 1 Comparative study on malware detection

Author/Year	Model	Dataset	Merit	Demerit
Zhiqiang et al. [18]/2021	DNN	ISCX 2012 CICIDS2017	Binary classification for zero-day malware detection	Dynamic classification for detecting attacks
Ahmed et al. [19]/2020	DNN	CTU-13	Neural network identifies the random noise instead of relationships	Detects only botnet attack
Gibert et al. [25]/2020	Multimodal DNN	Kaggle	Modality fusion for malware classification	More data modalities needed for analysis
Kim et al.[30]/2018	Multimodal DNN	Android APK	Multiple feature types are encompassed to detect malware	Dynamic feature extraction for malware detection
Cui et al. [22]/2020	CNN	CICIDS2017	Recursive feature elimination algorithm to reduce the dimensionality	Individual attack recognition for classification
Srinivasanet al. [20]/2021	CNN LSTM	URL	Character-level Keras embedding for URL malware detection	Auxiliary modules for robust solution
Sun et al. [29]/2020	CNN LSTM	CICIDS2017	Feature fusion for intrusion detection	LSTM unit quantity, training packet length, per-flow packet quantity, batch size and weight affects the performance
Zhang et al. [14]/2020	Gated-CNN Bi-LSTM	Real-world	Hashing trick to process heterogeneous information gated-CNN to transform high-dimensional hash feature for malware detection	Algorithm failed to work on virtual address
Darabian et al. [15]/2020	CNN ATT-LSTM	Cuckoo Sandbox	Attention network for calculating the word weight	Multi-view learning for classification

(continued)

Table 1 (continued)

Author/Year	Model	Dataset	Merit	Demerit
Amin et al. [23]/2020	Bi-LSTM	Android malware dataset	Automatic and feature engineering to detect malware in low power and memory-limited devices	Time-consuming for training
Jha et al. [26]/2020	RNN	Kaggle	Negative sampling technique to reduce the computational cost	Specialized feature extraction method to detect malware
Pei et al. [21]/2020	GCN IndRNN	DREBIN, AMD Lab-built, AndroZoo, Praguard	Multiple embedded for Android malware detection	Deep hybrid exploration for malware detection
Imtiaz et al. [24]/2020	DeepANN	CICInvesAndMal2019	Identify Android malware on Static and dynamic layer	Online malware detection service
Tian et al. [27]/2020	DBN	NSL-KDD UNSW-NB15	Sparsity penalty to avoid overfitting	Customization of parameter selection
Lu et al. [28]/2020	GRU DBN	Google Play, APKpure VirusShare, PRAGuard	Anti-obfuscation ability to extract malware feature	Limited computational resources for malware detection

5 Analysis and Discussion

This survey was analyzed and discussed on cyber security attacks and an exhaustive survey of deep learning strategies to detect malware in personal computer, Android smartphones, URL, and network traffic. The overwhelming deep learning techniques are compared with the traditional machine learning in the field of network protection with selected papers interrelated to this review topic. As for various deep learning techniques concerned, cyber security issues take place in various real-world issues like zero-day attack, malicious URL attack, botnet attack, malicious Android application attack, cryptomining malware attack, real-world Android malware attack, and intrusion in network traffic data attack with the means of intruders and hackers. Based on the above literature review, it is identified that deep learning approach is giving prominent solution for malware detection and its accuracy rate in Fig. 7. It is considered with conventional datasets and real-world datasets for identifying malware for a green sustainable environment.

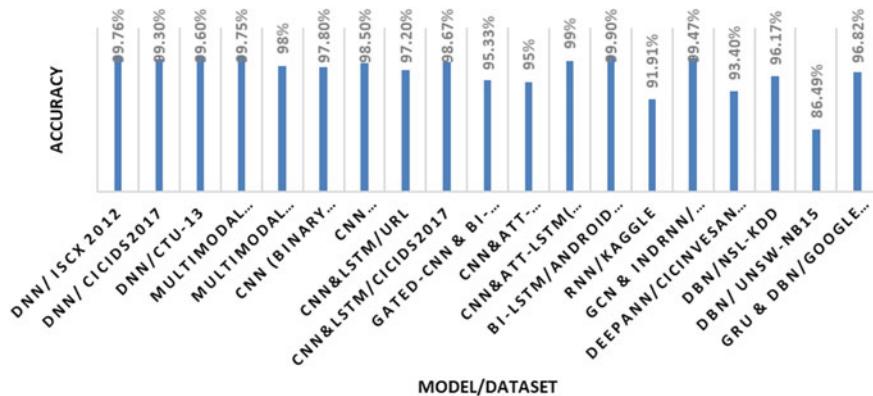


Fig. 7 Investigation report of detection accuracy on malware

6 Conclusion

In this paper, a relative investigation review of deep learning approaches for malware detection by means of static and dynamic features. Special analysis about deep learning techniques includes convolutional neural network, recurrent neural network, long short-term memory network, deep belief network, and auto encoder. In this way, it is noted that deep learning techniques play a vital role in analyzing the feature extraction for identifying malware in Android, URL, botnet, cryptomining, and intrusion in network traffic with a large datasets like CICIDS2017, CICInvesAndMal2019, Android ADK, Cuckoo Sandbox, and NSL-KDD for the detection of malware. The limitation of this work is to include a concentration on other malware attacks like DoS, phishing, and messaging attack. Hence, this review aims to provide a significant contribution to cyber-attack; based on the algorithms, merits and demerits, it is motivated and attracted to proceed the research using hybrid approaches of deep learning for malware detection.

References

1. Drewek-Ossowicka A, Pietrołaj M, Rumiński J (2020) A survey of neural networks usage for intrusion detection systems. *J Ambient Intell Humanized Comput*
2. Berman DS, Buczak AL, Chavis JS, Corbett CL (2019) A survey of deep learning methods for cyber security. *Information* 10(4):122
3. Wang Z, Liu Q, Chi Y (2020) Review of android malware detection based on deep learning. *IEEE Access* 8:181102–181126
4. Yan J, Qi Y, Rao Q (2018) Detecting malware with an ensemble method based on deep neural network. *Secur Commun Networks*
5. Mahdavifar S, Ghorbani AA (2019) Application of deep learning to cybersecurity: a survey. *Neurocomputing* 347:149–176

6. Alabadi M, Celik Y (2020) Anomaly Detection for cyber-security based on convolution neural network: a survey. In 2020 international congress on human-computer interaction, optimization and robotic applications (HORA). IEEE, pp 1–14
7. Sohn I (2020) Deep belief network based intrusion detection techniques: a survey. *Expert Syst Appl* 114170
8. Aldweesh A, Derhab A, Emam AZ (2020) Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues. *Knowl Based Syst* 189:105124
9. Ferrag MA, Maglaras L, Moschouyannis S, Janicke H (2020) Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. *J Inf Secur Appl* 50:102419
10. Kumar S (2020) An emerging threat Fileless malware: a survey and research challenges. *Cybersecurity* 3(1):1–12
11. Mercaldo F, Santone A (2020) Deep learning for image-based mobile malware detection. *J Comput Virol Hacking Tech* 1–15
12. Dixit P, Silakari S Deep learning algorithms for cybersecurity applications: a technological and status review. *Comput Sci Rev* 39:100317
13. Vinayakumar R, Barathi Ganesh HB, Poornachandran P, Anand Kumar M, Soman KP (2018) Deep-net: deep neural network for cyber security use cases. arXiv preprint [arXiv:1812.03519](https://arxiv.org/abs/1812.03519)
14. Zhang Z, Qi P, Wang W (2020) Dynamic malware analysis with feature engineering and feature learning. In: Proceedings of the AAAI conference on artificial intelligence, vol 34, no 01, pp 1210–1217
15. Darabian H, Homayounoot S, Dehghantanha A, Hashemi S, Karimipour H, Parizi RM, Choo KKR (2020) Detecting cryptomining Malware: a deep learning approach for static and dynamic analysis. *J Grid Comput* 1–11
16. Vinayakumar R, Soman KP, Alazab M (2020) A comprehensive tutorial and survey of applications of deep learning for cyber security
17. Choi YH, Liu P, Shang Z, Wang H, Wang Z, Zhang L, Zhou J, Zou Q (2020) Using deep learning to solve computer security challenges: a survey. *Cybersecurity* 3(1):1–32
18. Zhiqiang L, Zhijun L, Ting G, Yucheng S (2021) A three-layer architecture for intelligent intrusion detection using deep learning. In Proceedings of fifth international congress on information and communication technology. Springer, Singapore, pp 245–255
19. Ahmed AA, Jabbar WA, Sadiq AS, Patel H (2020) Deep learning-based classification model for botnet attack detection. *J Ambient Intell Humanized Comput* 1–10
20. Srinivasan S, Vinayakumar R, Arunachalam A, Alazab M, Soman KP (2021) DURLD: malicious URL Detection using deep learning-based character level representations. In Malware analysis using artificial intelligence and deep learning. Springer, Cham, pp 535–554
21. Pei X, Yu L, Tian S (2020) AMalNet: a deep learning framework based on graph convolutional networks for Malware detection. *Comput Secur* 101792
22. Cui W, Lu Q, Qureshi AM, Li W, Wu K (2021) An adaptive LeNet-5 model for anomaly detection. *Inf Secur J Global Perspect* 30(1):19–29
23. Amin M, Tanveer TA, Tehseen M, Khan M, Khan FA, Anwar S (2020) Static malware detection and attribution in android byte-code through an end-to-end deep system. *Futur Gener Comput Syst* 102:112–126
24. Imtiaz SI, ur Rehman S, Javed AR, Jalil Z, Liu X, Alnumay WS (2020) DeepAMD: detection and identification of android malware using high-efficient deep artificial neural network. *Future Gener Comput Syst* 115:844–856
25. Gibert D, Mateu C, Planes J (2020) HYDRA: a multimodal deep learning framework for Malware classification. *Comput Secur* 101873
26. Jha S, Prashar D, Long HV, Taniar D (2020) Recurrent neural network for detecting malware. *Comput Secur* 99:102037
27. Tian Q, Han D, Li KC, Liu X, Duan L, Castiglione A (2020) An intrusion detection approach based on improved deep belief network. *Appl Intell*
28. Lu T, Du Y, Ouyang L, Chen Q, Wang X (2020) Android malware detection based on a hybrid deep learning model. *Secur Commun Networks* 2020

29. Sun P, Liu P, Li Q, Liu C, Lu X, Hao R, Chen J (2020) DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system. *Secur Commun Networks* 2020
30. Kim T, Kang B, Rho M, Sezer S, Im EG (2018) A multimodal deep learning method for android malware detection using various features. *IEEE Trans Inf Forensics Secur* 14(3):773–788

RSSI Strength Measurement in Wireless Sensor Network with and Without Obstacles



Santosh Anand, Roshan Muralidharan, K. Manoj, C. Shreyas, and Hruthik Kariappa

Abstract In today's world, the influence of high-fidelity wireless sensor networks based on system on a chip (SoC) architecture has produced a robust framework of networking possibilities. These networks provide better latency and efficient power consumption in the WSN. This work presents a novel approach to analyze the effective variation in receiving signal strength indicator (RSSI) between two ESP8266 Wi-Fi microchips with or without the presence of obstacles. The proposed devices are configured to emit, scan and fetch the signal strength of one or more host networks at the same time. It can help to deploy the sensors in the network more effectively with respect to the distance. In this energy of nodes are getting optimized and communication link establishment between the neighbor nodes. This could assist in improving network switching capabilities by automatically analyzing the signal strength of the available networks from the most viable distance and connecting to them seamlessly in real time. This proposed architecture can be further enhanced by integrating with ESP-supported devices, which can be applied in exposure detection and contact tracing systems with the less power consumption in the WSN. This device can make the great change in the WSN deployment, so redundant nodes energy will be utilized properly with the help of RSSI values. So nodes can use the transmission energy with respect to the distance, which will make high energy optimization of the node, instead of using the same energy level of the short and long distance neighbor nodes.

Keywords IoT · RSSI · SoC · Wi-Fi · Arduino

1 Introduction

The evolution of wireless networking hardware systems has revolutionized the communication industry. Since its adaptation, there is been a wide range of applications that are focused on providing effective means to communicate and share information. This provided users to seamlessly interact and engage with multiple

S. Anand (✉) · R. Muralidharan · K. Manoj · C. Shreyas · H. Kariappa
Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India

communities and people from anywhere around the world. In fact, these systems have fostered the growth of service providers and established various networking tools and protocols. In today's scenario, wireless applications can be applied in multiple domains ranging from AI automation to threat detection monitoring systems. This has led to the immense rise of IoT frameworks that can interconnect multiple wireless networking systems with different functionalities and configurations in real time by sending RF signals. Most of us are leaned toward wireless fidelity network interface cards (NIC), which provide instant access to the internet and support data sharing with multiple external devices. It even provides users the accessibility to remotely store and fetch data from the cloud and control paired devices effortlessly. Progressive advancements in this field have accelerated the growth of portable consumer devices and other multi-purpose accessories.

A wireless sensor network (WSN) system comprises thousands of minuscule sensors that are placed in a systematic approach. These sensors communicate using radio signals and are liable for handling the recorded data to external networks. Each node is interconnected and shared bidirectionally when required. This makes it much more customizable and increases the overall power efficiency. WSN follows the conventional forms of topologies, making it more flexible and enabling it to cope with multiple node failures. The signal strength of WSN systems depends on the router as it measures the power of the emitted radio frequency and sends the generated signals to the device.

2 Objectives

The main objective is to measure the received signal strength indicator (RSSI) value of ESP8266 module with respect to the interference of external obstacles. This could potentially assist in enhancing the network detection capabilities and automatically switch to a stronger network with minimal packet loss in WSN. This developed system could identify multiple active and available networks and fetch RSSI strength of each network. Based on the value of the RSSI strength, it displays the computed fuzzy distance of each networks in real time. The serial monitors scan for further networks after each detection and synchronize with the updated. The dynamic range can be used based on the distance between the nodes to reach to the sink in WSN. In this, measurement of RSSI strength is the important factor as it determines the availability of a particular network when it is scanned during the sync rate.

3 Literature Survey

Anand and Sinha [1] provided a system that enables effective communication for the soldiers who are on the battlefield to their base camp. This is done by embedding a GPS module with low-power wireless body sensors in the system. Communication

is set up in two ways using an antenna and using a GSM module. Anusha et al. [2] contribute a method to accurately find the location of wireless sensor networks by calculating the RSSI strength based on the GPS signal. This uses an advanced dimensional distance measuring technique that can predict the network propagation based on its position and arrangement. Manishankar et al. [3], in this work, applied an in-depth analysis of the quality of service in wireless sensor systems by contrasting the different kinds of existing available wireless networks. The mobility and speed are determined based on their respective functionality, mobility and receiving signal strength. Nagarjun et al. [4] propose the vulnerability of the existing wireless sensor networks. This focuses on determining nodes that are fake and could explicitly less reliable. The author analyzes the packet loss and remaining energy to estimate the node's functionality through simulation. Mukhopadhyay et al. [5] conduct a study that deals with network testing of multiple nodes in real time using a hypervisor technique. This was accomplished because of the lack of accuracy of network simulators which provide estimated results based on the given conditions. Manoj Kumar et al. [6] in this work, authors using state-of-the-art algorithm which focus on the sending the data to its neighbour node during the fluctuation in the network. The author provides several circumstances under which this occurs like variation in energy and efficiency. Manishankar et al. [7] state that this system is based on establishing and integrating both types of network simulators to provide a wide-scale environment for testing the functionality which would provide a better platform to choose from while testing networking applications and concepts that are advanced and require more testing scenarios. Prabhan et al. [8], in this paper, developed a way to identify nodes that are not working. The authors calculate the latency in communication between the respective nodes and evaluate the residual energy attained. Barai et al. [9] of this paper observe a way to utilize a wireless network to determine the position of an object in indoor spaces by using vector quantization techniques. Xue et al. [10] of this paper propose an algorithm that determines the exact location in indoor spaces based on the generated signal strength. Dobrilovic et al. [11], in this paper, focus on establishing anchor nodes that dynamically positions based on the location at which the initial signal has been generated. Sourya et al. [12], in this paper, propose an alternate method to help a person find his very own car from a parking location. This uses localized GPS for effective positioning. Lova Raju et al. [13] mainly focus on enhancing home automation in the field of IoT by implementing a prototype that could remotely control the systems from the internet and monitor the activities of users. Patchava et al. [14], in this paper, proposed a technique that allows surveillance systems to record when motion is detected. Anamul Haque et al. [15] of this paper focus on providing a cost-effective way to control indoor locations by establishing a dedicated micro-web server and monitor system that supports various protocols and plugins.

4 Proposed Methodology

4.1 *ESP-ARDUINO Configuration*

The ESP8266 is one of the most popularly used cost-effective Wi-Fi microchips which has a transmission protocol stack along with microcontroller capabilities and functionalities. The Arduino UNO is a microcontroller which is used as an interface to establish the network and display the output. The ESP is configured with the Arduino by overloading the bypass system. In order to attain this, the Arduino board is flashed with an empty code, and then using the breadboard as a medium, the power(5v) of Arduino is connected directly to the board using the connector pins. Once this is established, the ESP is connected as follows: The GND pin of ESP is connected to the GND of breadboard. The TX pin of ESP goes to the TxD port in the Arduino. The CH-PD pin is attached to the power line in the breadboard. The VCC port is directly connected to the streamlined power source. The GPIO 0 is connected to the GND in breadboard while flashing the code in the microcontroller. The RXD pin of ESP goes to the RX pin present in the ESP.

After this once the Arduino IDE is loaded, select the connected COM port and import the generic ESP package from the Arduino board manager. The imported firmware uses a baud rate of 115,200 which is modified in the serial monitor. When the applied code is compiled and uploaded, the GPIO 0 pin is removed from the ESP module. In the serial monitor to make sure to set the line ending to NL and CR, the updated changes can be verified when the AT command is passed on the message field in the serial monitor as this would return an OK message indicating a successful configuration between the two modules. The proposed connection enables both the hardware connected modules to adhere its functionalities and expands better portability.

4.2 *Configuring Access Point and Station Networks*

The proposed system consists of both the ESP8266 modules in which one is configured as an ACCESS POINT and the other as the STATION network.

Access Point. An access point acts as the host which generates the wireless network to which we can connect Wi-Fi compatible systems. It enables to remotely establish a bidirectional communication with multiple station networks.

Station Mode. When the ESP8266 module is set to station mode, the module acts as a portable device which can be used to connect to the access point. It can be configured according to a particular network and can act as the interface for the end user.

In this scenario, Module A acts as the access point and Module B acts as the station network. Module A generates the Wi-Fi signals; the Service Set Identifier (SSID) is set to “ESP8266” along with a password so that the station could access

that particular AP itself. The baud rate for ESP is set to default, and the ESP libraries are imported in the Arduino IDE so as to choose the “Generic ESP8266 Module.” The code is flashed then onto the module. The GPIO pin is disconnected from GND. The circuit is then powered via power bank in order to make it standalone and place it in a constant location. Module B is configured as the station network which gets connected to the ESP access point generated by Module A by verifying the SSID and password. The Module B is flashed with the potential to: Scan all AP networks rapidly, show the RSSI strength both in dBm and percentage generated by each network, show the estimated distance of the module with respect to the scanned access points.

The access point as shown in Fig. 1 is initiated so as to generate the ESP hotspot, after this, the Arduino IDE is booted up on a portable PC and the station module is connected to the PC via Arduino to USB. In this case, it is COM3 port. When the serial monitor is launched in the IDE, the station module as represented in Fig. 2 readily scans for nearby active networks and displays the RSSI value of each scanned network respectively as shown in Fig. 3. After the initial scan, the ESP station module frequently scans for new networks and updates it onto the serial monitor.

The scanning is done until the station module is disconnected from the PC. In Fig. 3, we have converted the RSSI strength to percentage in order to gain a better insight. The observations carried on the literature survey enabled us to resolve the initial problems such as flashing the ESP module and understanding the working process of both the SoC modules as shown in Fig. 4.



Fig. 1 Access point Module A



Fig. 2 Station network Module B

Figure 1 shows the Module A which is connected through the power; in this, range is the major factor. This module shows that when the distance between sender and receiver (Module B Fig. 2) increases, it decreases the signal strength which is measured in the decibel. This value can be converted in the watt and joule to check the lifetime of the node and wireless sensor networks, in which these modules are deployed for the purpose such as monitoring and controlling.

Figure 3 shows the signal strength with respect to the distance which is measured by.

Arudino IDE is used to implement the proposed work, output value of COM3 Port will be used for the communication. With the help of this value, sender can decide the selection of the neighbor node for the effective communication, which also shows that node which cannot able to communicate with the sender directly because of threshold values of signal strength required for the communication. It is represented in the tables. Fig. 4 shows the architecture of the system, which provides a much more further briefing into the process of the communication between the two micro-WSN modules. Here, when a node sends information as RF signals, the signal passes through the TCP/IP protocol layer and sends the RF packet to the receiving node which generates the output based on what has been programmed in the microcontroller. After this process, the microcontroller displays the output and transmits back RF signals to scan to the receiver network.

COM3

```
15:47:01.441 -> gauthamRSSI dbM:-91
15:47:01.441 -> 18% )Distance:6.31metres
15:47:06.453 -> Wifi scan started
15:47:08.054 -> Wifi scan ended
15:47:08.054 -> 3 networks found
15:47:08.054 -> 1)
15:47:08.054 -> ESP8266RSSI dbM:-42
15:47:08.054 -> 100% )Distance:0.27metres2)
15:47:08.054 -> JioFiber-DycQ6RSSI dbM:-90
15:47:08.054 -> 20% )Distance:10.80metres3)
15:47:08.102 -> iPhoneRSSI dbM:-53
15:47:08.102 -> 100% )Distance:0.63metres
15:47:13.093 -> Wifi scan started
15:47:14.680 -> Wifi scan ended
15:47:14.680 -> 4 networks found
15:47:14.680 -> 1)
15:47:14.680 -> ESP8266RSSI dbM:-40
15:47:14.680 -> 100% )Distance:0.33metres2)
15:47:14.728 -> DIRECT-Lu-BRAVIARSSI dbM:-91
15:47:14.728 -> 18% )Distance:6.31metres3)
15:47:14.728 -> iPhoneRSSI dbM:-54
15:47:14.728 -> 100% )Distance:0.75metres4)
15:47:14.728 -> JioFiber-DycQ6RSSI dbM:-92
15:47:14.728 -> 16% )Distance:6.68metres
15:47:19.727 -> Wifi scan started
15:47:21.311 -> Wifi scan ended
15:47:21.311 -> 5 networks found
15:47:21.311 -> 1)
15:47:21.311 -> ESP8266RSSI dbM:-39
15:47:21.359 -> 100% )Distance:0.40metres2)
15:47:21.359 -> JioFiber-DycQ6RSSI dbM:-90
15:47:21.359 -> 20% )Distance:4.17metres3)
15:47:21.359 -> iPhoneRSSI dbM:-56
15:47:21.359 -> 88% )Distance:0.87metres4)
```

Fig. 3 RSSI strength based on the network distance

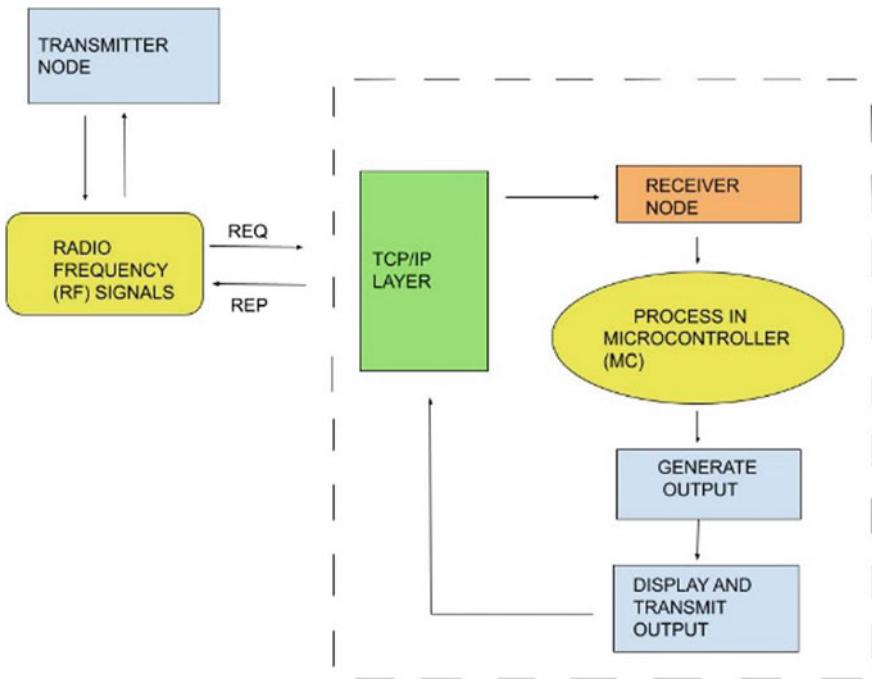


Fig. 4 System architecture

4.3 Steps to Configure the Access Point

- Step 1: Start Algorithm
- Step 2: Read SSID and Password
- Step 3: Set the ESP module to its default baud rate
- Step 4: Set Wi-Fi mode to Access Point
- Step 5: Set a Software Access point and add parameters from Step 3
- Step 6: Initialize IP and fetch IP address then Print IP
- Step 7: End Algorithm

4.4 Steps to Configure the Station Network

- Step1: Start Algorithm
- Step 2: Declare SSID, Password as character pointers
- Step 3: Declare Rssi_high, Rssi_low, Distance A and N as variables
- Step 4: Check connection for given SSID and Password
- Step 5: On Connection, Print IP Address
- Step 6: Scan for connections,

```

Print WiFi SSID and WiFi RSSI, dBmToPercentage(rssi)
Calculate distance, Print distance end if, delay (10)
Step 7: function dBmToPercentage(int dBm), Declare integer variable
if(dBm is less than or equal to Rssi_low) then, set quality to Nil
else if dBm is greater than Rssi_high set quality to 100 else
set quality to (2 * (dBm+100)) end if Print Status
Step 8: End Algorithm

```

5 Mathematical Model of the System

The received signal strength indicator (RSSI) value is measured in dBm, in order to establish an effective relationship between the difference in variation of the signal strength with respect to distance which is being measured during the transmission process. In the proposed work, received signal strength will be in the form of the Decibel-Milliwatts (dBm), which can be calculated by representing the ratio of two quantities of power; the referenced power is 1 milliwatt which implies:

$$\text{dBm} = \frac{10 \log P_1}{1 \text{ mw}} \quad (1)$$

From Eq. 1, received power of the signal can be calculated in the watt; if we invert this equation in terms of Power P_1 , then, according to this, P_1 will be as follows,

$$P_1 = 1 \text{ mw} \cdot 10^{\frac{\text{dBm}}{10}} \quad (2)$$

For an efficient path loss model in free space, Friss equation is revamped here as the gain in both transmitting and receiving antennas is negligible. R is set to be the distance across the transmitter and receiver antennas. P_r, P_t is the power at transmitting and receiving antennas and C is represented as the speed of light traveled.

This helps in mapping the distance of our proposed model.

$$R = \frac{c}{4\pi f} \sqrt{\frac{P_t}{P_r}} \quad (3)$$

By establishing a relationship between the distance, power of transmitter and power of receiver with respect to the frequency used in the communication as shown in Eq. (3), the energy with respect to distance as shown in the figures, how the distance varies with respect to the signal strength. As per the findings, it is evident that the RSSI strength decreases with increase in distance which when formulated results as shown in Eq. 4.

$$R \propto \frac{1}{P_r \cdot P_t} \quad (4)$$

If we consider a variable n as the maximum range in the proposed methodology according to Eq. 7, the summation is applied to the distance R which is inversely related with the power factor. This will help to calculate the distance between source and destination in the WSN.

$$\sum_0^{n-1} R = \frac{1}{P_r \cdot P_t} \quad (5)$$

By substituting the values from (2) to Eq. (6), energy is the product of power times distance per the constant speed of light in free space we get,

$$\text{Energy} = \frac{1 \text{ mw} \cdot 10^{\frac{\text{dBm}}{10}}}{f \cdot \lambda} \cdot \frac{1}{P_r \cdot P_t} \quad (6)$$

When both the parameters are proliferated, we get our contemplated equation:

$$\text{Energy} = \frac{1 \text{ mw} \cdot 10^{\frac{\text{dBm}}{10}}}{f \cdot \lambda \cdot P_r \cdot P_t} \quad (7)$$

Therefore, as per the Eq. (7). Under these circumstances, it is evident that energy is inversely proportional to the distance R , which in fact means that there is a relationship between these two entities that transcribes to the variation in RSSI strength as per the recorded distance. Nature of sensors in the WSN is randomly deployed as well as redundant nodes. So these techniques can be used to enhance the lifetime of the node as well as WSN, to make the nodes transmission according to the distance. So nodes will use the energy with respect to the distance of neighbor nodes instead of using the same transmission power for the near and far nodes.

6 Result and Analysis

In this, RSSI strength is evaluated depending on the distance and the number of obstacles which is shown in Table 1. The substantial variation in the values depends on many factors. ESPs have a brief history of hanging onto multiple networks for a longer range; however, there is a sufficient packet loss when it is redirected above the estimated threshold distance as referred in Table 1 in which there are no obstacles, and both the modules are kept at line of sight (LoS).

Both module A and B kept at the different distance without any obstacles as shown in Table 1 and measured the RSSI value with respect to the distance, which clearly shows that distance increases which makes the week signal strength. The obstacles in this scenario are blocks of “walls,” so both the AP and station are kept in between a particular plain wall at a linear angle such that the RSSI strength and distance are measured appropriately. When both the modules are powered, the station

Table 1 RSSI-distance measured under 0 obstacles

SL. No.	Phase-I		
	Distance in meters	No of obstacles	RSSI strength (dBm)
1	20	Nil	-67
2	40	Nil	-72
3	60	Nil	-78
4	80	Nil	-83
5	100	Nil	-89
6	120	Nil	-94
7	160	Nil	-104
8	200	Nil	-116
9	260	Nil	-132
10	300	Nil	-144

module starts the scanning and detects the access point (AP) module and establishes a connection. Figure 5 shows the graphical representation of Table 1. Table 2 records the variation in signal strength as per distance under the influence of 1 obstacle. Figure 6 indicates the graphical representation of Table 2.

The noted deviation as shown in Fig. 6 in RSSI strength indicates a semi-stable decrement as the distance is increased. Table 3 shows the variation under 2 obstacles, and Fig. 7 represents it graphically.

The recorded observations and visual depictions from Table 4 and Fig. 8 justify the proposed equation that as distance increases, the RSSI strength decreases based

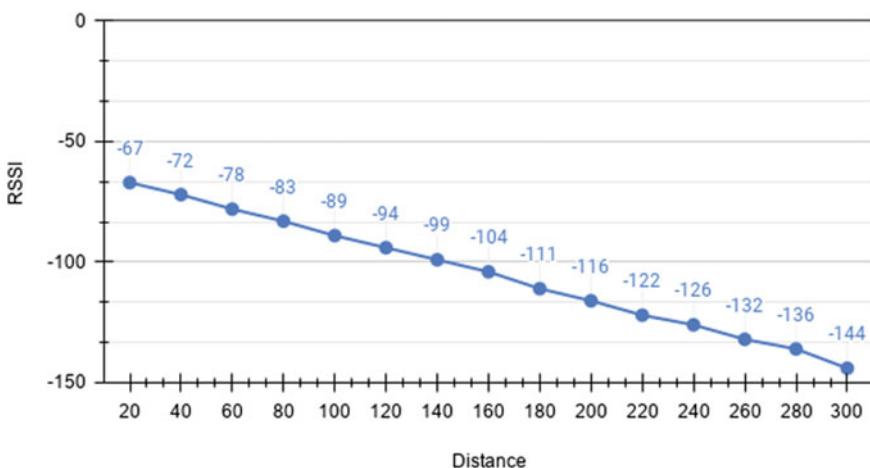
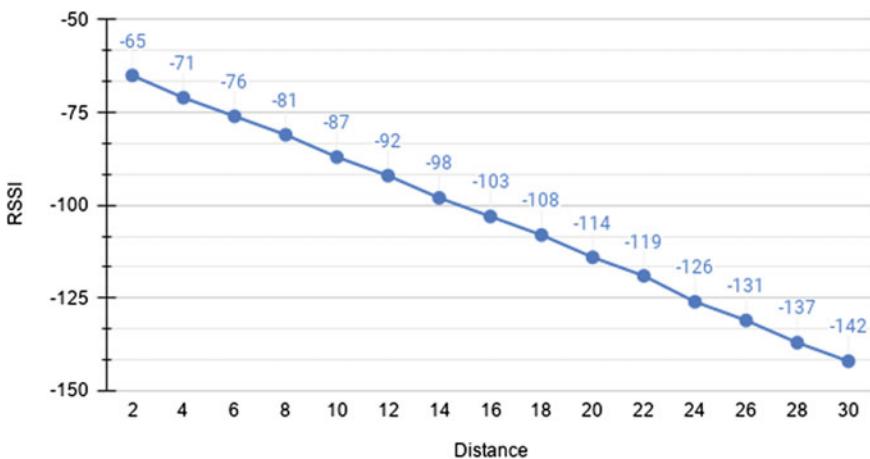
**Fig. 5** Graph of RSSI without obstacles with respect to the distance

Table 2 RSSI-distance under 1 obstacle

SL. No.	Phase-II		
	Distance in meters	No of obstacles	RSSI strength (dBm)
1	2.0	1	-65
2	4.0	1	-71
3	6.0	1	-76
4	8.0	1	-81
5	10.0	1	-87
6	12	1	-92
7	16	1	-103
8	20	1	-114
9	24	1	-126
10	28	1	-137
11	30	1	-142

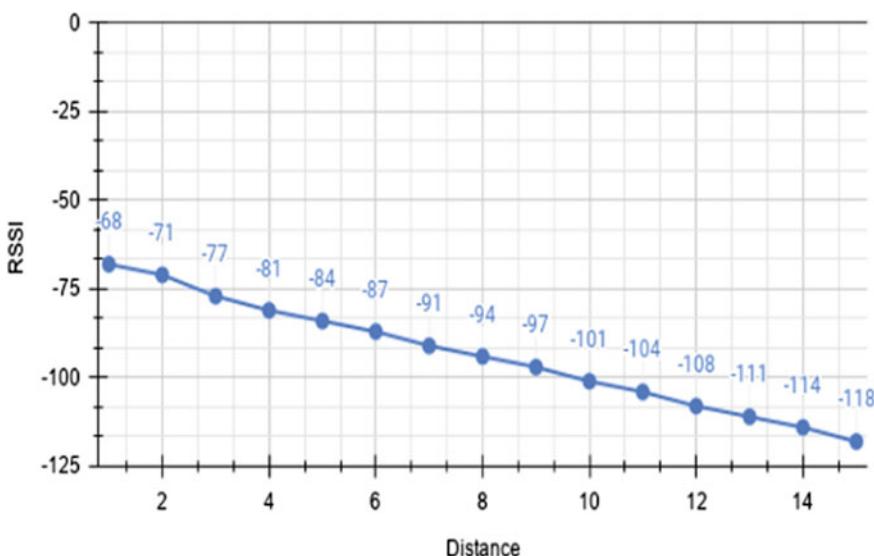
**Fig. 6** RSSI-distance of Table 2

on multiple factors and obstacles being one of them. The overall readings and plotted representation substantially relate with the suggested findings.

It is observed that the variation in distance steadily decreases the RSSI strength by a tier fraction. The graph depicts the behavioral pattern of the signals under the influence of obstacles and variations in the distances. This compromises with the proposed findings. However, other constraints such as variation in the RF signals and frequency due to shadowing effect could restrain the readings scalability.

Table 3 RSSI-distance under 2 obstacles

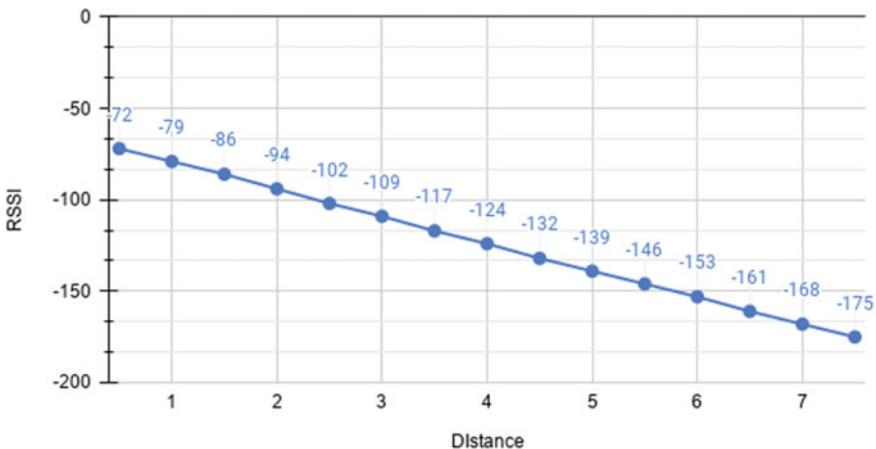
SL. No.	Phase-III		
	Distance in meters	No of obstacles	RSSI strength (dBm)
1	1.0	2	-68
2	2.0	2	-71
3	3.0	2	-77
4	4.0	2	-81
5	6.0	2	-87
6	8.0	2	-94
7	10.0	2	-101
8	12.0	2	-108
9	14.0	2	-114
10	15.0	2	-118

**Fig. 7** RSSI-distance of Table 3

Finally, this module can be taken to deploy the sensor nodes according to their range with respect to the distance, so when distance is less, node will reduce the range. When the distance between the sensor nodes are more than range of the node can be increase for the transmission. This helps to save the energy of the nodes instead of using the same power for all distances.

Table 4 RSSI-distance under 3 obstacles

SL. No.	Phase-IV		
	Distance in meters	No of obstacles	RSSI strength (dBm)
1	0.5	3	-72
2	1.0	3	-79
3	1.5	3	-86
4	2.0	3	-94
5	2.5	3	-102
6	3.0	3	-109
7	4.0	3	-124
8	5.0	3	-139
9	5.5	3	-146
10	6.0	3	-153

**Fig. 8** RSSI distance of Table 3

7 Conclusion

This research work explores on determining RSSI strength of a wireless fidelity hardware microchip under the influence of obstacles with respect to distance. However, there are multiple factors and variables that are above the standards of this research. The RF signals vary on the measured attenuation element, and this recalibrates the threshold distance. The optimal and stable source is discovered and the overall implementational parameters are evaluated. The network mode is performed in Arduino IDE using the compatible SoC packages. This system has a lot of room for further add-ons and enhancements which could result in a much more robust and superior RSSI strength indication level, lower latency time and automatic network shifting

seamlessly in real time. Major advantage of this work is that nodes can be deployed effectively to enhance the lifetime of the network based on the range of the node in WSN. This work can make the great change in the WSN world to enhance the lifetime of the node as well as network, because in these, the nodes will be operated based on the neighbor nodes distance.

References

1. Anand S, Sinha S (2019) 24X7 Lifeline Chip for Soldiers. In: International conference on recent trends in advanced computing, pp 573–581
2. Anusha PC, Anand S, Sinha S (2019) RSSI-based localization system in wireless sensor network. *Int J Eng Adv Technol (IJEAT)* 1765–1768, ISSN: 2249–8958
3. Manishankar S, Srinithi CR, Joseph D (2017) Comprehensive study of wireless networks qos parameters and comparing their performance based on real time scenario. International conference on innovations in information, embedded and communication systems (ICIIECS), pp 1–6
4. Nagarjun S, Anand S, Sinha S (2019) A research on the malicious node detection in wireless sensor network. *Int J Eng Adv Technol (IJEAT)* 8(5):1802–1805, ISSN: 2249–8958
5. Mukhopadhyay A, Anoop A, Manishankar S, Harshitha S (2020) Network Performance Testing: A Multi Scenario Contemplate. In: International conference on emerging trends in information technology and engineering (ic-ETITE), pp 1–7
6. Manoj Kumar T, Manishankar S, Ranjitha PR (2017) Data integration into cloud with efficient offloading of data from multiple nodes. In: International conference on intelligent computing and control (I2C2), pp 1–6
7. Manishankar S, Harshitha S, Mukhopadhyay A, Anoop A (2019) Technologies for network testing: a hybrid approach. In: Proceedings of the third international conference on computing methodologies and communication (ICCMC), pp 1–6
8. Prabhan AP, Anand S, Sinha S (2019) Identifying faulty nodes in wireless sensor network to enhance reliability. *Int J Recent Technol Eng (IJRTE)* 8(2):1727–1731, ISSN: 2277–3878
9. Barai S, Biswas D, Sau B (2017) Estimate distance measurement using node MCU ESP8266 based on RSSI technique. IEEE CAMA, Tsukuba, Japan, pp 170–173
10. Xue W, Qiu W, Hua X, Yu K (2017) Improved Wi-Fi RSSI measurement for indoor localization. Pukyong National University, Busan 48513, Korea, pp 1–22
11. Dobrilovic D, Stojanov Z, Stojanov J, Malic M (2020) Tools for modelling distance estimation based on RSSI. Novi Sad, Serbia, pp1–10
12. Sourya A, Dutta S, Chandra A, Prokes A, Kim M (2020) Find my car: simple RSS-based UWB localization algorithms for single and multiple transmitters. Niigata-shi, Japan, pp1–7
13. Lova Raju K, Chandrani V, Shahina Begum SK, Pravallika Devi M (2019) Home automation and security system with node MCU using internet of things. In: International conference on vision towards emerging trends in communication and networking (ViTECoN), pp 1–6
14. Patchava V, Kandal HB, Ravi babu P (2015) A smart home automation technique with raspberry Pi using IoT. In: International conference on smart sensors and systems (IC-SSS), pp 45–76
15. Anamul Haque SM, Kamruzzaman SM, Ashraful M (2010) A system for smart-home control of appliances based on time and speech interaction. In: Proceedings of the 4th international conference on electrical engineering, pp 1–5
16. Miao Y, Wu H, Zhang L (2018) The accurate location estimation of sensor node using received signal strength measurements in large-scale farmland. In: National engineering research center for information technology in agriculture, Beijing 100097, China, pp 1–11
17. Chowdhury SR, Rahman MM, Rapvez SA (2015) A multi-step approach for RSSi-based distance estimation using smartphones. *Networking Syst Secur (NSySS)* 7:104–345

18. Sundari Battula R, Khanna OS (2013) Geographic routing protocols for wireless sensor networks: a review. *Int J Eng Innov Technol* 2(12):1–4
19. Kurhe PS, Agrawal SS (2013) Real time tracking and health monitoring system of remote soldier using ARM 7. *Int J Eng Trends Technol* 4(3): 311–315, ISSN: 2231–5381
20. Willig A, Matheus K, Wolisz A (2005) Wireless technology in industrial networks. *Proc IEEE* 93(6):1130–1151
21. Mohapatra SK, Choudhury RR, Das P (2014) Proposed paper called the future directions in evolving Wi-Fi: technologies, applications and services. *Int J Next-Gene Networks (IJNGN)* 6(3):13–22
22. Al-Karaki JN, Kamal AE (2005) Routing techniques in wireless sensor networks: a survey. *IEEE Wireless Commun* 5(2):1–38
23. Mathew M, Anand S, Sinha S (2019) A hybrid algorithm for balancing load in wireless sensor network. *Int J Recent Technol Eng (IJRTE)* 8(2):1717–1721 ISSN: 2277–3878
24. Verma L, Fakharzadeh M (2013) Wifi on steroids: 802.11AC and 802.11AD. *Wireless Commun* 20(6): 30–35, IEEE
25. Anusha K, Yashaswini C, Segmentation M (2016) Retail mobile market using HMS algorithm. *Int J Electr Comput Eng (IJECE)* 6(4):1818–1827, ISSN: 2088–8708. <https://doi.org/10.11591/ijece.v6i4.10295> ρ 1818 Journal homepage: <http://iaesjournal.com/online/index.php/IJECE>
26. Rani BKS, Bhat S, Mukhopadhyay A (2017) A survey of wireless technologies and vertical handoff techniques from the perspective of telemedicine scenarios. In: 2017 international conference on communication and signal processing (ICCP), pp 1246–1251.<https://doi.org/10.1109/ICCP.2017.8286580>

Anatomy of Virtual Machine Placement Techniques in Cloud



Chayan Bhatt and Sunita Singhal

Abstract With consistent advancement in virtualization techniques, organizations are building up enhanced datacenters that are capable of maintaining impactful resource management, high-performance benchmarks, and controlled power consumption for eco-friendly computing. Virtual machine placement problem has a significant part in designing the datacenters. Placing virtual machines in the cloud can be very profitable and beneficial but it can be the cause of many new problems, if not performed properly. It comprises multiple complex relations and designing factors that directly affect the operating cost of datacenters. It bridges up the customers with cloud administrators to obtain preferences and SLA of both, to bring out an optimal solution. As optimization has given a boost to many businesses, an efficiently optimized VM placement technique has got a lot of potentials and can bring a drastic rise for many organizations running toward the cloud. A comprehensive study has been performed to bring out the important traits of different VM placement solutions by surveying cloud literature. By assessing the capabilities and objectives of different approaches and techniques, this paper presents an in-depth comparison, unveiling the drawbacks, and suggestions to improvise the methods in this direction.

Keywords Cloud computing · Virtual machine placement · Resource utilization · Quality of service · Performance

C. Bhatt (✉) · S. Singhal

Department of Computer Science and Engineering, School of Computing & Information Technology, Manipal University Jaipur, Jaipur, Rajasthan, India

S. Singhal

e-mail: Sunita.singhal@jaipur.manipal.edu

1 Introduction

Cloud computing empowers convenient and worldwide access to multiple virtual resources that are configurable and provisional, to fulfill the need for accommodating varying workload requirements. It has emerged out as a key for providing flexibility, portability, hardware independence, reliability, cost-savings, and many more, to facilitate dynamic scaling of the platform, infrastructure, and applications for its customers and users. Cloud service providers (CSPs) make these resources available to the users or customers as per the “pay-as-you-go” model [1].

To enhance the total cost of ownership (TCO), companies have adopted the flourishing features of IAAS that accomplish maximum resource utilization and cost-cutting [2]. To provide infrastructure services, the IaaS model works to deal with virtual machines (VM) requests as directed by the cloud users. But, due to the increasing demand for resources and variable workload, degradation in the application’s performance has been observed. To resolve this problem, rapid developments of virtualization techniques have been done. Virtualization entitles cost-saving cloud computing with flexibility and maximized CPU utilization. It illustrates the virtuality of computing over the real world. With the help of virtualization, the resources and assets of a physical system can be cloned, to create VM that works precisely the same and obtains similar results [1]. Thus, a VM is a software program that is built to run applications and operating systems, just like a physical machine (PM). Similar to PM, VM also requires memory, CPU, bandwidth, and many other resources to run given tasks [3]. They have a specified set of resource characteristics to emulate the behavior of a computing system and to frame traditional computing into a scalable one. VM can be advantageous to its clients as compared to a conventional physical computer. The mechanism of optimally mapping the VM among respective PM, as per their resource requirements, keeping various objectives and constraints of cloud into consideration, is termed as “virtual machine placement” (VMP) [4]. VM placement can be performed offline (static) or online (dynamic). In offline VMP, datacenters collect the inputs, and placement decisions are made considering various constraints, to satisfy requests of multiple users. In the online approach, after performing initial placement, the datacenter periodically gathers the data and makes decisions of VM placement shuffling when required. Provisioning of VM in the cloud is a highly complex task but holds a very important role in maintaining the cloud performance. Due to the unpredictable arrival pattern of VM instantiation requests and the large size of the datacenter along with its load, finding the optimal or near-optimal solution for VM placement emerges out as an NP-hard problem [5]. To achieve an efficient VM placement, few constraints need to be considered that form an integral part of placing VM to an appropriate PM [6]. Under capacity constraint, the total capacities of all VM resources, allocated to a particular physical machine, should not be more than the overall resource capacity of that PM. The placement constraint confines that a virtual machine should be not be allocated to more than one physical machine. Lastly, SLA constraint ensures that the parameters enlisted under QoS as per the user request should not be violated during optimization. VM placement mechanisms act

on targeted objectives and utilize different placement factors to improvise resource utilization, enhanced income, and minimized expenditure. This optimization can be tested and verified using different metrics and tools. This review renders a detailed study and evaluation of placing virtual machines in a cloud environment. The works of researchers have been reviewed in detail by classifying them as per their VM placement approach, to bring out their significant characteristics and traits along with their merits and demerits. Moreover, various placement factors have also been deeply focused on and investigated. By conducting this survey, current trends and emerging issues of placing virtual machines in a cloud environment have been broadly explored. Section 2 illuminates the challenges of the VM placement problem in the cloud. Section 3 differentiates different techniques for placing VM. Lastly, Sect. 4 provides concluding remarks.

2 Problem Description

VM has become popular due to its ample number of benefits. Deployment of VM can be seen as a simple task but it requires a lot of tedious background computations. To deliver guaranteed performance and throughput, placement of VM requires a planned and precise framework, to get placed in the server underlying the cloud. The nature of the workload has an immense effect on VM assignment decisions. Most of the workload today is either dynamic-centric or resource-centric.

- **Non-Resource-Centric**—involves a continuous change in resource requirements, thereby either put servers under distress of too much load or make them remain idle and underutilized.
- **Resource-Centric**—involves the utilization of one resource more than the other, thereby providing load to one resource too much as compared to the other (e.g., CPU intensive or I/O intensive).

Hence, hosting VM on dedicated servers under such workloads causes low resource utilization. Optimal placement of VM can bring up a significant boost in memory as well as CPU utilization. Sometimes, allocating VM on a PM causes performance degradation if a server is found to be loaded or underutilized. To handle such a situation, VM migration can be done. But, performing migration can be costly and may cause the application to slow down when the process of migration is underway. VM migration in the cloud needs to have refrained under the following cases-

- Migration cost exceeds the cost of running underutilized servers.
- Presence of sensitive VM data
- Prohibition by SLA for migration
- Rise of application downtime.

The placement of VM to its suitable PM has brought numerous remarkable researches that target various challenges and issues faced during VM placement.

Different strategies and mechanisms have been proposed with various goals and assumptions to handle complicated virtualization technologies and huge datacenters.

Figure 1 depicts various important aspects of VM management research that has been identified and presented in form of taxonomy. Organization follows different ideologies as per their needs and preferences. Some believe in cost-saving while some wants to stick on high performance, irrespective of its cost. Thus, a few adjustments

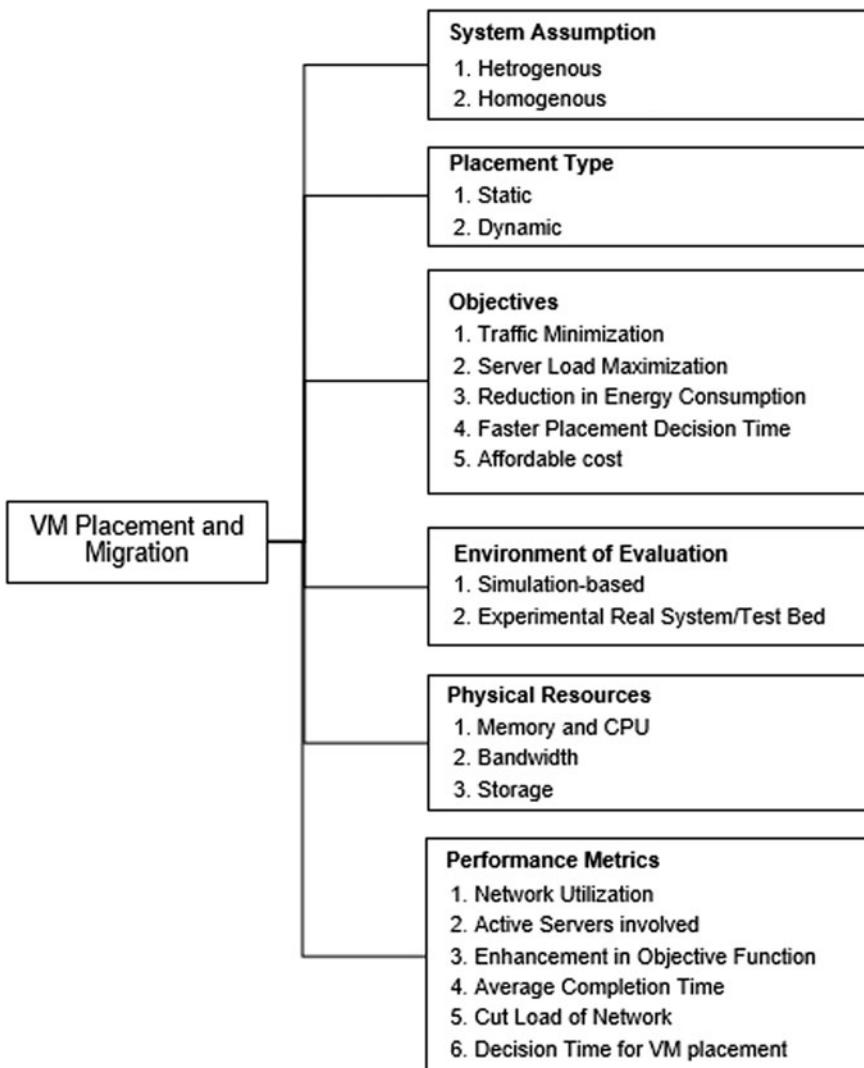


Fig. 1 Taxonomy of virtual machine placement

and refinement in strategies of VM placements can help in achieving preferential optimization goals. Economical providers will emphasize more on resource utilization, whereas high-performance providers will work on keeping the up time of running applications and will rule out the consideration of utilizing the resources more.

2.1 Categorization of VM Placement Algorithms

Mapping VM to their suitable PM as per their resource requests depicts VM placement. It plays a notable role in discovering the appropriate hardware for hosting different types of virtual machines. It can be seen as a periodic process rather than a one-time process. The placement approach can be classified as per two major goals, i.e.,

- **QoS-based**—maps VM to PM to attain the maximal quality of service.
- **Power-based**—works on VM-PM mapping to maintain energy efficiency of the datacenters.

But both goals contradict each other, making only one of the two, achievable at a time. A VM placement that desires to conserve power may not provide high-standard quality service to its clients. On the other hand, if the VM placement focuses on meeting all quality requirements to satisfy its clients, it may lead to high power consumption. VM placement approaches are used by cloud service providers as per their needs and preferences.

Analysis and comparison of VMP mechanism bring out the significant traits and merits of a particular VMP approach, along with their working limitations, thereby giving rise to future work and enhancement in this field. Typically, in placing VM onto an appropriate PM, the provider needs to fulfill the criteria for cost-efficient resource utilization and must assure that user requirements, as per the service-level agreement (SLA), are met. This creates two interrelated major problems that involve the selection of VM as per their resource characteristics, i.e., virtual machine configuration and how the VM will be mapped to their respective PM, i.e., virtual machine placement. Both the problems need to be tackled separately as they form an integral part of research in cloud computing.

2.2 Challenges of VMP

In a cloud datacenter, server consolidation has brought a noticeable benefit to datacenters by placing many virtual machines on a lesser number of physical servers, thereby minimizing resource wastage, energy, and improvising system availability, and performance [5]. But still, some problems have been observed in the IAAS environment that affects the efficiency of the cloud.

1. A tradeoff among performance, consumed energy, and resource utilization of VM allocated on a single server. By consolidating the VM to a few PM, PM resources have to be shared among the VM allocated to that PM, causing an increase in execution time, performance degradation, and energy consumption.
2. With continuous change in resource demands, frequent VM migration causes significantly longer execution time during consolidation and raises non-negligible overheads due to network traffic and PM load due to continuous updating of VM's CPU state, storage, main memory, and network connections.
3. It becomes quite complicated to predict energy consumed by an individual server. There is a different type of machines, some might be old, and some might be the new ones. PM having different power efficiency will execute an instruction with different energy consumption, and hence, better power efficiency PM would be favored for good VM placement.
4. With different PM load power characteristics, it is difficult to analyze the execution of a particular VM consumes more energy on PM 1 or PM 2.
5. Server consolidation may incur congestion of PM resources due to which the accommodated VM may find difficulty in obtaining the required number of resources, which thereafter results in application performance degradation of those VM, thereby violating service-level objectives. Hence, resources of data-center, i.e., storage, memory, and bandwidth, need to be equally focused as energy consumption, thereby satisfying user-specified quality of service (QoS) requirements.

Focusing on the challenging aspects of VM placement, many effectual mechanisms and strategies for server consolidation have been proposed that address notable placement factors and objectives to deal with frequent issues of variable workload and heterogeneous environment of virtual machines.

3 Related Work

As server virtualization brings improvisation in resource utilization and cost-cutting of cloud datacenters, a rise in substantial interests of practitioners and researchers in the virtual machine placement problem has been observed. Silva Filho et al. [7] have studied various VM placement and migration approaches and presented problem description, merits, and demerits of relevant research in detail. Past research works have emerged out many goal-based algorithms, emphasizing mainly load balancing, reduced SLA violation, service downtime, energy and cost-saving, network delays, and congestion. On considering the type of optimization techniques, provisioning of VM can be largely classified into the following categories.

3.1 Linear Programming

The placement of VM is framed into a linear programming problem. It reviews different constraints and requirements through a linear relationship from practical applications. Rym et al. [8] proposed a two objective integer linear programming model that aimed maximization of server's resource usage while preserving its power efficiency. Configurations of two heterogeneous datacenter were examined to analyze their impact on the performance and efficiency of the proposed model.

3.2 Constraint Programming

This optimization technique is based on a constraint satisfaction problem. Hélène et al. [9] designed a VM placement approach for a particular case of hybrid cloud. Different placement objective functions and constraints were proposed to automate software deployment on hybrid clouds. Zhenpeng et al. [10] designed a location-constrained virtual machine placement algorithm that worked on placing VM onto PM on the basis of location constraints such as the selected PM shares at least one link among each other. LCVP performed well in terms of feasibility, computation time, and blocking probability.

3.3 Heuristic Bin Packing

The VM placement problem is framed as a vector bin packing. VM is considered as small items that are required to be tightly packed in a lesser number of bins, and those bins are assumed as PM. Various heuristic techniques have been proposed for this problem optimization. Packing algorithm, being the most popular and simpler one, comes up with first fit, best fit, worst fit, etc., for efficiently and optimally placing the multiple VM [11]. These algorithms assure to deliver at most 70% optimum results. Moreover, the approximation ratio can be improvised by using first fit decreasing and best-fit decreasing, in which sorting of items are being done in decreasing order of their weights. A modified best-fit decreasing algorithm proposed by Beloglazov and Buyya [12] performed sorting of available physical servers in decreasing order of their resource utilization and then placed VM to its suitable PM. Through this mechanism, power consumption was minimized and dynamic VM consolidation was achieved. Wang et al. [6] proposed the energy-efficient and QoS-aware VM placement (EQVMP) technique that involved three main modules, i.e., hop reduction, energy-saving, and load balancing. Max-min multi-dimensional stochastic bin packing along with best-fit decreasing (BFD) has been used for placing VM such that the number of servers of the datacenters is reduced. Dong et al. [13] have portrayed a significant relationship between energy and performance of VM. Salient features accountable

for energy consumption reduction were examined as a multi-dimensional bin packing problem, thereby uncovering many research issues. Mahyar et al. [14] developed a dynamic mapping algorithm that used simulated annealing for mapping VM onto a small set of PM to satisfy customer satisfaction and prevent SLA violations. Saurav et al. [2] constructed a VM placement problem to optimize two objective functions problem and used simulated annealing to produce optimal results. The technique aimed and provided a reduction in power budget and improvised revenue. In most of the VM placement approaches, usually, CPU and memory resources are mainly considered but networking issues have been neglected. Limited network bandwidth plays an important issue in the allocation of VM. Guo et al. [15] have built a mechanism involving a shadow routing approach for adaptively mapping VM to PM, to achieve high resource utilization, energy, and cost-saving and performing auto-scaling of VM, handling user requests, and dynamic applications. But, due to the high complexity of cloud datacenters, a more enhanced approach is needed for VM placement. Son et al. [16] proposed priority-aware VM allocation (PAVA) that works on information derived from network topology for allocation of VM on the host, as per the type of applications.

3.4 Biology-Based Optimization

Many new nature-inspired metaheuristic algorithms have been designed in the last few years to effectively place VM while maintaining the load balance in a dynamic workload environment. Mirjalili et al. [17] designed a multi-objective grey wolf algorithm with the addition of two new components. A fixed-sized archive was used for saving the non-dominated solutions. The algorithm follows the leadership ranking of grey wolves comprising of α , β , δ , and ω wolves. The algorithm has been tested on different benchmark functions. But the algorithm does not provide optimal results for problems having more than four objectives. Riahi et al. [4] came out with a multi-objective genetic algorithm that minimizes resource wastage and utilization rate of active servers in the cloud. It also works on the reduction of operational costs. To verify the adaptivity of the proposed work, Bernoulli simulations have been performed. However, with the high growth rate of data, it is necessary to solve the aforementioned problem but this work lacks handling of big data efficiently. Zahoor et al. [18] designed a smart grid mechanism based on cloud fog by hybridizing artificial bee algorithm with ant colony optimization for managing cloud resources efficiently. The proposed algorithm works for efficiently balancing the cloud datacenters load and enhancement of VM' time to respond and process. But the mechanisms involved in balancing the load fail to cope with the complexity of datacenters in the cloud. Liu et al. [10] build up an approach to place VM effectually. With the inspiring performance of ant colony optimization, the proposed mechanism focuses on minimizing the PM' utilization rate and efficiently saving energies and cost of the datacenters. Operational cost and number of active servers were significantly reduced. But resource management needs to be enhanced in the initial phase

of VM placement. Wu et al. [19] build up an enhanced modified GWO algorithm that holds characteristics of grey wolf optimization along with levy flight, since the simple grey wolf algorithm experiences local minima problems. Therefore, to work on this hurdle, the modified grey wolf optimization aims to solve real-time and global optimization problems. The approach has been tested under complex functions to check its compatibility. The proposed algorithm works well for the single-objective problem but, to resolve problems of global optimization, optimization of more than one objective function is required. Jensi et al. [20] designed a simple particle swarm optimization along with levy flight, to deal with global optimization problems. Tracing the significant drawbacks of particle swarm optimization like premature convergence rate and local optima trapping, the authors designed this framework to face the above-mentioned issues. The proposed work has been verified by twenty-one benchmark functions. However, the proposed algorithm is capable of solving only single-objective problems. Guerrero et al. [21] developed a multi-objective non-dominated sorting genetic algorithm. Virtualized Hadoop was used for the minimization of power consumption and resource wastage of PM. A 1.9% reduction was observed in total power consumption. However, the rapid increase of cloud users cannot be optimally handled by the proposed method. Md Hasanul Ferdaus et al. [22] presented a metaheuristic ACO approach, for effectively placing VM to PM. According to historical experience, the VM is grouped together and then artificial ants are used to search the fittest place of the host. In every search iteration, the infeasible solution was revised to reduce the convergence time. As the iteration number grows, the solution space was reduced. The infeasible solution was turned feasible exchanging process and migrating VM from overloaded PM, thereby reducing the convergence time. Sivaraman et al. [23] have solved VM placement problem using intelligent water drops optimization technique. This metaheuristic is nature-based and mimics natural flow of river in sea. The strength and properties of the water drops find the optimal solution of the problem, i.e., minimization of resource wastage and power consumption. Dilip et al. [24] designed a bi-objective VM placement technique using combination of genetic and particle swarm optimizations. The proposed work focused on reduction of active servers and minimization of resource wastage. VM migrations could be used for enhancement. Madhusudan et al. [25] proposed an energy and fault-aware VM placement technique using neural network to minimize power consumption and execution time of datacenter. Genetic algorithm was used to train artificial neural network. It worked well to improvise energy consumption and execution time but SLA violations became the loophole for the approach. On the basis of optimization techniques used to solve the VM placement problem, past proposed work has been categorized into four main optimization approaches as shown in Table 1 and a summary of past research has been presented in Table 2 that describes the observed objectives, constraints, advantages, and limitations of various VM placement techniques.

Table 1 Categorization of VM placement techniques

Deterministic	Linear programming [8] Constraint programming [9, 10]
Heuristic	Modified best-fit decreasing [12] Energy-efficient and quality of service-aware VM placement [6] Shadow routing [15] Next fit allocation [26] Priority-aware VM allocation using FFD [16] Adaptive threshold energy and SLA-aware algorithm [27] Learning-based approach [28]
Metaheuristic	Ant colony optimization [29] Profit-based simulated annealing [2] Multi-objective grey wolf optimization [1] Enhanced modified grey wolf optimization [13] Levy-based particle swarm optimization [20] Binary particle swarm optimization [30] SLA-based simulated annealing [14] Energy-efficient ant colony system [29] Multi-objective genetic algorithm [21] Genetic algorithm [4] Intelligent water drops (IWD) [23] Resource-aware genetic algorithm [31] Enhanced Firefly Algorithm [32]
Hybrid	Hybrid of ant colony and artificial bee colony [18] Best-fit VM placement policy using machine learning [33] Hybrid of genetic algorithm and particle swarm optimization [24] Genetic algorithm with machine learning [25]

4 Performance Metrics

The needs and requirements of the cloud user and provider play an important role in solving VM placement problems. Every VM placement approach performs well under specified conditions and environment. The performance of VM placement mechanisms depends on parameters and constraints, specified to these algorithms and is evaluated at system and application levels by using performance metrics.

4.1 Power and Energy Consumption

Enormous energy consumption by datacenters has become a great hindrance for service providers in cloud computing. To cope with power-related issues, power-aware VM placement schemes have been formulated to optimize energy consumption and increase the server's power efficiency of cloud datacenters [12]. The following factors need to be reviewed while determining the power and energy consumption of a VM placement.

Table 2 Summary of research on VM placement algorithms

Year	Technique	Constraints	Objective	Advantages	Limitation
2014	Energy-efficient and quality of service-aware VM placement [6]	CPU utilization, No. of servers, congestion	Optimal energy-saving, VM Allocation leveraging QoS	Minimization of datacenter servers	More VM migration cost
2014	Ant colony optimization [29]	CPU utilization, No. of servers	Load balancing of multi-dimensional resources	Effective migration of VM reduced the convergence time	More computation time
2014	SLA-based simulated annealing [14]	VM requests, SLA	To manage dynamic nature of VM requests from cloud customers	Demand satisfaction achieved keeping considered SLA	Execution time need to be lowered down
2016	Multi-objective grey wolf optimization [1]	CPU utilization, No. of servers	Aims to find solution for multi-objective problems	Convergence rate reduced to provide optimal solution	Not worthy for more than four objectives
2016	Enhanced modified Grey Wolf Optimization [13]	CPU utilization, No. of servers	To upgrade the performance of simple GWO algorithm with levy distribution	Real-time and global optimization problems can be solved	More computation time
2016	Levy-based particle swarm optimization [20]	CPU utilization, No. of servers, execution time	Mechanism to deal with premature convergence problem	Problems of Premature convergence and trapping optimized	Solves only single objective problems
2017	Binary PSO [30]	Resource Utilization, power consumption	To enhance resource usage and save power	Power consumption reduced as idle servers were removed	Not applicable for dynamic placement of VM
2017	Profit-based simulated annealing [2]	Revenue, power budget	To place maximum VM with less power consumption	Worked well to achieve less no. of servers, energy used, and low execution time	Need to enhanced for dynamic workload
2018	Energy-Efficient Ant Colony System [29]	No. of servers	PM utilization rate minimized to save energies and cost	Reduction in usage of PM and cost	Resources are not efficiently managed

(continued)

Table 2 (continued)

Year	Technique	Constraints	Objective	Advantages	Limitation
2018	Multi-objective genetic algorithm [21]	No. of servers, resource utilization	Effectual placement of VM	Minimization of active servers	Overhead caused due to large searching spaces
2018	Shadow routing [15]	Application requests, No. of servers	Auto-scaling and rational placement of Virtual Machines to minimum number of servers	Reduction in cost and consumption of energy	Applicable only for private clouds
2018	Cloud fog-based smart grid [18]	Resource utilization	For managing the cloud resources efficiently	Processing time and response of the VM enhanced	Complex cloud datacenters cannot be handled
2018	Genetic Algorithm [4]	No. of servers, Resource utilization	Allocation of VM along with template selection	Reduction in power consumption and minimization resource wastage	Lacks efficiency for rapidly growing users
2019	Priority aware VM allocation using FFD and dynamic flow scheduling [16]	Bandwidth, Network congestion	Transferring High Priority data in high traffic, bandwidth allocation	Reduction in network congestion due to other tenants	Need to support multi-degree priorities

1. The state of CPU, i.e., being idle, busy, over-utilized
2. The power consumption of various server components, i.e., CPU, memory, storage, network, etc.
3. Determining the states of PM (ready, running, sleep, off) and their power consumption
4. The minimum distance among VM
5. Power used by the datacenter involving.
 - Cooling energy
 - Peak power
 - The energy of the base server
 - The energy of dynamic server.

4.2 Network

Both inter- and intra-network traffic of datacenter give a great impact on the performance, SLAs, and the profits of the cloud providers [16]. Network-aware VM placement has emerged out with remarkable importance in the cloud due to the rise of communication applications. They aim to evenly utilize the links in datacenter networks to avoid congestion and address other traffic-related problems and issues experienced during VM placement. Following factors are considered to build effective network-aware VM placement mechanisms.

1. Minimizing the total data transfer time among the VM
2. Shortening the average distance between the VM to reduce network traffic and enhance the performance of communication by slowing down the network congestion
3. Slow down the traffic incur in the federated clouds and among the physical machines.

4.3 Resource Usage

Virtual resource management holds an important issue of research in cloud computing. Execution of different sorts of application requires VM to obtain and access variety of resources, and therefore resource-aware VM placement mechanisms are required and provide optimal placement decision by considering hardware resource requirements of each VM [31, 34]. An efficient resource-aware placement solution aims to achieve maximization of overall resource utilization, thereby placing VM on the PM optimally and efficiently.

4.4 Cost Estimation

Minimizing the overall operational cost comes up as a critical task. The tradeoff between cost and response delay makes this task more complex. Thus, cost-aware VM placement schemes are needed for diminishing operational and maintenance costings of datacenters for the cloud providers, taking consideration of SLAs and the quality of service of cloud services simultaneously [14].

The following factors play an important role in formulating optimal cost-oriented VM placement mechanism-

1. Cost of VM during their creation in the data centers
2. The PM usage cost for specific period of time
3. Cost of the datacenter
4. Minimum distance between clients and VM
5. Cooling system cost.

Table 3 Categorization of influential parameters and their Source of Measure

Power	<ul style="list-style-type: none"> • No. of active servers • Power efficiency • Power consumption • Power usage effectiveness • Thermal dissipation • Temperature
Cost	<ul style="list-style-type: none"> • VM migration cost • VM and PM cost • Penalties cost • Cooling cost • Electricity cost • Cloud cost • Cost efficiency • Communication cost
Network	<ul style="list-style-type: none"> • Inter/Intra-VM Traffic • Delay • Service throughput • Execution time • Distance between VM • Link-load ratio
Resource	<ul style="list-style-type: none"> • Resource usage • Resource wastage • Average resource utilization • Load balancing • Network resource utilization
QoS	<ul style="list-style-type: none"> • SLA violation • Scalability • Reliability • Fault-tolerance • Availability • Scheduling time • Fairness • Success ability

On performing the survey of different existing VM placement approaches, influential parameters and their source of measure have been identified and described in Table 3. These parameters affect the performance of a datacenter and thus are very important for VM placement. On considering the performance metrics involving power, network, cost, and QoS, a broad comparison among the research works has been made as shown in Table 4.

5 Conclusion

An efficient VM placement has significantly helped in provisioning the VM with their suitable host. Variants of bin packing like FFD, MBFD have been more cautious

Table 4 Analytical summary of VM placement algorithms

Year	Technique	Power consumption	Network traffic	Cost estimation	QoS
2014	Energy-efficient and Quality of Service-aware VM placement [6]	Yes	No	No	Yes
2014	Ant colony optimization [29]	Yes	No	No	No
2014	SLA-based simulated annealing [14]	No	No	No	Yes
2016	Multi-objective grey wolf optimization [21]	Yes	No	No	No
2016	Enhanced modified grey wolf optimization [13]	Yes	No	No	No
2016	Levy-based particle swarm optimization [20]	Yes	No	No	Yes
2017	Binary PSO [30]	No	No	No	Yes
2017	Profit-based simulated annealing [2]	Yes	No	Yes	No
2017	Linear programming [8]	Yes	No	No	Yes
2018	Energy-efficient ant colony system [29]	Yes	No	Yes	No
2018	Multi-objective genetic algorithm [21]	Yes	No	Yes	No
2018	Shadow routing [15]	No	Yes	No	No
2018	Cloud fog-based smart grid [18]	No	No	No	Yes
2018	Genetic algorithm [4]	Yes	No	No	Yes
2019	Priority-aware VM allocation using FFD and dynamic flow scheduling [16]	Yes	Yes	No	No
2020	Next fit allocation [26]	Yes	No	No	Yes
2020	Intelligent water drops [23]	Yes	No	Yes	No
2020	Resource-aware genetic algorithm [31]	Yes	No	No	No

of power consumption of the datacenters which holds one of the biggest reasons for increasing the datacenters' operational costs. But, violating the guidelines of SLA has brought a loophole which needs to be improvised. Bio-inspired optimization techniques like PSO, ACO, grey wolf have proven to efficiently broaden the search space area of VM placement but due to their complexity, there has been rise in overhead cost and execution time. Many approaches have proposed an efficient resource management but due to continuous resource balancing, excess migration has contributed rise in monetary cost. Therefore, it is inappropriate to suggest or state the best approach among all, as every placement technique works on specific target and has diverse migration techniques and prominent resources. These approaches hold different areas of improvement due to tradeoffs between influential parameters like VM performance and energy consumption or between response delay and cost, and this can be considered for future enhancement, to deal with variable workload, and continuous change of applications demands. Moreover, one can contribute toward eco-friendly IT infrastructures by opting "greener cloud" by implementing hybrid techniques out of surveyed approaches.

References

1. Fatima A (2019) An enhanced multi-objective grey wolf optimization for virtual machine placement in cloud data centers. *Electronics* 8:1–32
2. Addya SK, Turuk AK, Sahoo B, Sarkar M, Bishwash SK (2017) Simulated annealing based VM placement strategy to maximize profit of cloud service providers. *Eng Sci Technol Int J* 20:1249–1259
3. Lin MH, Tsai JF, Hu YC, Su TH (2018) Optimal allocation of virtual machines in cloud computing. *Symmetry* 10:1–9
4. Riahi M, Krichen S (2018) A multi-objective decision support framework for virtual machine placement in cloud data centers: a real case study. *J Supercomputer* 74:2984–3015
5. Mann ZA (2015) Allocation of virtual machines in cloud data centers—a survey of problem models and optimization algorithms. *ACM Comput Surv* 48:1–34
6. Wang SH, Huang PPW, Wen CHP, Wang LC (2014) Energy-efficient and QoS aware virtual machine placement for software-defined data center network. In: Proceedings of the IEEE international conference on information networking (ICOIN), pp 220–225
7. Silva Filho MC, Monteiro CC, Inácio PRM, Freire MM (2018) Approaches for optimizing virtual machine placement and migration in cloud environments: a survey. *J Parallel Distrib Comput* 111:222–250
8. Regaieg R, Koubâa M, Osei-Opoku E, Aguilis T (2018) A two objective linear programming model for VM placement in heterogeneous datacenters. International symposium on ubiquitous networking, vol 11277, pp 167–178
9. Couillon H, Le Louet G, Menaud J-M (2017) Virtual machine placement for hybrid cloud using constraint programming. In: International conference on parallel and distributed systems, pp 326–333
10. Liu Z, Lu J, Su N, Zhang B, Li X (2020) Location-constrained virtual machine placement (LCVP) algorithm. *Sci Program* 2020:1–8
11. Kumaraswamy S (2019) Bin packing algorithms for virtual machine placement in cloud computing: a review. *Int J Electr Comput Eng* 9:512–524

12. Beloglazov A, Buyya R (2010) Energy efficient allocation of virtual machines in cloud data centers. In: Proceedings of the 10th IEEE/acm international conference on cluster, cloud and grid computing, pp 826–831
13. Dong J, Wang H, Cheng S (2015) Energy-performance Tradeoffs in IaaS Cloud with virtual machine scheduling. *Communications* 12:155–166
14. Amini M, Safavi NS (2014) A dynamic SLA aware solution for IaaS cloud placement problem using simulated annealing. *Int J Comput Sci Inform Technol* 6:52–57
15. Guo Y, Stolyar A, Walid A (2018) Online VM auto-scaling algorithms for application hosting in a cloud. *IEEE Trans Cloud Comput* 8:1–11
16. Son J, Buyya R (2019) Priority aware VM allocation and network bandwidth provisioning in software defined networking (SDN)-enabled clouds. *IEEE Trans Sustain Comput* 4:17–28
17. Mirjalili S, Saremi S, Mirjalili SM, De Coelho LS (2016) Multi-objective Grey wolf optimizer: a novel algorithm for multi-criterion optimization. *Expert Syst Appl* 47:106–119
18. Zahoor S, Javaid S, Javaid N, Ashraf M, Ishmanov F, Afzal M (2018) Cloud fog based smart grid model for efficient resource management sustainability. *Sustainability* 10:1–21
19. Wu Q, Ishikawa F, Zhu Q, Xia Y (2019) Energy and migration cost-aware dynamic virtual machine consolidation in heterogeneous cloud datacenters. *IEEE Trans Serv Comput* 12:550–563
20. Jensi R, Jiji GW (2016) An enhanced particle swarm optimization with levy flight for global optimization. *Appl Soft Comput* 43:248–261
21. Guerrero C, Lera I, Bermejo B, Juiz C (2018) Multi-objective optimization for virtual machine allocation and replica placement in virtualized Hadoop. *IEEE Tran Parallel Distrib Syst* 9:2568–2581
22. Ferdaus MH, Murshed M, Calheiros RN, Buyya R (2015) Network-aware virtual machine placement and migration in cloud data centers. *Emerg Res Cloud Distrib Comput Syst* 42–91
23. Eswaran S, Dominic D, Natarajan J, Honnavalli PB (2020) Augmented intelligent water drops optimization model for virtual machine placement in cloud environment. *IET Networks* 9:215–222
24. Kumar D, Mandal T (2017) Bi-objective virtual machine placement using hybrid of genetic algorithm and particle swarm optimization in cloud data center. *Int J Appl Eng Res* 12:12044–12051
25. Madhusudhan, Kumar S (2020) Energy and fault aware virtual machine allocation using machine learning for cloud infrastructure. *Int J Adv Sci Technol* 29:2472–2482
26. Sengupta J, Singh P, Suri PK (2020) Energy aware next fit allocation approach for placement of VMs in cloud computing environment. *Adv Inform Commun* 1130:436–453
27. Zhou Z, Zhigang H, Lin K (2016) Virtual machine placement algorithm for both energy-awareness and SLA violation reduction in cloud data centers. Hindawi Publishing Corporation *Sci Program* 1:1–11
28. Ghobaei-Arani M, Rahamanian AA, Shamsi M, Rasouli-Kenari A (2018) A learning-based approach for virtual machine placement in cloud data centers. *Int J Commun Syst* 32:1–18
29. Liu XF, Zhan Z, Deng JD, Li Y, GU T, Zhang J (2018) An energy-efficient ant colony system for virtual machine placement in cloud computing. *IEEE Trans Evol Comput* 22:113–128
30. Tripathi A, Pathak I, Vidyarthi DP (2018) Energy efficient VM placement for effective resource utilization using modified binary PSO. *Comput Commun Networks Syst Comput* J 61:832–846
31. Kumar J, Singh AK, Mohan A (2021) Resource-efficient load-balancing framework for cloud data center networks. *ETRI J* 43:53–63
32. Barlaskara E, Singha YJ, Issac B (2016) Energy-efficient virtual machine placement using enhanced firefly algorithm. *Multiagent GridSystems Int J* 12:167–198
33. Rawas S, Zekri A, Zaart AE (2018) Power and cost-aware virtual machine placement in geo-distributed data centers. In: Proceedings of the 8th international conference on cloud computing and services science, pp 112–123
34. Gupta MK, Jain A, Amgoth T (2018) Power and resource-aware virtual machine placement for IaaS cloud. *Sustain Comput Inform Syst* 19:52–60

35. Al-Moalmi A, Luo J, Salah A, Li K (2019) Optimal virtual machine placement based on grey wolf optimization. *Electronics* 8:1–32
36. Wang H, Tianfield H (2018) Energy-aware dynamic virtual machine consolidation for cloud datacenters. *IEEE Access* 6:15259–15273

Effectual Attendance Application for Remote Education During Era of COVID-19



Mohitsinh Parmar , Shailesh Khant , and Atul Patel

Abstract Coronavirus pandemic COVID-19 has affected many sectors like business sector and education sector. Specifically in education sector, there are lots of changes in methods of teaching as well as learning. Online attendance calculation is the major issue for teacher because many different learning management system (LMS) tools are used. Many of them are providing the report of attendance and many are not. An android-based application for mobile user is proposed and designed for managing the attendance records. During online lectures, QR code is appeared and it is scanned by the application users to mark their attendance. A small effort is made here for marking the attendance which can reduce a big burden of the presenter. A report can be generated at client side and owner side for analysis of the attendance. Enhancements are suggested at the end for improving the application by additional report generation where duration of lecture or class attended can be maintained.

Keywords Mobile application · Android · QR code · Education · Attendance

1 Introduction

During this COVID-19 era, all educational institutes suffer losses in terms of quality teaching. Various online learning management tools are available which can keep the track of attendance. The problem that arises with system are tracking students' performance, students' attendance, students' concentration, etc. Many such problems can be addressed and managed by many researchers nowadays.

M. Parmar · S. Khant · A. Patel
Smt. ChandabenMohanbhai Patel Institute of Computer Applications, Charusat University,
Changa, India
e-mail: mohitsinhparmar.mca@charusat.ac.in

S. Khant
e-mail: shaileshkhant.mca@charusat.ac.in

A. Patel
e-mail: atulpatel.mca@charusat.ac.in

One of the attempt is made here to solve the problem of students' attendance. This system is based on generating QR code by faculties using smart attendance app and scanning the same QR code by students using the same application. The idea which is implemented here is based on generating QR code and paste it on specific slides or pages of the presentation during ongoing lecture. Students studying from home can scan this code, and report will be sent to teacher.

Various activities diagrams are presented to support the base of this system. Android studio is used to develop the mobile application as a frontend tool while on backend side, MySQL and Java are used. PHP web service is also used for the development of the system. A development can also be suggested for developing same application for IOS. Furthermore, various security aspects can also be imparted with this application but it can be implemented in next version of the application.

2 Proposed Model

See Fig. 1.

Keeping a record of presence in any sector is vital, and it should be maintained in a proper place. The most common way to keep track of students' attendance is to let them sign the attendance list. However, in this COVID-19 situation, to take student's signature on paper is not feasible. So here in the proposed system, attendance can be taken through remote mode study.

Another problem is to track the total time spent by the student in specific class because many students cannot able to join the classes for continuous time. So a minimum period must be required for marking their presence. It can be imposed by scanning the students' attendance during starting and ending of the class or lecture.

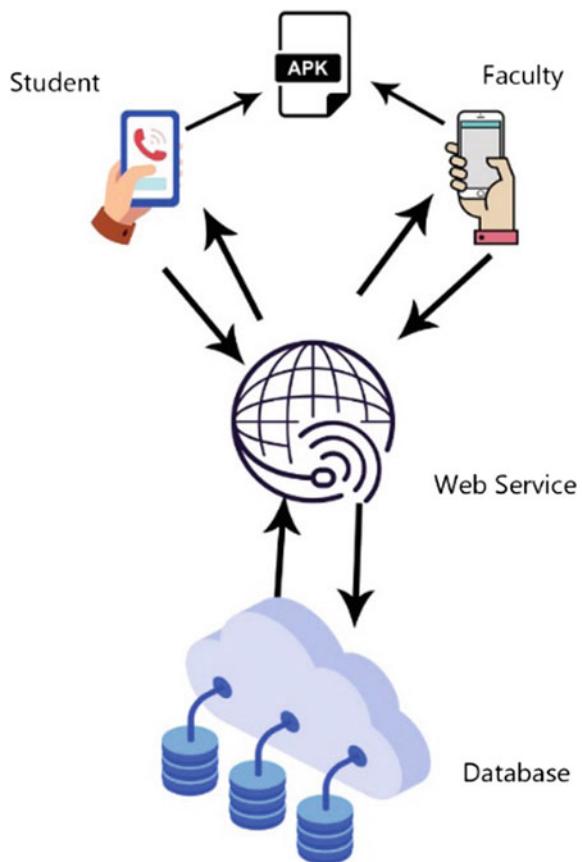
The proposed system architecture shows that how the application mechanism works. There will be two users one will be faculty, and the second user will be the student. Both can use the android application for maintaining attendance. There will be different roles which both the users have to perform. In that, the faculty role is to generate a QR code using the application by a particular subject. The student has to scan the QR code which will be going to share by faculty during the lecture time to make their present for that lecture.

For the communication between application and database, the web service is used. The request is sent by the post method. Each request is sent by only an authorized user with a unique id number. For fast performance, JSON is used as the data format for web service response.

Figure 2 displays the use case diagram of the system. If we look into a diagram them we will find the use cases (Login/Registration, Generate QR Code, Scan QR Code, attendance report, Share Application, and Profile). So fundamentally the diagram covers the functional requirements of the system. It also captures the actors of the system and the relationship among the use cases and actors.

Figure 3 shows the activity diagram for faculty. It describes how activates are coordinated to the service. In Fig. 3, the first notation shows the start point from

Fig. 1 Proposed system architecture



where the system starts. Next, the user has to login using credentials, and if a user is not registered, then the user must have to register inside the system by providing valid details. In the next activity, the wall is displayed to the user after login in inside the system. The next activity will depend on the user's choice whether the user wants to generate a QR code, shows the attendance report, or shows the profile.

Figure 4 shows the activity diagram for students. It describes how activates are coordinated to the service. In Fig. 4, the first notation shows the start point from where the system starts. Next, the user has to login using credentials, and if a user is not registered, then the user must have to register inside the system by providing valid details. In the next activity, the wall is displayed to the user after login in inside the system. The next activity will depend on the user's choice whether the user wants to scan a QR code, shows the attendance report, or shows the profile.

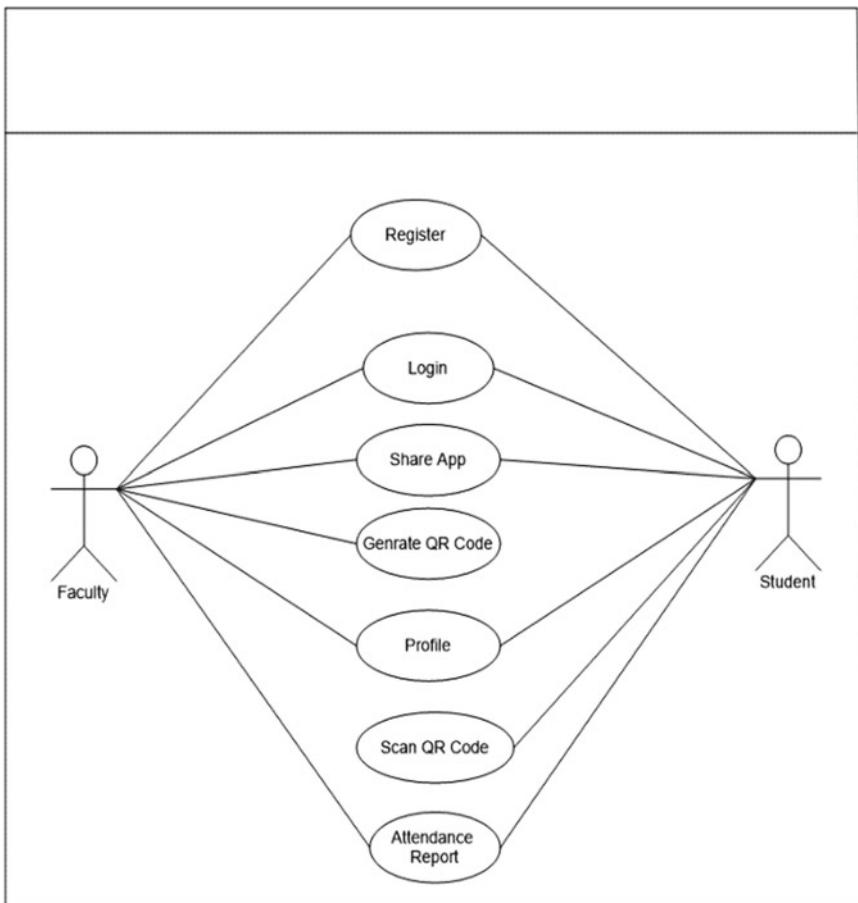


Fig. 2 Use case diagram

3 Words from Literature

Jun Lio (2016) had developed an attendance system using a mobile device and web application. A teacher can prepare the student list and upload the data of registration. Students can mark the presence during lectures using their registered application on mobile using selfie or signature. Various types of existing methods for marking attendance are also discussed. A combination of mobile device and web server is used [1].

Sadat Duraki et al. (2019) had developed a mobile application to take the attendance in wireless environment. To maintain the attendance records efficiently, they have developed a system which is based on fingerprint sensors and Zigbee modules.

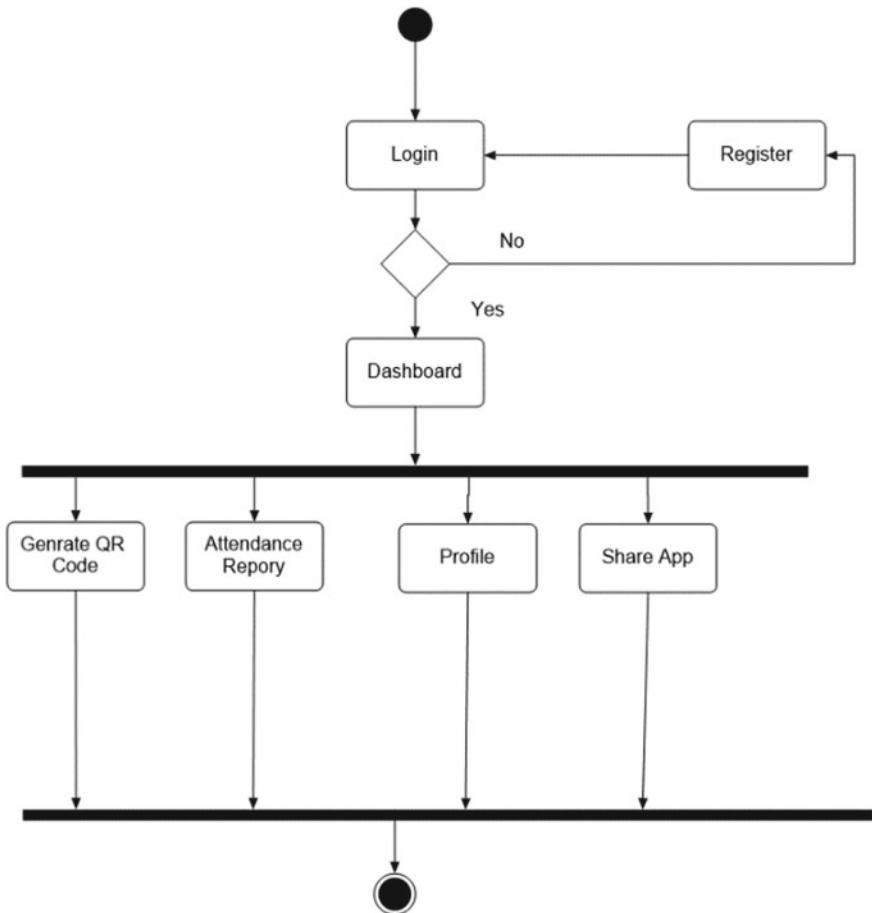


Fig. 3 Activity diagram [Faculty]

Attendance was taken using fingerprint sensor, and through wireless Zigbee module, it is uploaded into server database [2].

Milon Islam et al. (2017) had developed a smart phone-based attendance system for tracking the students in the class. Teacher is an administrator for attendance who can create attendance records and connect it to smart phone through MySQL database. This system can take attendance and generate report of attendance in terms of percentage. Once the attendance is updated and email or message is sent to students from the system [3].

Refik Samet et al. (2017) had worked on classroom attendance system based on face recognition. It uses the concepts and techniques of face recognition like eigenfaces, local binary pattern, and fisher faces. Three different versions of mobile applications are developed. They are for students, parents, and teachers. System architecture is explained in terms of application layer, communication layer, and

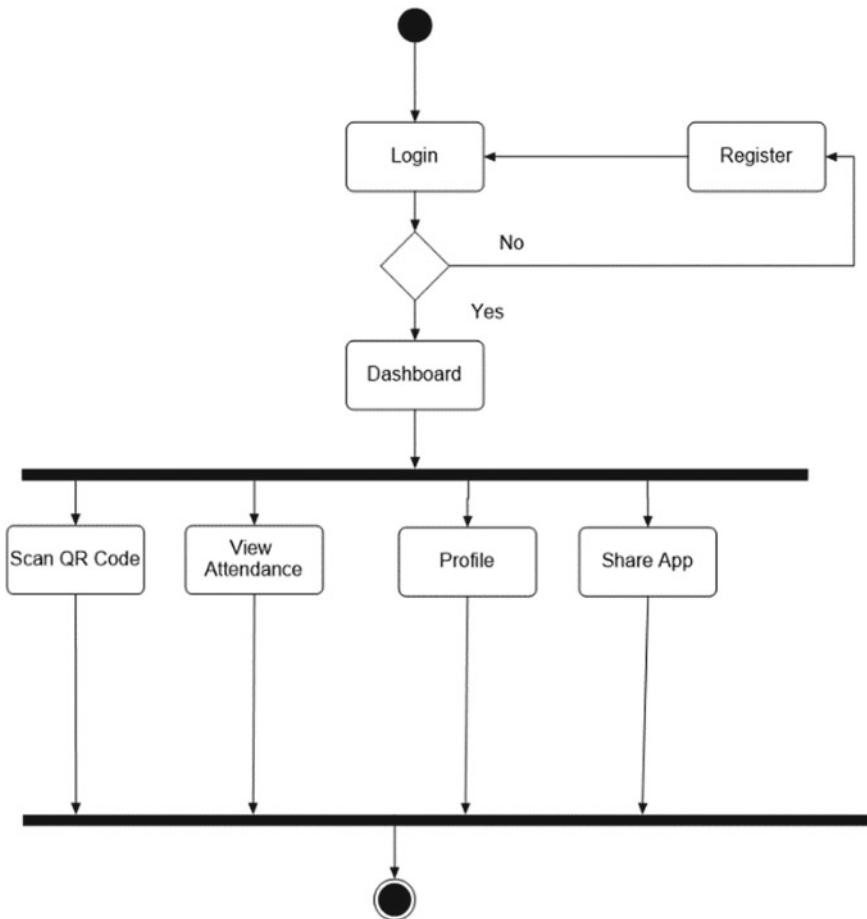


Fig. 4 Activity diagram [Student]

server layer. Feature base appearance-based approaches are used to extract features here [4].

Sanha Marvic Sijat et al. (2016) had developed a system for tracking students and maintaining the attendance. The system uses Muffin application with Arduino board and Bluetooth module. A desktop application is developed for maintaining the attendance. Data login is required through Arduino board three different buttons are shown in muffin application after successful login. They are students, teacher, and classes. By selecting appropriate options, attendance can be created or modified by teachers [5].

Mohammad Ausaf Anware et al. (2015) had designed and implemented smart phone-based attendance system. In this system, android platform is used. Teacher can enter the details of subject and can able to mark the presence of specific student

on specific date and time of lecture. Summary report can be generated for specific date [6].

Khaled Mahfouz et al. (2020) had developed a biometric attendance system which is implemented in school buses to maintain the entry and exit of the students. It comprises of smart table, automatic door opener, and fingerprint-based sensor interfaced using microcontroller. A route-based report is also generated which informs the driver whether specific student has boarded or exited or not [7].

A Kassem et al. (2010) had developed RFID-based attendance and monitoring system. The system is implemented in Notre Dame University on small scale. It was time-saving reliable and easy to upgrade. It can be further implemented in payment systems and quality control system [8].

Upanya Singh et al. (2016) had developed an android application for faculty of the college. It was implemented for JIIT Institute, Noida. The application allows teachers to create their profile, upload, and view the attendance and uploading the study material. Teachers can able to post the assignment and see the activities done by the students in feed [9].

Ghosh Swamendu et al. (2018) had developed smart attendance system. The system is developed using Arduino Uno controller, fingerprint sensor, Bluetooth controller, and LCD display. The system scans the data from students and sends it to mobile application developed for tracking the attendance system [10].

4 Software and Tools Technology

4.1 *Fronted Tools and Technology*

Android Studio

Android Studio is the authority incorporated development climate (IDE) for Android application development. It depends on the IntelliJ IDEA, a Java coordinated development climate for programming and incorporates its code altering and designer tools.

To help application improvement inside the Android working framework, Android Studio utilizes a Gradle-based form framework, emulator, code layouts, and GitHub reconciliation. Each task in Android Studio has at least one modality with source code and asset documents. These modalities join Android application modules, Library modules, and Google App Engine modules.

Android (XML)

The frontend is composed utilizing XML. It is very much like HTML in that it is a markup language that utilizes settled labels as its programming structure. Android utilizes a few XML documents to make the application's front end. There is, at any rate, one XML design record for every action (or a few on the off chance that you are supporting numerous gadget sizes), just as format documents for custom

perspectives. XML is additionally used to characterize consistent strings that will be set in the designs, like the content on a catch.

4.2 Backend Tools and Technology

JAVA

Java is a programming language and stages autonomous. Java was intended to have the look and feel of the C++ language; however, it is less difficult to use than C++ and authorizes an item arranged programming model. Java can be utilized to make total applications that may run on a solitary PC or be conveyed among workers and customers in an organization. It can moreover be used to amass a little application. Java is a programming language that produces programming for different stages. At the point when a developer composes a Java application, the accumulated code (known as bytecode) runs on most working frameworks (OS), including Windows, Linux, and Mac OS. Java derives a great deal of its syntax from the C and C++ programming lingos.

Java is a largely used programming language expressly expected for use in the spread environment of the web. It is the most standard programming language for Android wireless applications.

Java was intended to be not difficult to utilize and is along these lines simple to compose, assemble, investigate, and learn than other programming dialects. This permits you to make secluded projects and reusable code. Quite possibly the main benefit of Java is its capacity to move effectively starting with one PC framework then onto the next.

MYSQL

MYSQL is a database, and it is the most mainstream open-source data set framework. The free MYSQL data set is frequently utilized with PHP. The information in MYSQL is put away in data set items called tables. In the event that the PHP worker does not have an MYSQL data set, MYSQL can be downloaded effectively and unreservedly. Something incredible about MYSQL is that it tends to be downsized to help inserted database applications. Maybe it is a direct result of this standing that numerous individuals accept that MYSQL can just deal with little to medium-sized frameworks. Actually, MYSQL is the true standard data set for sites that help immense volumes of both information and end clients (like Friendster, Yahoo, and Google).

Web Service

PHP

Web administration offers interoperability between two distinct dialects. PHP is a worker-side scripting language. It must be deciphered on a worker that has PHP

introduced. It has inbuilt help for working connected at the hip with MySQL. It additionally upholds significant conventions like POP3, IMAP, and LDAP.

Web Server

XAMPP is a free and open-source cross-stage web laborer course of action stack group, involving generally of the Apache HTTP Server and go-betweens for substance written in the PHP and Perl programming dialects. It is straightforward, lightweight, and simple for designers to make a nearby web worker for testing and sending purposes.

5 Experiment and Result

The application mentioned in the above section is created using Android studio software. The results are shown and analyzed for various phones listed as follows.

These results are useful for students in understanding the use of the application. Various methods and parameters are involved in results are QR code generation and fetching attendance from scanning QR code. An activity diagram is also used to show the activities of the application. Performance of the application is measured by varying input parameters like validation of registration and login page.

Figure 5 shows the selection for faculty or student. The user will select as per role for use of the application. The user will be able to use the application according to his/her role.

Figure 6 shows a registration page. Through the figure, the user will be able to see the form fields like general id number, email address, password, which are to be used for account creation of students as well as for faculties. The registration form includes validations for the empty field as well as for a unique id number that checks the faculty and student id number.

Figure 7 shows a login page. Through the figure, the user will be able to see the form fields like email address, password which are to be used for login to students as well as for faculties. The login form includes validations for the empty field as well as for valid credentials.

Figure 8 shows four buttons one is for generating QR code, and the second one is for attendance report and other buttons are for assignment and profile view.

Figure 9 shows a QR code generation page. Through the figure, the user will be able to see the text box for writing the text and the save button for saving QR code into the phone. After writing text into textbox, you will be able to generate a QR code for the lecture or lab by clicking on generate button. After generating a QR code, you can save that and share that QR code on any platform.

Figure 10 shows an Attendance Faculty page. Through the figure, the user will be able to see the report subject-wise for a particular student in a percentage format.

Figure 11 shows four buttons one is for go to scan QR code, and the second one is for attendance report and other buttons are for assignment and profile view. The result after clicking on that buttons is explained in below figures.

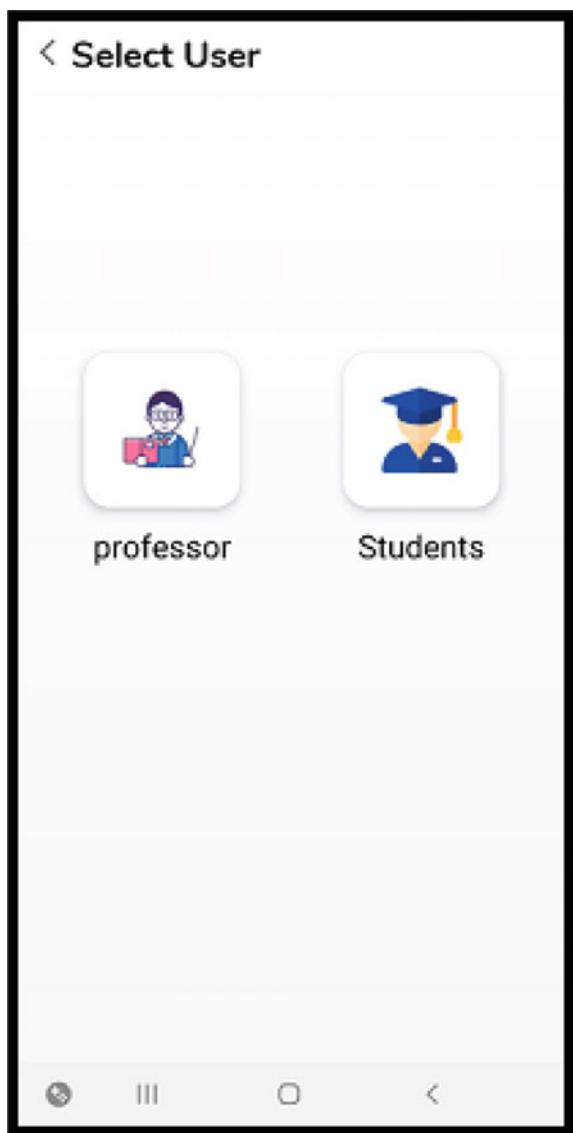
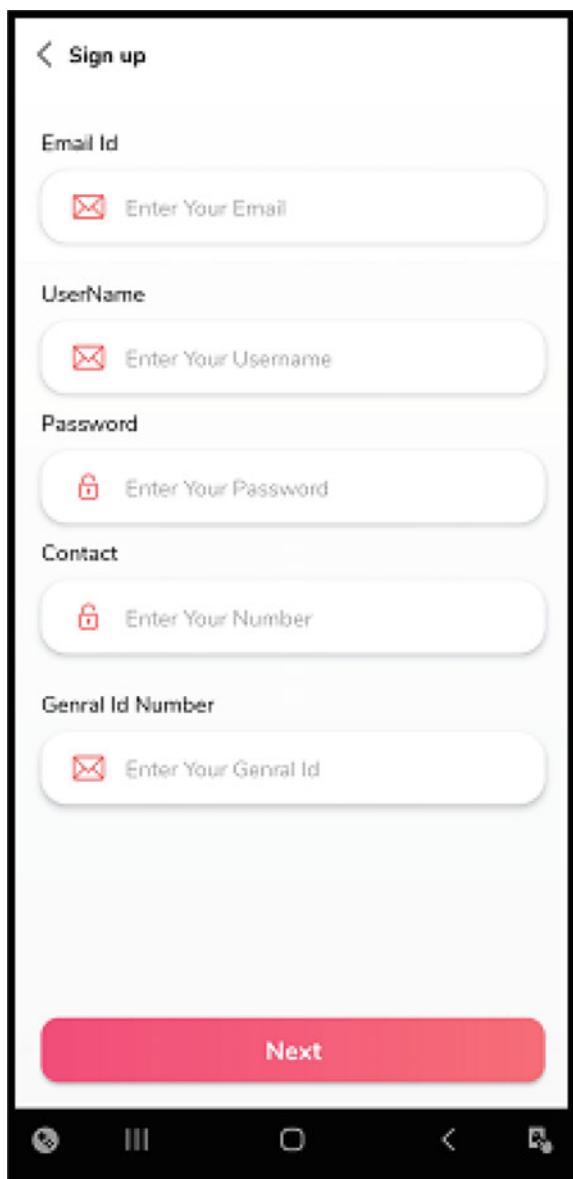
Fig. 5 Selection screen

Figure 12 shows a scan QR code button on the page. By clicking on the button student will be able to scan the QR code and mark attendance for the lecture.

Figure 13 shows an attendance report page. Through the figure, the user will be able to see the report subject-wise in a percentage format.

Fig. 6 Registration screen

Testing the Application

One popular method is black-box testing. In black-box testing, the tester only knows the inputs that can be given to the system should give. This form of testing is also called functional or behavioral testing. The most obvious functional testing procedure is exhaustive testing. One criterion for generating test cases is to generate them

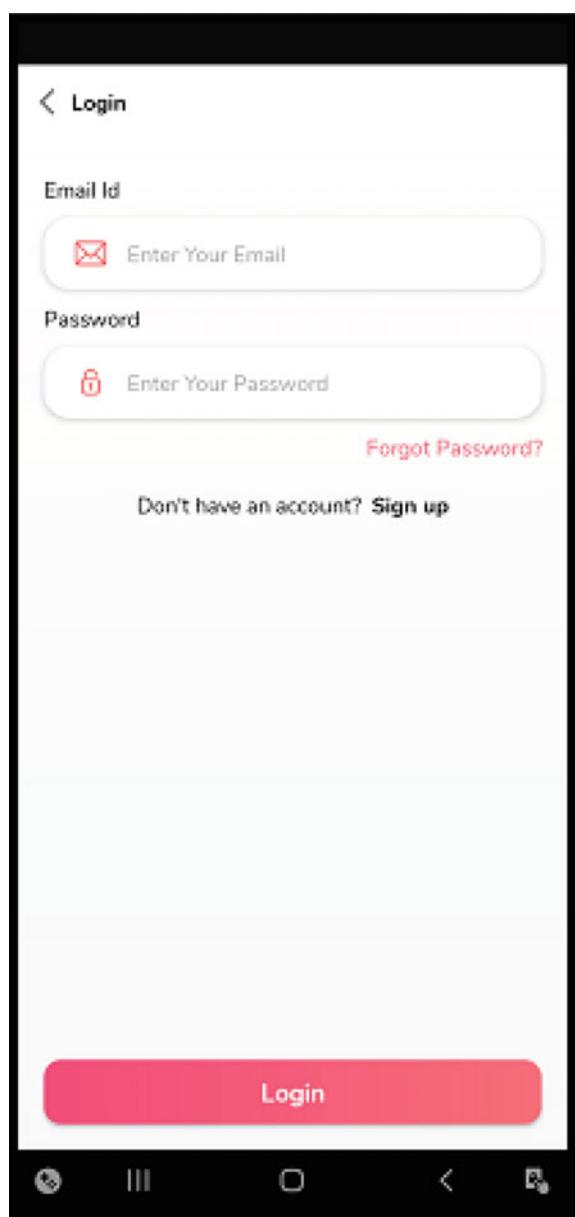
Fig. 7 Login screen

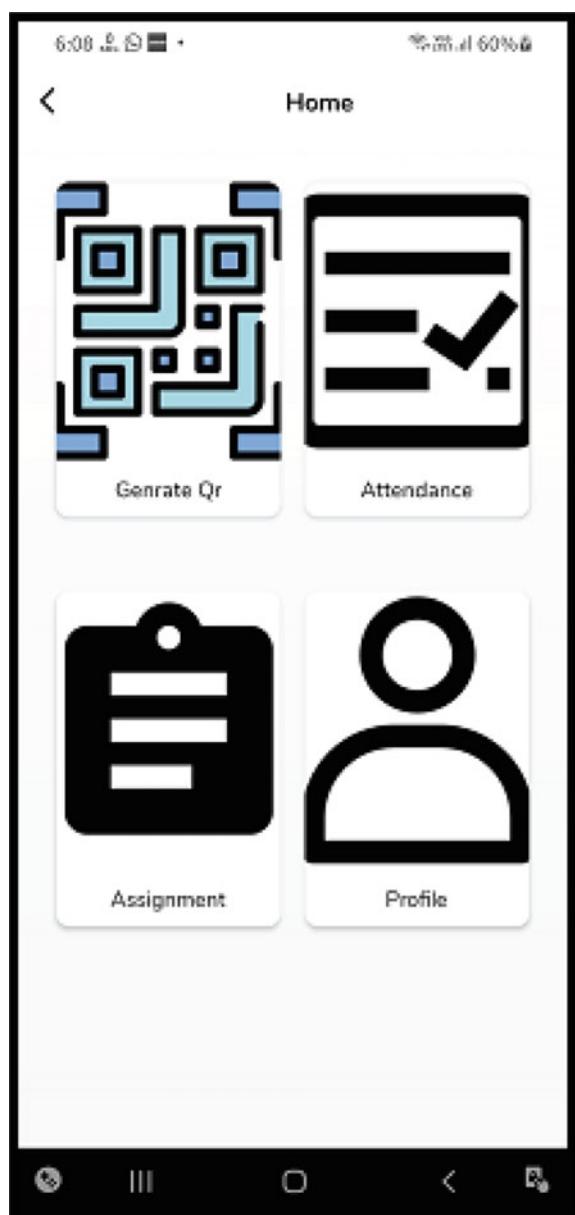
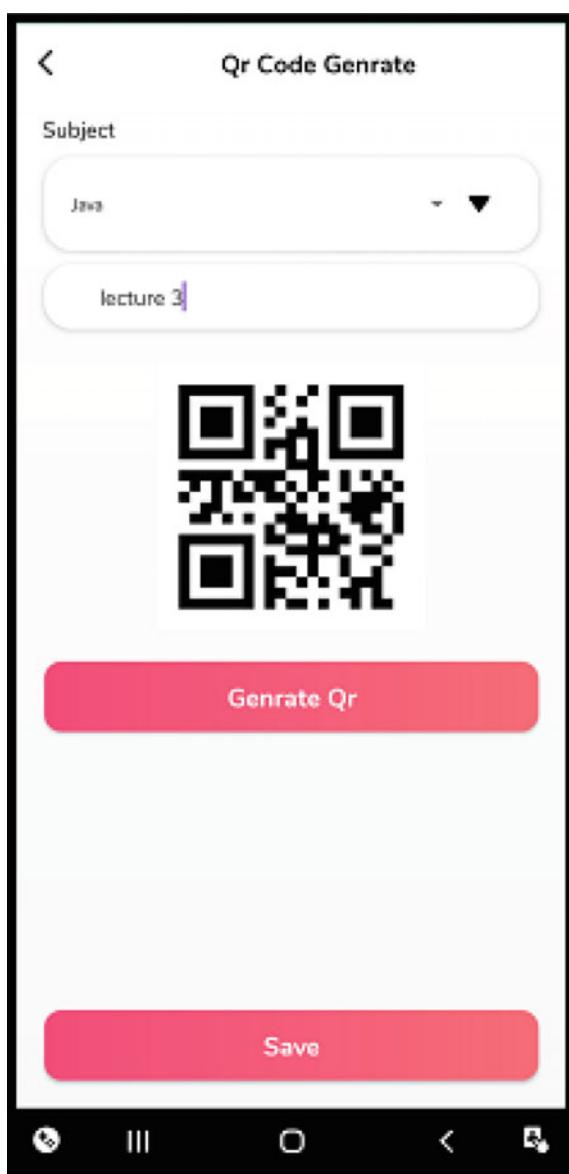
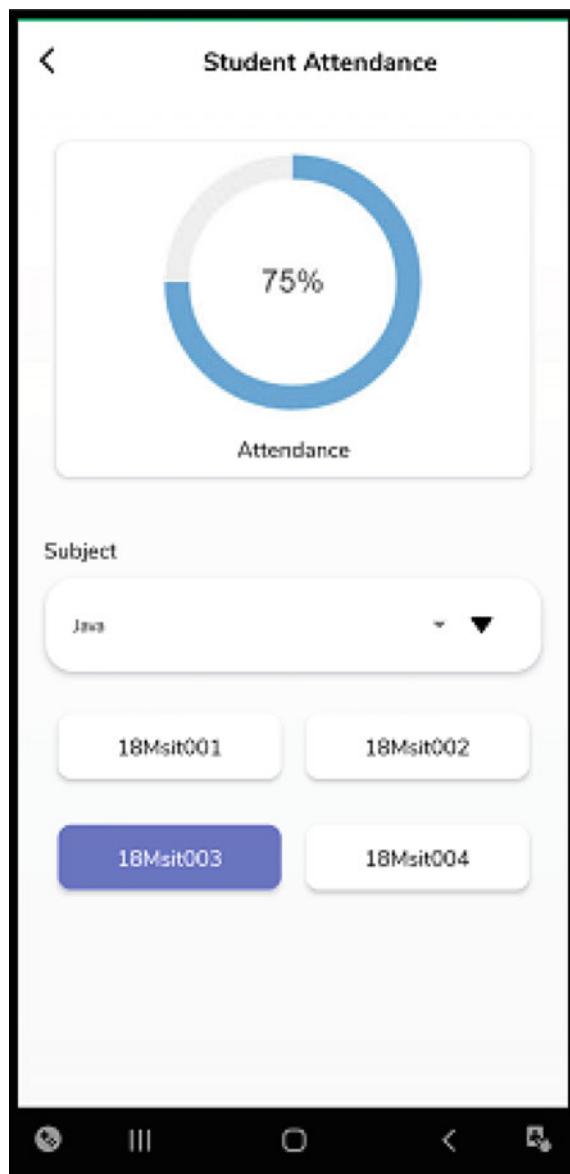
Fig. 8 Faculty dashboard

Fig. 9 QR code generation

randomly. There are formal rules for designing test cases for functional testing. In fact, there are no precise criteria for selecting test cases.

The details of validation which make the application work efficiently are shown in Table 1. It shows the input validation, expected output, and actual output. Various test cases at different application pages are login check, filling registration form,

Fig. 10 Attendance report

generating, and scanning QR code. The expected and actual outputs are the same when actual testing is done.

Fig. 11 Student dashboard

6 Conclusion and Future Scope

In COVID-19 era, it is very difficult for education sectors to maintain their quality. An attempt is made here to maintain the standard by developing an android application for attendance in the class. This system is based on generating and scanning QR code

Fig. 12 Scan QR code

for tracking the attendance. A teacher can generate the report of a class as well as of n individuals. An IOS version of this application is planned to implement in the future.

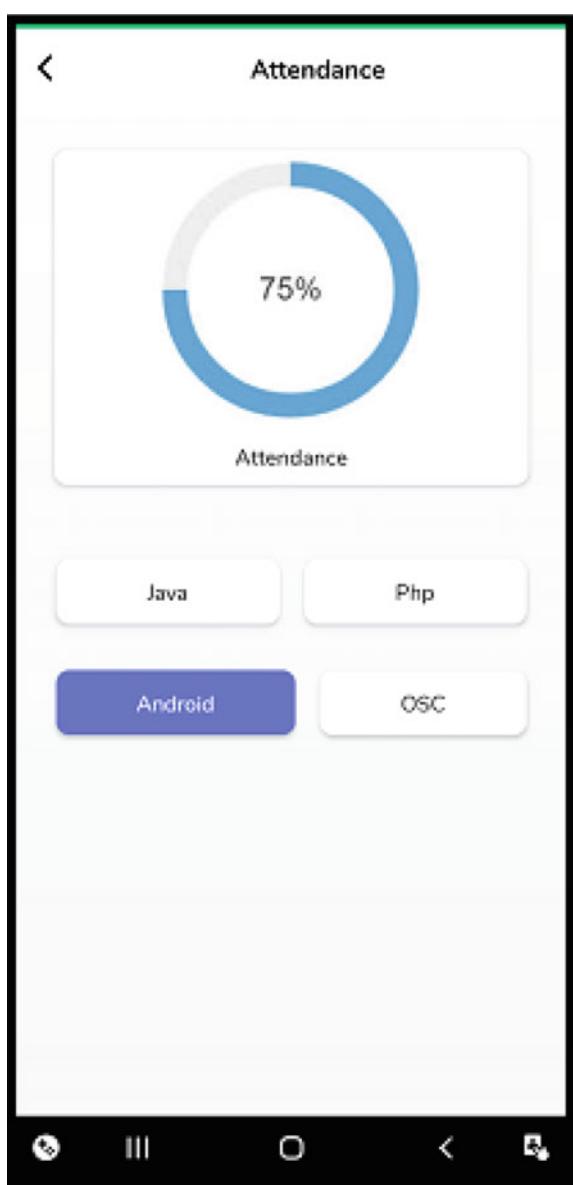
Fig. 13 Attendance report

Table 1 Testing with validation

Sr. No.	Test cases	Input	Expected output	Actual Output
1	Check login	Email Id and password	Move to the home page	Move to the home page
2	Fill registration form	Contact No More than 10 digit	Don't able to input	Don't able to input
3	Fill registration form	Fill the all fields with the correct details	Move to the login page	Move to the login page
4	Generate QR code	Fill the detail and click on generate button	Show the generated QR code	Show the generate QR code
5	Scan QR code	Scan the QR code	Show the message attendance done	Show the message attendance done

Acknowledgements The authors would like to thank Principal CMPICA and the management of the Charusat University for their constant motivation and support. Without his constant motivation and support, the work does not become so worthy.

References

1. Iio J Attendance management system using a mobile device and a web application. In: 2016 19th international conference on network-based information systems (NBiS), pp 510–515
2. Duraki S, Mehrat A, Demirci S A Mobile application for wireless attendance system. In: 2019 1st international informatics and software engineering conference (UBMYK), pp 1–6
3. Islam MM, Kamrul HM, Billah MM, UddinMM Development of smartphone-based student attendance system. In: 2017 IEEE region 10 humanitarian technology conference (R10-HTC), pp 230–233
4. Samet R, Tanriverdi M Face recognition-based mobile automatic classroom attendance management system. In: 2017 international conference on cyberworlds (CW), pp 253–256
5. Cisar SM, Pinter R, Vojnic V, Tumbas V, Cisar P Smartphone application for tracking students' class attendance. In: 2016 IEEE 14th international symposium on intelligent systems and informatics (SISY), pp 227–232
6. Anwar MA, Gangodkar D Design and implementation of mobile phones based attendance marking system. 2015 communication, control and intelligent systems (CCIS), pp 120–123
7. Mahfouz K, Rameshi SM, Rafat M, Elsayed M, Sheikh M, Zidan H Route mapping and biometric attendance system in school Buses. In: 2020 advances in science and engineering technology international conferences (ASET), pp 1–4
8. Kassem A, Hamad M, Chalhoub Z, El Dahdaah S An RFID attendance and monitoring system for university applications. In: 2010 17th IEEE international conference on electronics, circuits and systems, pp 851–854
9. Singh U, Srivastava N, Kumar A JIIT-EDU: an android application for college faculty. 2016 ninth international conference on contemporary computing (IC3), pp 1–6
10. Ghosh S, Mohammed SKP, Mogal N, Nayak P, Champaty B Smart attendance system. In: 2018 international conference on smart city and emerging technology (ICSCET), pp 1–5

Anomaly Detection in Thermal Images of Perishable Items Using Deep Learning



G. Ramyapriyanandhini, T. Bagyammal, Latha Parameswaran, and Karthikeyan Vaiapury

Abstract In this article, deep learning architecture has been used to detect anomalies on thermal images of perishable items like fruits and vegetables. Three experiments have been conducted, and their performances have been studied. In first experiment, two different deep learning models have been applied for classifying presence or absence of anomaly and Deep-CNN (Deep-Convolutional Neural Networks) performs well. In second experiment, YoLoV4 and YoLoV5 have been used to classify and localize the anomalies; YoLoV5 localizes the anomalies effectively with a mAP (Mean Average Precision) of 95.67%. In third experiment, YoLoV5 and Faster-RCNN (Faster Region Based Convolutional Neural Network) have been trained on the dataset in which each image contains multiple objects with and without anomalies to detect and localize them; Faster-RCNN gives mAP of 96.55% for detecting and localizing anomalies. From the experiments, it is observed that proposed deep learning models can be used for anomaly detection in thermal images.

Keywords Anomaly detection · Perishable items · Convolutional Neural Networks (CNNs) · Transfer learning · YoloV4 · YoloV5 · Faster-RCNN · Detectron

G. Ramyapriyanandhini · T. Bagyammal (✉) · L. Parameswaran
Computer Science Department, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: t_bagyammal@cb.amrita.edu

G. Ramyapriyanandhini
e-mail: cb.en.p2aid19024@cb.students.amrita.edu

L. Parameswaran
e-mail: p_latha@cb.amrita.edu

K. Vaiapury
TCS Research and Innovation, IITM Research Park, Chennai, India
e-mail: karthikeyan.vaiapury@tcs.com

1 Introduction

Retail is a huge trillion-dollar industry in a major price war. In past decade, it has seen important changes and continues to evolve quickly. Retailers need to adapt to new changes in attitudes, demographics and consumer preferences as the market dynamics are changing faster than ever. In this context, successful retailers will be able to improve operational performance and boost sales by focusing on technology and innovation and leveraging meaningful insights for new product development. Artificial intelligence (AI) is preferred mostly by retailers to boost their sales; and particularly, computer vision plays a major role from manufacturing unit to large departmental stores and even in online shopping platforms. Some applications of computer vision in retail includes stock visibility, cashier less stores, marketing and merchandising. The retail industry generates large quantities of data by collecting all the information during their daily operations using many modern technologies like IoT (Internet of Things). This data is used to make better decisions and competing on this data is the key to success of a retailer. Retailers who were the first to use this data with advances in technology like AI always beat the competition. Anomaly detection is one such advanced technique used in retail industry for the purposes like quality inspection of the objects in manufacturing, detecting abnormalities in planogram compliance, detecting abnormalities in perishable items like fruits and vegetables. Anomaly detection is one of the tasks in computer vision which detects the irregularities in the targeted group of data. In computer vision, an anomaly detection problem can be approached in three stages: classification, classification with localization and object detection.

This research work aims to propose a deep learning model for detecting anomalies in thermal images of perishable items like fruits and vegetables. This algorithm can be implemented as an anomaly detection system in retail stores where a vision system can capture and analyse the thermal images for anomalies. The research starts with collecting the dataset and preparing the dataset based on the models to be trained. The next phase is to train CNN based deep learning models that comes under three categories of computer vision tasks that are classification, classification with localization and object detection; the last phase is to compare the results obtained using the evaluation metrics and to choose the appropriate model that suits well for each task.

2 Related Work

Computer vision tasks are divided into four types they are: Classification, classification and localization, object detection and object segmentation. Classification can be done by using normal CNN architectures that are mentioned above either by training from scratch or by transfer learning. The remaining tasks are implemented by stacking

several CNN architectures together. Newer concepts called Bounding box and annotations are primarily used. Sultana et al. [1] have given a brief review of object detection models based on CNN. This work compares the performance of various CNN based architectures on PASCAL VOC 2007, 2010, 2012 and COCO 2015 and 2016 datasets. Cireşan et al. [2] have mentioned the importance of adding dropout layers in CNN architecture for classification purpose which is the head part of object detection models. Duth and Jayasimha [3] explored the Convolution Neural Network (CNN) and came up with an efficient vegetable recognition system that detects the category of the vegetables by extracting the features from the images. From the results of the evaluation, intraclass vegetables recognized accurately with 95.50% and efficiently using deep learning. Aloysius and Geetha [4] have done a thorough literature survey and discussed about the widely. Keeping AlexNet as the base model, they have reviewed all the variations emerged over time to suit various applications. Alias et al. [5] used transfer learning technique to classify high-resolution remote sensing images. Dharneeshkar et al. [6] used YOLOV3 model to detect potholes in Indian roads.

2.1 Anomaly Detection

Knowledge on these object detection models paved the way to the idea of approaching anomaly detection problem using deep learning models particularly using CNN architectures. Anomaly detection is an important research problem in various domain. Many anomaly detection techniques have been proposed and developed for certain application domains, while others are more generic. An extensive study of traditional anomaly detection methods as well as open challenges can be found in this survey [7]. Various methods have been proposed to detect the anomalies especially in image datasets using image processing techniques, deep learning techniques and machine learning techniques. Minhas and Zelek [8] proposed transfer learning-based anomaly detection on cement crack dataset and evaluated with MNIST and CIFAR10 datasets. In [9], the authors have proposed an algorithm using image processing techniques and machine learning models to detect the changes in the arrangement of objects present in the shelf of retail stores. Bagyammal et al. [10] proposed a statistical based algorithm for detecting changes in same scene captured in different timings. Chong et al. [11] performed three types of CNN models to classify different types of products collected from internet and tested on the shelf images of those products and found that the CNN models performed well. Vung et al. [12] proposed a road crack detection model using Faster-RCNN on Detectron2 library. Natasa et al. [13] proposed convolutional autoencoder model to detect low dimensional representation of the images. Many research works have been done using autoencoders and Generative Adversarial Networks (GANs) which are constructed using basic convolutional layers. These networks would reconstruct the images, train on them and detect anomalies in original images. From the literature, it is inferred that anomaly detection is still a challenging problem to be solved in various domain. Motka and

Parameswaran [14] have proposed anomaly detection in smart environments using thermal images.

In this paper, we have proposed unified frameworks for anomaly classification, anomaly classification with localization and anomaly detection.

3 Proposed Work

In the proposed method, first step is to prepare data for anomaly detection model. Next step is to choose a CNN algorithm as the basement for anomaly detection that acts as a feature extractor. Then, in the third step, various object detection algorithms or frameworks that come under one and two stage approaches are used to build bounding box detection and classify the anomalies. Finally, a brief comparison is done over these deep learning models using a set of evaluation metrics and concluded with the winning algorithm. Figure 1 shows the overall architecture of the proposed anomaly

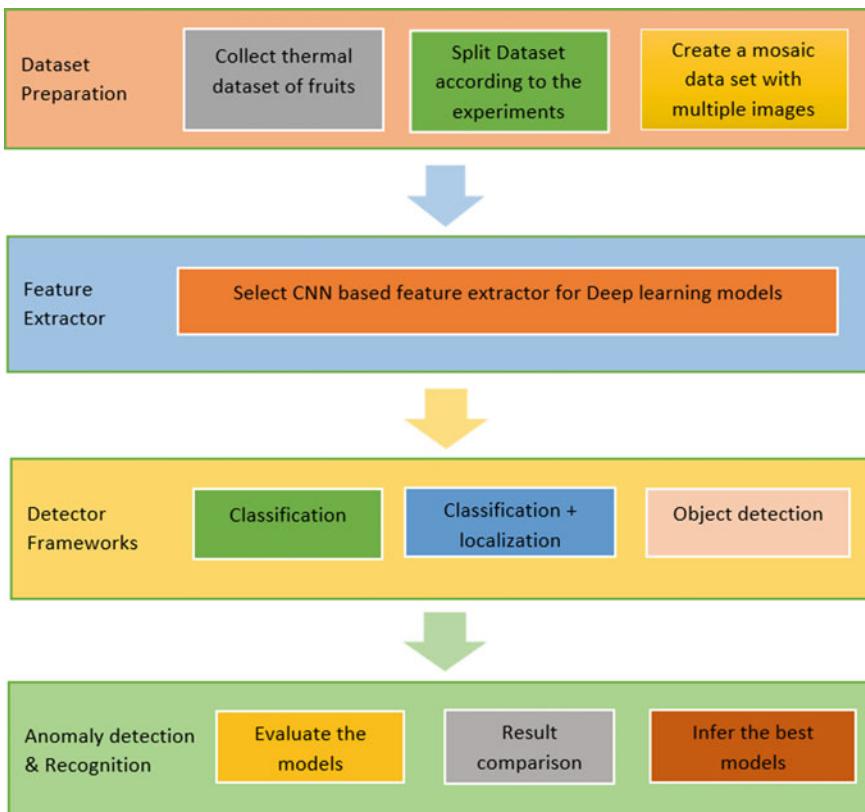


Fig. 1 Architecture diagram of the proposed anomaly detection system

detection system. The system goes through 4 stages Dataset Preparation, Feature Extractor, Detector Frameworks and Anomaly Detection and Recognition.

3.1 Dataset

The dataset used in this research is collected from the research work done by authors in [15] which contains thermal images of apples, bananas and cabbage. For the purpose of this research, a new set of mosaic dataset which contains only good and bad apples has been prepared using the existing dataset. Each image in the mosaic dataset has a mixture of good and bad apples which looks like a closely packed fruit basket or shelf image. The mosaic dataset is prepared in the following manner.

- Each image contains 100 annotations that has two labels “good_apple” and “bad_apple.”
- The annotation tool used is Makesense.ai [16] and polygon annotation as method.
- A total of 59 such images are prepared and augmentation techniques have been followed for increasing size of the dataset.
- Four types of augmentation techniques are used: Hue, saturation, exposure and noise. After augmentation, a total of 141 images are obtained with 13496 annotations.
- Of these images, 123 images have been chosen for training, 12 for validation and 6 for testing.

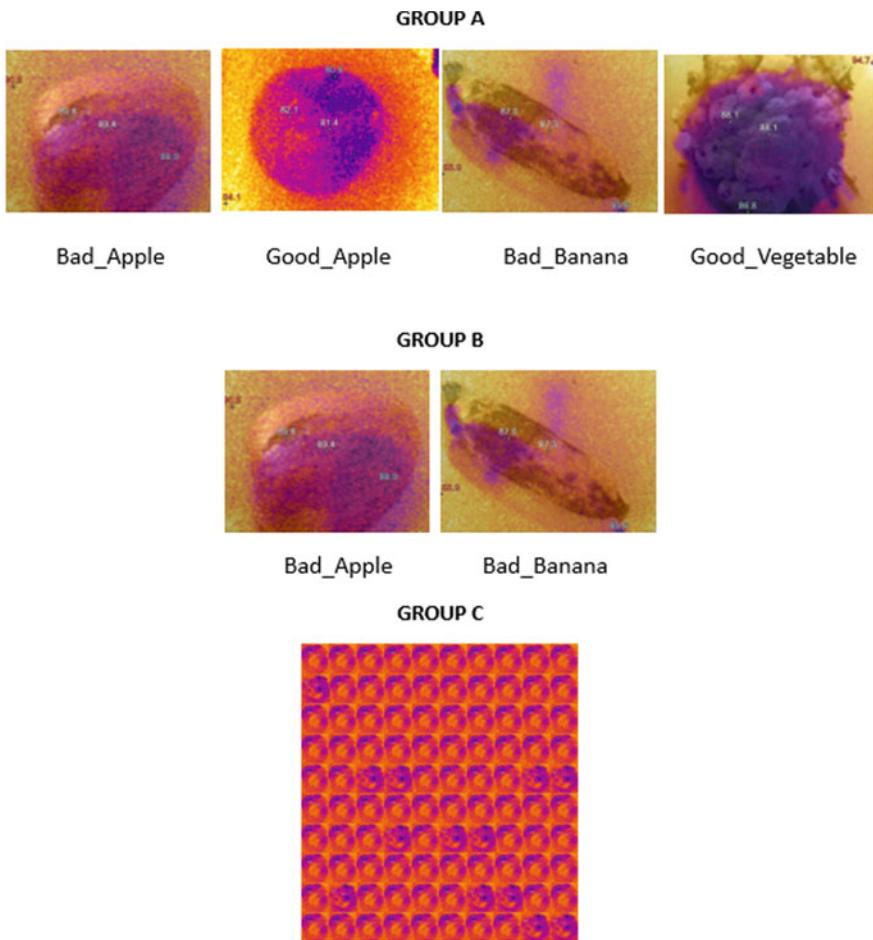
For experimentation purposes, the dataset has been grouped into three following ways:

- I. GROUP A: Thermal dataset which has single images of good apple (738), bad apple (1059), banana (695) and good vegetable (393).
- II. GROUP B: Single-thermal images of bad apples and bad bananas.
- III. GROUP C: Prepared mosaic dataset which contains mixture of only good apples and bad apples within an image.

Sample images of each group of datasets have been shown in Fig. 2. Group A contains 4 classes, Group B contains 2 classes and GROUP C contains 2 classes.

3.2 Methodology and Algorithm

This research work approaches the anomaly detection problem using the three basic tasks of computer vision that are classification, classification with localization and object detection. To do these tasks, three different combinations of datasets are created from dataset and six different deep learning models have been experimented. Two different deep learning models have been applied for classifying presence or absence of anomaly and two other models to localize anomalies. YoLoV5 and Faster-RCNN



Single image with combination of thermal images of good and bad

Fig. 2 Sample input images for each group of datasets

have been trained on the dataset in which each image contains multiple objects with and without anomalies to detect and localize them. Each phase of the research work is an experiment of applying a deep learning model on the dataset and analyzing the results. Thus, three experiments are conducted and explained in the further sections.

3.2.1 Experiment 1—Anomaly Classification

The first experiment is to approach anomaly detection as classification problem. For this purpose, we choose two CNN based deep learning models and trained them

over GROUP A dataset. First, we choose Deep-CNN model which is built from scratch using TensorFlow 2.0 and Keras 2.1.0 platform. The model is designed by stacking up max pooling layer on convolutional layer and created 4 such combination. The output from the final layer is flattened and added dropout layer with 0.5 as value. Adding dropout layer is a kind of regularization method which can avoid overfitting. The activation function used for convolution layers is Rectified Linear Unit (ReLU) and at flatten layer softmax is used. Since the classes are categorical values categorical_crossentropy loss function is used to compute the loss and Root Mean Square Propagation (RMSprop) is used as the optimizer with learning rate of 0.0001. The model is trained for 25 epochs with batch size of 10, and the metrics used for evaluation is accuracy.

Generally, it is good practice to prefer pretrained models rather than training a large neural network from scratch so we choose transfer learning model as the second model. In transfer learning model, the lower layers of a pretrained model are reused such that the weights are non-trainable so that gradient descent won't modify them and this method is called freezing. For our experiment purpose, we use Inception-v3 [17] model, which is pretrained on ImageNet dataset. The model consists of symmetric and asymmetric building blocks, like convolutions, max pooling, average pooling, concats, fully connected and dropouts layers. Batch normalization is used extensively throughout the model and applied to activation inputs. Softmax is used for computing loss. After few experiments on freezing and unfreezing the layers, we choose Mixed7 layer as the last layer and the lower layers are freezed, and the layers above that are removed and our model is trained using the GROUP A dataset for 25 epochs with batch size 10. TensorFlow and Keras are used for building this model. While training, we use categorical_crossentropy as loss function and RMSprop as optimizer with learning rate of 0.0001 as mentioned. It is good practice to reduce the learning rate when unfreezing the layers, and the metrics used for evaluation is accuracy. Figure 3 shows the transfer learning process of the proposed experiment. The input images are given to the model and higher level layers are kept freeze and the feature vectors are extracted from the middle layers and given to the fully connected layer and then SoftMax layer for final classification.

3.2.2 Experiment 2—Anomaly Classification with Localization

In this experiment, we approach anomaly detection as a classification with localization problem. For this task, we choose CNN based deep learning models which can perform classification of anomalies and also at the same time localize them using bounding box. So, we choose YoLoV4 and YoLoV5 for our experiment and are trained on GROUP B dataset. Dataset comprises of 340 annotated apple thermal images with anomalies and 360 annotated banana thermal images with anomalies. As mentioned in [18] YoLo is based on a single Convolutional Neural Network (CNN). It will look the image once and divide it into boundary boxes using CNN and predicts probabilities for each box. These probabilities are associated with the

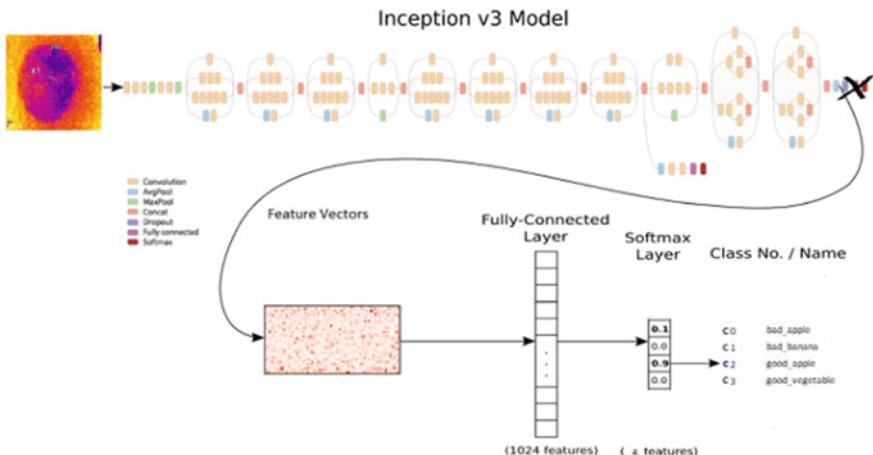


Fig. 3 Architecture of the transfer learning model using inception v3 model

classes and the model looks for the class with higher probability. YoLoV4's architecture is composed of Cross Stage Partial network (CSPDarknet53) as a backbone, spatial pyramid pooling additional module, Path Aggregation Network (PANet) neck and YoLoV3 head.

CSPDarknet53 is a novel backbone that can enhance the learning capability of CNN. PANet is used as the method for parameter aggregation for different detector levels instead of Feature pyramid networks (FPN) for object detection used in YoLoV3. PANet is used as neck to get feature pyramids. Feature pyramids help models to generalize well on object scaling. It helps to identify the same object with different sizes and scales. Feature pyramids are very useful and help models to perform well on unseen data. In YoLo v5 model, head is the same as the previous YoLoV3 and V4 versions. The model head is mainly used to perform the final classification part. It applies anchor boxes on features and generates final output vectors with, objectness scores, class probabilities and bounding boxes. YoLoV5 has 4 versions small, medium, large and xlarse in this research small version is used for all the groups of datasets. Figures 4 and 5 show the output from YOLOV4 and YOLOV5, respectively, which locates the defective part in the fruit, draws a boundary around the defective part and classifies whether the fruit is bad_apple or bad_banana. The ultimate aim of the system is to identify the defective fruit among a group of fruits and classify them into good or bad.

3.2.3 Experiment 3—Anomaly Detection

In this experiment, we approach anomaly detection as object detection problem. Here, we detect anomalies in an image which contains multiple objects (classes). For this, we train two types of object detection models, one stage object detector

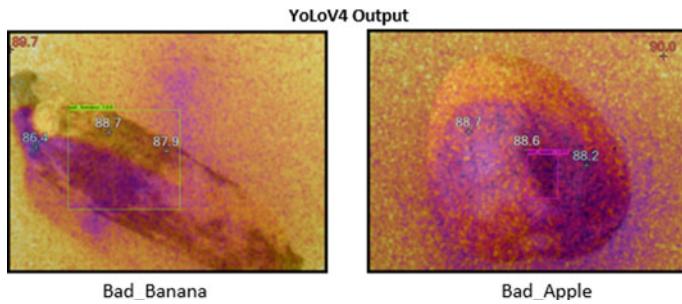


Fig. 4 Output from YOLOV4

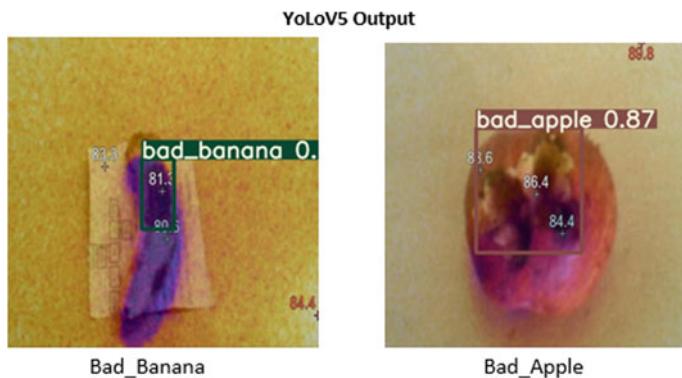


Fig. 5 Output from YOLOV5

and two stage object detectors. First, we choose YoLoV5 for one stage detector and trained on GROUP C dataset. The challenge in this experiment is that the dataset has closely packed objects. For training YoLoV5, the same platform and specification mentioned in the experiment 2 has been chosen and for evaluating the performance mAP has been used. Second, Faster-RCNN which is a two stage object detection model has been chosen. Instead of developing a Faster R-CNN model from scratch, Detectron2 library has been used to speed up the development cycle. Detectron2 is Facebook AI Research's next generation software system that implements state-of-the-art object detection algorithms. It is also a common practice to use a base model pretrained on a large image set (such as ImageNet) as the feature extractor part of the network. Detectron2 provides many such base-models. However, for Faster R-CNN, two commonly used models are R101-FPN and X101-FPN. These two pretrained models have good Faster R-CNN box Average Precision (AP) compared to others which are 42% and 43%, respectively. Though X101- FPN has better box AP on the ImageNet benchmark, it takes longer time to train/predict and might be overfitting in some cases. Hence R101-FPN has been selected. Figure 6 shows the output of YOLOV5 and Faster-RCNN with bounding box and corresponding class

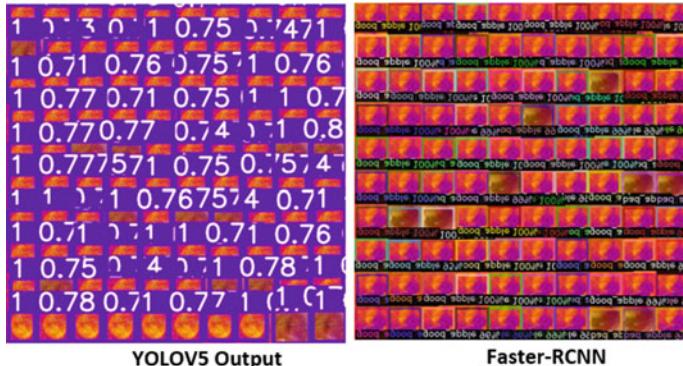


Fig. 6 Output of YOLOV5 and Faster-RCNN

probabilities. The aim of the system is to detect defected/bad fruits among a group of fruits.

4 Results and Discussion

4.1 Metrics and Evaluation

Metrics which are used to evaluate the performance of the algorithms used in the proposed approach are explained in this subsection. Precision is used to measure the correct predictions which can be calculated using Eq. (1) [19]. The outcomes where the model correctly predicts the anomalies are called True Positives (TPs). The outcomes where the model correctly predicts the non-anomaly class are called True Negatives (TNs). The outcomes where the model incorrectly predicts the anomalies are called False Positives (FPs). The outcomes where the model incorrectly predicts the non-anomaly class are called False Negatives (FNs). Recall is used to calculate the true predictions from all correctly predicted data, which has been calculated using Eq. (2). F1 score combines both recall and precision into a single measure that captures both properties. F1 score is mainly used to compare the classifier models and can be calculated using Eq. (3). Intersection over Union (IoU) is a metric that finds the difference between ground truth and predicted bounding boxes. Threshold value is given to select only appropriate probably higher valued bounding box. Equation (4) shows how to select the IoU threshold. mAP gives the precision for the entire model and compute average AP for each class and calculate the mean [19]. Accuracy is used only for classification models and calculated using Eq. (5).

$$\text{Precision} = \text{True positives}/(\text{true positives} + \text{false positives}) \quad (1)$$

Table 1 Comparison of the experiment results

Experiment	Proposed method	Model type	Training dataset	Performance metrics	IoU	Values (%)
Anomaly	Deep-CNN	Classifier	GROUP A	Accuracy	–	96.84
Classification	Inception v3(transfer learning)	Classifier	GROUP A	Accuracy	–	82.98
Anomaly classification with localization	YoLoV4	One stage object detector	GROUP B	mAP	0.5	79.27
	YoLoV5	One stage object detector	GROUP B	mAP	0.5	95.67
Anomaly detection	YoLoV5	One stage object detector	GROUP C	mAP	0.5	64.71
	Faster-RCNN	Two stage object detector	GROUP C	mAP	0.5	96.55

$$\text{Recall} = \text{Truepositives}/(\text{truepositives} + \text{falsenegative}) \quad (2)$$

$$\text{F1score} = 2 * (\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \quad (3)$$

$$\text{IoU} = \text{Areaofoverlap}/\text{areaofunion} \quad (4)$$

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (5)$$

Table 1 shows the comparison of the experiment results which contains model type, performance metrics used and their values.

4.1.1 Inference

From the experiments conducted for the three tasks, anomaly classification, anomaly classification with localization and anomaly detection using various deep learning frameworks and models, the following observations and inferences are made:

- Deep Convolutional Neural Network performs well by achieving 96.84% accuracy on GROUP A dataset for classification of anomalies rather than transfer learning which is performed on Inception-v3 model.

- Though YoLoV4 trained from scratch on GROUP B dataset, it could achieve only 79.27% mAP. But YoLoV5 achieves 95.67% mAP by using pretrained weights trained on Common Objects in Context (COCO) dataset. Therefore, YoLoV5 performs well for anomaly classification with localization task.
- For anomaly detection task on GROUP C dataset YoLoV5 performs poorly when compared to Faster-RCNN via Detectron2. From this, we can infer that two stage detectors perform much better than one stage detector on closely packed objects image dataset.

5 Conclusion

In this work, a novel approach to detect anomalies in thermal image dataset using deep learning has been proposed and experimented with various deep learning models starting from deep-CNN architecture, transfer learning using inception-v3 model and extended to CNN based object detection models like Faster-RCNN, YoLoV4 and YoLoV5 implemented on three groups of datasets. From the experimental results, it is observed that Faster-RCNN performs well based on the evaluation metrics. From this detailed analysis and comparison, it may infer that anomaly detection can be viewed as object detection problem. Since anomaly detection is a problem for detecting and localizing the abnormalities in a dataset, object detection model gives the feasibility to locate the anomalies and classify them as what type of anomaly. These types of anomaly detection systems are in a huge demand in retail industry particularly for perishable items where retailers need to worry about monitoring the perishable items for the whole day. This proposed algorithm is useful to identify and locate anomalous objects from the given thermal image. Deep-CNN performs well by achieving 96.84% accuracy for classification purpose of GROUP A dataset rather than transfer learning which is performed on Inception v3 model. Though YoLoV4 has been trained on GROUP B dataset for 6000 iterations, it could achieve only 79.27% mAP. But YoLoV5 achieves 95.67% mAP with 99 iterations. For GROUP C dataset, YoLoV5 performs poorly when compared to Faster-RCNN via Detectron2. From this, we can infer that two stage detectors perform much better than one stage detector on closely packed objects image dataset. Proposed method can be extended to perform more comparisons on one stage and two stage object detectors using the available thermal image dataset and propose a deep learning framework suited for such constrained dataset as an anomaly detector for detecting anomalies in thermal images of perishable items like fruits and vegetables.

References

1. Sultana F, Sufian A, Dutta P (2020) A review of object detection models based on convolutional neural network

2. Cireşan D, Meier U, Masci J, Gambardella LM, Schmidhuber J (2011) High-performance neural networks for visual object classification. *Comput Res Repository (CoRR)*
3. Duth PS, Jayasimha K (2020) Intra class vegetable recognition system using deep learning. In: 4th international conference on intelligent computing and control systems (ICICCS), pp 602–606
4. Aloysius N, Geetha M (2017) A review on deep convolutional neural networks. In: International conference on communication and signal processing (ICCSP), pp 0588–0592
5. Alias B, Karthika R, Parameswaran L (2018) Classification of high-resolution remote sensing images using deep learning techniques. In: International conference on advances in computing, communications and informatics (ICACCI), pp 1196–1202
6. Dharneeshkar J et al (2020) Deep learning based detection of potholes in Indian roads using YOLO. In: 2020 International conference on inventive computation technologies (ICICT), pp 381–385
7. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. *ACM Comput Surv (CSUR)* 41(3):1–58
8. Minhas M, Zelek J (2019) Anomaly detection in images. arXiv preprint [arXiv:1905.13147](https://arxiv.org/abs/1905.13147)
9. Muthugnanambika M, Thirumurthy B, Parameswaran L, Vaipury K (2018) An automated vision based change detection method for planogram compliance in retail stores. In: Computational vision and bio inspired computing. Springer, Cham, pp 399–411
10. Thirumurthy B, Parameswaran L, Vaipury K (2018) Visual-based change detection in scene regions using statistical-based approaches. *J Electr Imaging* 27
11. Chong T, Bustan I, Wee M (2016) Deep learning approach to planogram compliance in retail stores. *Semant Scholar* 1–6
12. Pham V, Pham C, Dang T (2020) Road damage detection and classification with Detectron2 and faster R-CNN. arXiv preprint [arXiv:2010.15021](https://arxiv.org/abs/2010.15021)
13. Sarafijanovic-Djukic N, Davis J (2019) Fast distance-based anomaly detection in images using an inception-like autoencoder. In: International conference on discovery science. Springer, Cham, Oct 2019, pp 493–508
14. Motka H, Parameswaran L (2019) A vision based approach for anomaly detection in smart environments using thermal images. *Int J Innovative Technol Exploring Eng* 8(7)
15. Mishra C, Thirumurthy B, Parameswaran L (2021) An algorithm design for anomaly detection in thermal images. In: Innovations in electrical and electronic engineering. Springer, pp 633–650, Singapore
16. Maier W, Eschey M, Steinbach E (2011) Image-based object detection under varying illumination in environments with specular surfaces. In: 18th IEEE international conference on image processing, Sept 2011, pp 1389–1392
17. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
18. Bochkovskiy A, Wang C-Y, Liao H-Y (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934)
19. Powers DMW (2020) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint [arXiv:2010.16061](https://arxiv.org/abs/2010.16061)

Median Filtering Detection Using Ensemble Methods



Sajjad Ahmed and Saiful Islam

Abstract Passive image forgery techniques are aimed at detecting image tampering without any a prior knowledge about the images. One of the major challenges in the field of image forensics is detection of basic operators such as application of median filtering. Median filter is one of the basic operators that is being used for malicious intent of hiding history of operations applied on an image. The studies related to detection of median filtering are becoming very popular in the field of digital image forensics. A large number of methods have been proposed for detection of median filtering. The paper presents improved percentage streak area based on streak effect in median filtered images. An ensemble classifier is used for detection of median filtered images by performing a multi class classification using bagged method with decision tree as weak learner. The results show that the classifier performs better than state-of-the-art methods for multiclass classification between original, median filtered and postJPEG compressed median filtered images.

Keywords Image forensics · Median filter · Median filter detection · Machine learning · Ensemble methods

1 Introduction

Median filter is a very popular filter in signal and image processing. In digital images, it is applied for removal of spikes there by smoothing out images without degrading edge quality. As median filter operation is a nonlinear operation, it has found its way into anti forensics. Median filtering is widely used for concealing trail left by other image forgery techniques. The forgers in the area apply median filter to hide evidences left by tampering methods such as in [1, 2].

S. Ahmed (✉)
Baba Ghulam Shah Badshah University, Rajouri, India
e-mail: sajjadahmed@zhcet.ac.in

S. Islam
ZHCET, Aligarh Muslim University, Aligarh, India

In various works related to median filtering detection, support vector machine (SVM) or neural network based binary classifier is used. None of the work has done multi class classification for the purpose. The contribution of our work is improvement of feature vector proposed in [3] and then applying ensemble methods for classification of original unfiltered, median filtered and postJPEG compressed median filtered images.

The rest of the paper is organized as follow: the background is provided in Sect. 2, the proposed improved feature vector is described in Sect. 3, Sect. 4 discusses experimental setup, the results are discussed in Sect. 5, and finally, the work is concluded and future work is described in Sect. 6.

2 Background

Median filtering is defined for two-dimensional data, such as digital images that are represented as a 2-dimensional matrix of intensity value, as a pixel-based operation. The median filter is applied on images by taking median of pixel intensity enclosed in a $w \times w$ neighborhood of the pixel under processing. Every pixel is processed one by one and a new median filtered image is obtained. Details of the median filter and its properties may be found in [4, 5]. When median filtering is applied on an image, it introduces streaks that is a run length of pixels with same or almost same intensity value pointed out by Bovik [6]. Steaking effect is show in Fig. 1. This may be used as an artifact or characteristics finger print for detection of median filtering and is used in works such as [3, 7, 8]. The work in [3] measured streaking artifact as percentage streak area and introduced a feature vector based on percentage streak area detect median filtering of digital images. In our work, we propose an improved percentage streak area-based feature vector for better detection of median filtering and then applied ensemble methods for multi class classification task.

2.1 Ensemble Methods

Ensemble learning is a powerful technique for supervised learning. In these methods classification is performed by training a large number of classifiers and then results are integrated from these classifiers for prediction. Ensemble methods consist of two phases. First phase is to generate large number of classifiers and then combining the results. The ensemble classifier may consist of same or different types of base classifiers. Homogeneous classifiers are common choice in literature. The homogeneous classifiers are trained using different set of instances of input dataset [9]. A large number of ensemble method are available in literature with their own advantages and disadvantages. Some of the most popular methods are Bagging, boosting, subspace and random undersampling boosting (RUboost) as ensemble method and are briefly discussed as follow.

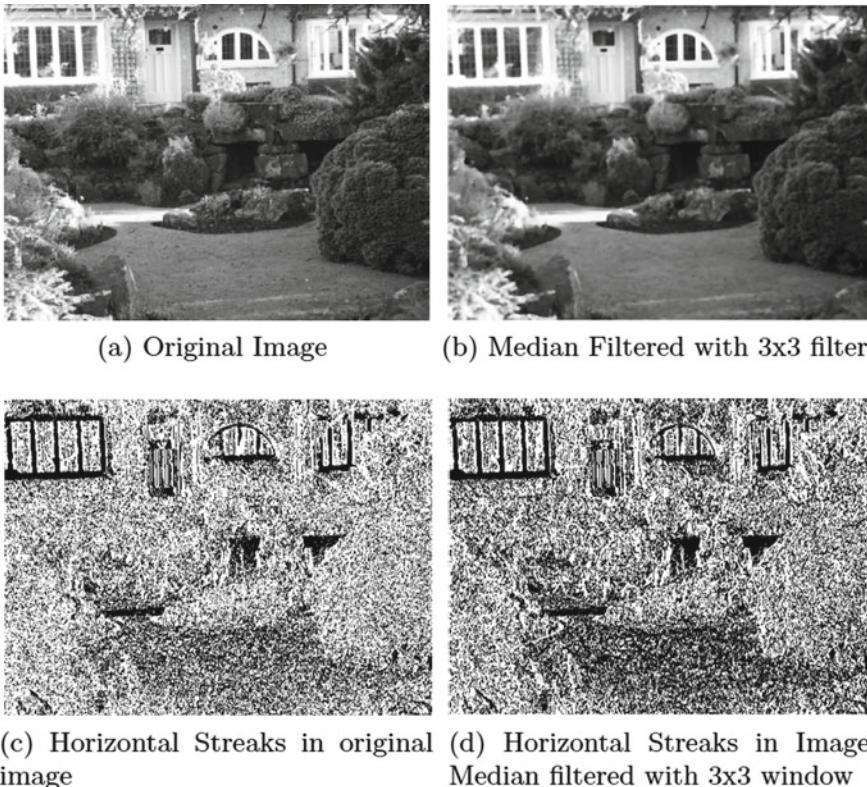


Fig. 1 Streaking Artifact in image UCID00025.tif taken from UCID data set

Bagging Bootstrap aggregating also known as bagging was proposed in [10]. Bagging works by training learners on samples of training dataset with replacement. The base weak learners are then trained on these samples and majority voting is done to combine the results of all induced classifiers to create a high-precision classifier.

Boost Boosting [11] is an important machine learning algorithm. The idea behind boosting is that the models are trained with weights assigned to training sets. The weights assigned are dynamically updated based on classification results in i th iteration. The weights of n th training set are increased if the model makes classification error. Thus, focusing more on models that misclassify thereby creating an ensemble of classifiers that are then combined to create final classifier. AdaboostM2 [12] is an extension to Adaboost for multiclass classification.

Random Undersampling boosting Random under sampling, RUSBoost, [13] are particularly effective for dataset where different classes have different number of instances. The size of sample depends on the number of instances in class with smallest number of instances of data. The other classes are then undersampled to construct samples for training of weak classifiers. AdaboostM2 is then used to construct ensemble of weak classifiers.

Subspace methods The random subspace ensemble method works by taking subsets of feature and then each subset of feature is used to train a weak classifier. The results are then combined to generate a high-precision classifier. Usually, subspace is used along with discriminant analysis and k-nearest neighbor (KNN) as weak classifiers [14].

3 Proposed Method

A large number of methods has been proposed for detection of median filtering in digital images. Our work is based on detection of median filtering of digital images using streaking artifact. The streaking effect was measured using percentage streak area by Ahmed et al. in [3]. The author in [3] used percentage streak area as a measure of streaking effect and developed a feature vector based on it for binary classification. Our proposed method is an improvement over the method given in [3]. The authors calculated percentage streak area using left to right and top to bottom streaks. We proposed that the measurement of streaking artifact may be further improved by using streaks along main diagonal and off diagonal of the digital image. Main diagonal streaks are streaks measured from top left to bottom right and off diagonal a streak are the streaks measured from top right to bottom left. Let I is a digital image in gray scale mode and \hat{I}_w is the median filtered image filtered with a window of size $w \times w$. Percentage streak area is defined in [3] as

$$psa(I) = \frac{\sqrt{2}(HSA(I) + VSA(I))}{2MN} \times 100 \quad (1)$$

where $HSA(I)$ is number of pixels involved in streaking when horizontal streaks are considered and $VSA(I)$ is number of pixels involved in streaks when streaks are measured in vertical direction and $M \times N$ is size of image.

The percentage streak is be further improved by considering main diagonal $MDSA(I)$ and off diagonal streaks. $ODSA(I)$. The proposed percentage streak area $ipsa(I)$ may be defined as

$$ipsa(I) = \frac{\sqrt{2}(HSA(I) + VSA(I) + MDSA(I) + ODSA(I))}{4MN} \times 100 \quad (2)$$

where $MDSA(I)$ is number of pixels involved in streaking measured along main diagonal and $ODSA(I)$ is number of pixels involved in streaking measured along off diagonal of the image.

Finally, a 3-dimensional feature vector extracted by applying $w \times w$ filter window size is derived from improved percentage streak area ($ipsa$) as follow:

$$Ifv_w(I) = \left[\left(ipsa(I) - ipsa(\hat{I}_w^1), ipsa(\hat{I}_w^1) - ipsa(\hat{I}_w^2), ipsa(\hat{I}_w^2) - ipsa(\hat{I}_w^3) \right) \right] \quad (3)$$

where I is the test image and \hat{I}_w^i is i -times median filtered image using $w \times w$ filter. The improved feature vector $Ifv_w(I)$ is better at differentiating among original unfiltered, median filtered and postJPEG compressed images. The 3-dimensional feature vector was extracted from images in [3] using following equation.

$$fv_w(I) = \left[\left(psa(I) - psa(\hat{I}_w^1), psa(\hat{I}_w^1) - psa(\hat{I}_w^2), psa(\hat{I}_w^2) - psa(\hat{I}_w^3) \right) \right] \quad (4)$$

4 Experimental Setup

Ensemble of classifiers was employed for multi class classification task. In ensemble classification, a group of weak classifiers are trained on the input dataset and then either averaging or majority voting is done for final classification results. Thus, in ensemble method, many weak learners are used to generate a high-quality ensemble predictor. We explored various weak learners and ensemble method for the classification of original images, median filtered images and postJPEG compressed median filtered images for multiclass classification. We have reported results for Bag, AdaboostM2, subspace and random undersampling boosting (RUboost) as ensemble method and decision tree, discriminant and K-nearest neighbors (KNN) as weak learners.

For bagged, AdaboostM2 and RUBoosted decision tree were used as weak learner and subspace ensemble method was used with discriminant and Nearest neighbors as weak learners. For experimentation purposes, the standard datasets UCID [15] and BOSS [16] were used. The UCID dataset contains 1338 colored images and BOSS dataset contains 10,000 images and were transformed to gray scale before extracting features using Eq. 3. A dataset $DS_{orig} = \{UCID, BOSS\}$ was thus constructed with total of 11,338. From DS_{orig} median filtered dataset DS_w were constructed by applying median filter with filter window size of $w = 3 \times 3$, and $w = 5 \times 5$ on original and unfiltered images on DS_{orig} . PostJPEG image dataset DS_w^{qf} was constructed from DS_{orig} by first applying median filter with window size $w \times w$ and then applying JPEG compression at different quality factor of $qf = \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$.

From dataset $\{DS_{orig}, DS_w, DS_w^{qf}\}$ training and test datasets were created. The training ratio was set at 75% and remaining 25% images were used for testing purpose. A system with intel core i7 10th generation processor with 8 GB RAM was used for classification.

5 Results and Discussion

For accessing the accuracy of our proposed improved percentage streak area using ensemble methods, the dataset DS_{orig} , DS_w and DS_w^{qf} were processed and feature vectors, using Eq. 4 and improved feature vector using Eq. 3 were extracted. The multiclass dataset constructed from three classes DS_{orig} , DS_w and DS_w^{qf} for $w = 3$ and $qf = \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$. The above datasets were split in to training and test sets with train ratio of 75% and remaining 25% were used for testing. Multi class classification was performed using different ensemble methods using improved feature vector in Eq. 3.

The ensemble methods Bootstrap aggregating (bagging) for classification after comparison with other popular ensemble methods based on accuracy. Table 1 shows the results when data was classifier using different ensemble methods such as Ada boost, subspace discriminant, subspace KNN and RUS Boosted methods. It can be clearly inferred for the Table 1 that among the popular ensemble methods the bagged ensemble method with decision tree as weak learner performs better with respect to classification accuracy. The bagged method is also fast as it is classifying large number of samples per second without compromising on accuracy. For comparison with method in [3], the feature vector was extracted using Eq. 4 from dataset DS_{orig} , DS_w and DS_w^{qf} . The data was split into training and test sets with 75% data in training set and remaining 25% reserved for testing purpose. The SVM configured with linear kernel was used for classification and results are reported in Table 2. The training data was then used to train bagged ensemble method and decision tree as weak learner. The optimal results were obtained with 22 learners. Table 2 shows results of the testing with remaining 25% data. The True positive rate, P_{tpr} , and average decision error, P_e , is calculated using Eq. 5 and Eq. 6, respectively, from multiclass confusion matrix. P_{tpr} is defined as.

$$P_{tpr} = \frac{\text{True Positives}}{\text{True Positive} + \text{False Negative}} \quad (5)$$

Table 1 Comparison of various ensemble methods for multi class classification for original versus MF3 versus MF3JPEG90

Method	Accuracy (%)	Prediction speed (object/s)	Training time (s)
Bagged	99.40	11,000	11.29
AdaBoost	99.10	12,000	12.57
Subspace discriminant	93.40	7,300	9.44
Subspace KNN	99.20	5,000	10.46
RUS Boosted	98.50	12,000	14.93

Table 2 Multiclass classification of ensemble bagged classifier and [3]

Proposed method		[3]		
QF	TPR (%)	P_e (%)	TPR (%)	P_e (%)
90	99.18	0.18	98.98	0.48
80	99.05	0.16	98.62	0.48
70	99.00	0.16	98.4	0.52
60	98.92	1.8	98.203	0.54
50	98.93	0.20	98.036	0.54
40	98.46	0.11	97.75	0.56
30	98.02	0.205	97.747	0.56
20	98.51	0.093	98.3	0.57

and average decision error (P_e) is defined as

$$P_e = \min\left(\frac{P_{fpr} + 1 - P_{tpr}}{2}\right) \quad (6)$$

where P_{fpr} denotes false positive rate and P_{tpr} represents true positive rate calculated 3×3 confusion matrix. Clearly, our improved feature vector when classified using bagged ensemble method with decision tree as weak learner performs better than the work of [3] for all postJPEG compression quality factors. Average decision error (P_e) is also low for our method.

6 Conclusion and Future Work

We performed median filtering detection using various ensemble methods and proposed an improved feature vector for median filtering detection which is based on streaking effect. The results shows that the bagged method with decision tree as weak learner performs better than other the methods proposed in [3]. The various ensemble methods were compared for the purpose and found out that the bagged method decision tree as weak learner performs better than other ensemble methods. In future, there is a need to perform the study with bigger dataset. Also, performance of the proposed method needs to be tested against other filters such as Gaussian filter, average filter, etc. There is a need for testing localization of median filtered region in unfiltered images. The performance against detection of median filtering in low resolution is also needed.

References

1. Kirchner M, Bohme R (2008) Hiding traces of resampling in digital images. *IEEE Trans Inf Forensics Secur* 3(4):582–592
2. Stamm MC, Liu KR (2011) Anti-forensics of digital image compression. *IEEE Trans Inf Forensics Secur* 6(3):1050–1065
3. Ahmed S, Islam S (2018) Median filter detection through streak area analysis. *Digital Invest* 26: 100–106 [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S174287617303109>
4. Tukey J (1971) Exploratory data analysis. Addison-Wesley, MA
5. Justusson BI (1981) Median filtering: statistical properties. Springer, Berlin, Heidelberg, pp 161–196. [Online]. Available: <https://doi.org/10.1007/BFb0057597>
6. Bovik AC (1987) Streaking in median filtered images. *IEEE Trans Acoust Speech Signal Process ASSP-35(4):181–194*
7. Cao G, Zhao Y, Ni R, Yu L, Tian H (2010) Forensic detection of median filtering in digital images. In: IEEE international conference on multimedia and expo (ICME), pp 89–94
8. Kirchner M, Fridrich J (2010) On detection of median filtering in digital images. In: IS&T/SPIE Electronic Imaging, pp 754 110–754 110
9. Rokach L (2010) Ensemble-based classifiers. *Artif Intell Rev* 33(1):1–39
10. Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
11. Freund Y, Schapire RE et al (1996) Experiments with a new boosting algorithm. ICML 96:148–156, Citeseer
12. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139
13. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A (2008) Rusboost: improving classification performance when training data is skewed. In: 2008 19th international conference on pattern recognition, pp 1–4
14. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844
15. Schaefer G, Stich M (2003) Ucid: an uncompressed color image database. *Electr Imaging* 2004:472–480
16. Bas P, Filler T, Pevny T (2011) Break our steganographic system: the ins and outs of organizing boss. International workshop on information hiding, pp 59–70

Creation and Segmentation of Image Dataset of Mung Bean Plant Leaf



Akruti Naik , Hetal Thaker , and Nirav Desai

Abstract Automated plant disease identification is an enduring research subject. Leaves are available for most of the season, and they have a flat (2d) surface that is why practically, it is phisible to detect disease symptoms using image analysis. Data collection and pre-processing are the most significant and crucial stages to obtain the data that can be taken as accurate and appropriate for further processing. Machine learning techniques require a large amount of data for training. The present paper focuses on process standardization for the creation of an image dataset of Mung bean plant leaves and pre-processing steps to enhanced captured images. The diseases in leaves result in loss of economic and production status in the agricultural industry worldwide. The identification of disease in leaves using image processing reduces the reliance on the farmers for the safeguard of agricultural crops. In this paper, creation and segmentation process of Mung bean plant leaf is performed. Present dataset will be available to be used by researchers to save their time, efforts, and cost related to dataset creation. Segmentation of images will intensify the accuracy of the identification of various diseases.

Keywords Mung bean · Leaf · Image analysis · Image dataset · Disease identification · Pre-processing · Segmentation

1 Introduction

Pulses play important role in nutritional requirements. Pulses help to reduce inanition among the poor masses. They provide minerals, vitamins, energy, dietary fiber, and the protein required for the health condition. Pulses contain substantial amounts of essential nutrients like calcium, iron, and lysine [1]. Latest research studies suggested that consumption of pulses may have likely health benefits as well as reduced risk of hypertension, gastrointestinal disorders, cardiovascular diseases, cancer, diabetes, and osteoporosis [2].

A. Naik · H. Thaker · N. Desai

Department of Computer Applications, Atmiya University, Rajkot, Gujarat, India

Gaston and O'Neill [3] projected possibility of plant species identification using artificial intelligence and digital image processing techniques. Ever since many studies have proposed various methods for automated plant and plant disease identification, Rzanny et al. (2017a) explored many approaches for image acquisition and pre-processing to improve the quality of plant organ images to train classifiers for the classification process.

This paper proposes an image dataset of Mung bean plant leaves to carry out an image-based plant disease identification and classification. There are no standard plant leaves image dataset for Mung bean leaves available. The database is created manually by capturing mung leaves images using various smart mobile phones in a controlled environment. How leaf images are acquired and pre-processed does have a substantial effect on the accuracy of the classifier trained on them.

2 Literature Review

Various effective and novel methods have been projected in recent times for the automatic identification of plant and plant organ diseases. Methods are exploring visual cues present in almost all of those parts, like fruits [4–6], stems, roots [7], kernels [8], and leaves. Amruta Ambatkar et al. [9] proposed a method for rose diseases detection using an 8-connected boundary detection algorithm for edge detection. Sannakki et al. [10] compared binary morphology and Sobel edge detector algorithms that detect edges and proved that morphology is more effective compared to others. Sabu et al. [11] used histogram of oriented gradients (HoG) and speeded up robust features (SURF) together with a KNN classifier to identify plants. Wang et al. [12] aimed at a new algorithm that segments a single leaf from real-time video and achieved clear and accurate edges. Kumar et al. [13] conducted the research that considered both front and backside of leaves with fresh and dried leaves and extracts features and tests them using support vector machine (SVM) and multi-layered perceptron (MLP) classifiers. Dahigaonkar and Kalyane [14] done related work by extracting various features including geometric, texture, shape, and color using SVM classifier. Nisale et al. [15] achieve 93% accuracy by extracting geometric features of a leaf for detecting various stages and deficiencies in the plant. Arivazhagan et al. [16] proposed an algorithm that detects and classifies an unhealthy region of leaves and segmented only diseased region with the help of an SVM classifier and obtained 94.74% accuracy. Venkataraman and Mangayarkarasi [17] perform classification and identification of plants using various statistical parameters, texture features, and SVM. Aitwadkar et al. [18] used artificial neural network (ANN) for automatic identification of plants. Batvia et al. [19] used convolution neural network (CNN) for automatic identification of plants.

A detailed study of the research work done during the last few years on leaf images is summarized in Table 1. From the information presented in Table 1, main point noticeable is that researches in the field of plant disease identification mostly focus on a single plant organ leaf. Also, the researchers are forming a custom dataset for

Table 1 Summarize of the researches carried out in recent times

Researchers	Culture	Primary feature	No. of images considered	Plant organ	Classifier/Techniques	Image acquisition/Dataset	Accuracy
(Narmadha and Arulvadiu)	Paddy	Shape, color	NA	Leaf	K-means	Custom (smartphones or digital camera)	NA
Hidayatuloh et al. [20]	Tomato	Color	1400	Leaf	CNN	Custom (smartphone)	86.92%
Kawacher Ahmed et al. [21]	Rice	Color	480	Leaf	Decision tree	Existing ("Rice leaf diseases dataset," https://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases)	97.91%
Le et al. [22]	Canola radish and barley	Texture	30,000	Leaf	SVM	Custom (on-semi VITA 2000 camera sensor)	91.85%
Sridhathan et al. [23]	Multi-species	Color	NA	Leaf	K-mean	Custom (digital camera or mobile phone)	98.27%
Dhingra et al. [24]	Basil	Color	400	Leaf	SVM	Custom (EOS 5D Mark III, 22.3 megapixel CMOS sensor)	98%
Saleem et al. [25]	Multi-species	Color	1600 625	Leaf	KNN	Existing (Flavia)	97.6%
Sun [26]	Tea plant	Texture	1308	Leaf	SVM	Custom (digital SLR camera)	96.1% 98.5%

(continued)

Table 1 (continued)

Researchers	Culture	Primary feature	No. of images considered	Plant organ	Classifier/Techniques	Image acquisition/Dataset	Accuracy
Sivasakthi [27]	Greenhouse crop	Color, texture	NA	Leaf	SVM, ANN	Custom (camera)	92% 87%
Majid et al. [28]	Rice	Color	NA	Leaf	PNN	Custom	91.46%
Arvind et al. [29]	Maize	Texture	2000	Leaf	Multiclass SVM	Existing (Plant Village)	83.7%
Suryawati et al. [30]	Tomato	Color	18,160	Leaf	CNN	Existing (Plant Village)	94%
Suresha et al. [31]	Rice	Color	NA	Leaf	KNN	Custom (digital camera)	76.59%
Saradhambal et al. [32]	Multi-species	Color	75	Leaf	k-means	Custom	NA
Tucker et al. [33]	Sunflower and oat	Shape	40	Leaf	Thresholding	Custom (TMC-76 color CCD)	NA
Zhang et al. [34]	citrus	Color, texture	500	Leaf	AdaBoost	Custom (digital camera)	88%
Wang et al. [35]	Wheat and grape	Color, texture, and shape	185	Leaf	PNN	Custom (digital camera)	94.29%
Zhang et al. [36]	Cucumber	Color	100	Leaf	SVM	Custom	92% Approx
Quin et al. [37]	Alfalfa	Color, texture, and shape	899	Leaf	SVM	Custom (digital camera)	80% Approx
Dey et al. [38]	Betel vine	Color	12	Leaf	Otsu	Custom	NA
Es-saady et al. [39]	Vegetable crop	Color, texture, and shape	284	Leaf	SVM	Custom (digital camera)	87.805

(continued)

Table 1 (continued)

Researchers	Culture	Primary feature	No. of images considered	Plant organ	Classifier/Techniques	Image acquisition/Dataset	Accuracy
Ali et al. [40]	Citrus	Color and texture	199	Leaf	Bagged tree classifier	Custom (DSLR camera)	99.9%
Tippannavar et al. [41]	Multi-species	Color	500	Leaf	KNN, PNN	Custom (digital camera)	75.04% 71.24%
Kaur et al. [42]	Multi-species	GLCM Features	NA	Leaf	SVM	NA	95.16–98.38%
Mondal et al. [43]	Okra and bitter gourd	Texture	79(Okra) 75(Bitter gourd)	Leaf	Naïve Bayes classifier	Custom (digital camera)	NA
Ma et al. [44]	Cucumber	Color	93	Leaf	Color map	Custom (digital camera)	NA
Al-Otaibi et al. [45]	Basil and parsley	Statistical feature	30	Leaf	NN	Custom (digital camera)	80%
Manimegalai et al. [46]	Apple	GLCM features	NA	Leaf	SVM	NA	98.46%
Chouhan et al. [47]	Plant leaf	Region growing	276	Leaf	NN	Existing (Plant Village)	86.21%
Zhang et al. [48]	Apple and cucumber	Color	150 (apple) 150 (cucumber)	Leaf	k-means	Custom	90.43% (Apple) 92.15% (Cucumber)
Picon et al. [49]	Wheat	Color	8178	Leaf	Deep convolution	Custom (mobile phones)	>98%
Junior et al. [50]	Multi-species	Shape	600	Leaf	RNN	NA	88.92%

(continued)

Table 1 (continued)

Researchers	Culture	Primary feature	No. of images considered	Plant organ	Classifier/Techniques	Image acquisition/Dataset	Accuracy
Sunny et al. [51]	Citrus	Texture	100	Leaf	SVM	Custom (digital camera)	NA
Nababa et al. [52]	Oil palm	Probability function	NA	Leaf	Naïve Bayes	NA	80%
Fuentes et al. [53]	Tomato	Color	5000	Leaf	NN	Custom (digital camera)	96%
Sabu et al. [11]	Multi-species	SURE, HOG	200	Leaf	KNN	Custom	NA
Vijayashree and Gopal [54]	Multi-species	Texture	127	Leaf	Dissimilarity	Custom	NA
Pushpa et al. [55]	Multi-species	Shape and edge	208	Leaf	NA	Custom	93.75%
Kumar and Talasila [56]	Multi-species	Shape, texture, and color	500	Leaf	Unique ID	Custom	NA
Kumar et al. [13]	Multi-species	Color and texture	1200	Leaf	SVM	Custom (scanned images)	94%
Dahigaonkar and Kalyane [4]	Multi-species	Color, texture, and shape	128	Leaf	SVM	Custom	96.66%
Venkataraman and Mangayarkarasi [7]	Multi-species	Texture	260	Leaf	SVM	Custom	NA
Aitwadkar et al. [18]	Multi-species	Edge, color	50	Leaf	ANN	Custom	75%
Batvia, et al. [19]	Multi-species	Shape	4000 approx	Leaf	CNN	Custom	NA
Venkataraman and Mangayarkarasi [57]	NA	Shape	5	Leaf	ANN, SVM	Custom	NA
Arun and Christopher Durairaj [58]	Multi-species	Color and texture	250	Leaf	SVM	Custom (digital camera)	98.7%

SVM support vector machine, ANN artificial neural networks, PNN probabilistic neural networks, KNN k-nearest neighbors, CNN convolutional neural network

Table 2 Existing plant image datasets

Dataset	Organ	No. of species	Culture	No. of images
Flavia	Leaf	32	Multi-species	1907
Plant Village	Leaf	3	Bell paper, potato, tomato	15,442
Oxford_flower102	Flower	102	Flowers	7000 +
Swedish	Leaf	15	15 tree classes	1125
New Plant Disease	Leaf	14	Fruits and vegetables	87,000
Coffee-dataset	Leaf	1	Coffee	1747

their research work as there is no standard dataset available for Mung bean plant organs. The abbreviations used are summarized in the last row of Table 1. Table 2 contains a list of some existing plant image datasets.

The main point to note in Table 2 is that none of the above plant organ image datasets are dedicated to the Mung bean plant leaf organ. This research addresses the need for a benchmark dataset for Mung bean plant organs.

3 Materials and Methods

3.1 Dataset Collection

The crucial necessity for accurate plant disease identification is a standard dataset of plant organ images. The dataset creation consists of stages as follows:

- Plant selection
- Capturing images
- Dataset creation

For this research, the Mung bean plant is under consideration as it is a local crop of the South Gujarat Region. In the present work, the leaf dataset consists of four types of healthy and diseased Mung bean leaf images; these are Cercospora Leaf Spot, Yellow Mosaic Virus, and Powdery Mildew. These were collected from the Navsari Agriculture University at Navsari, Gujarat, India, for reflective study. A pictorial assessment of the above-mentioned study site is shown in Fig. 1.

Leaf samples are acquired indoor to minimize the effect of lighting conditions. Leaves were digitally captured in a controlled environment using Oppo A5 13 MP and MI Note 8 Pro 64 MP smartphones.

The database consists of 1500+ images which include 400+ healthy and 1000+ diseased leaves. The diseases considered are Cercospora Leaf Spot, Powdery Mildew, and Yellow Mosaic Virus. Figure 2 represents the healthy and diseased Mung bean leaves.



Fig. 1 Study site of Mung bean plants



Fig. 2 Healthy and diseased Mung bean leaves

3.2 System Model and Discussion

The system model is consisted of four crucial steps as follows:

- (1) Pre-processing: Pre-processing helps to bring out useful information from an image.

- (2) **Segmentation:** Segmentation is used for locating objects in the image and to detect bounding lines of the image, background subtraction.
- (3) **Feature extraction:** In this phase, unique characteristics of an object or group of objects are collected.
- (4) **Classification:** Classification is the phase where training and testing take place. It is where the decision takes place using features extracted from the previous phase.

From the above four phases first, two phases have been discussed in detail in the following sub-sections, and the remaining two phases will be implemented in the future. For implementation, OpenCV an open-source computer vision library with Python is used.

- (a) **Pre-processing:** After image acquisition, the pre-processing phase takes place. In this phase, image enhancement will be done. For this, various operations are carried out in a series: RGB image acquisition and color transformation, normalization/ resize of image size, augmentation, masking green pixels, and segmentation. This phase makes changes in the image and makes it appropriate for segmentation.

Resize an image

Resizing refers to the scaling of an image. It helps to reduce or increase number of pixels from an image. Figure 3 represents the image resize phase.

Augmentation

Augmentation encompasses a wide range of techniques used to generate new training samples from the original ones. It helps us to increase the size of the dataset for training. Image augmentation artificially creates training images through a combination of multiple transformations. The result of image augmentation is displayed in Fig. 4.

- (b) **Segmentation:** Image segmentation is the first step in image analysis and pattern recognition; it is a critical and essential step and is one of the most



Fig. 3 **a** Original Image, **b** resized image

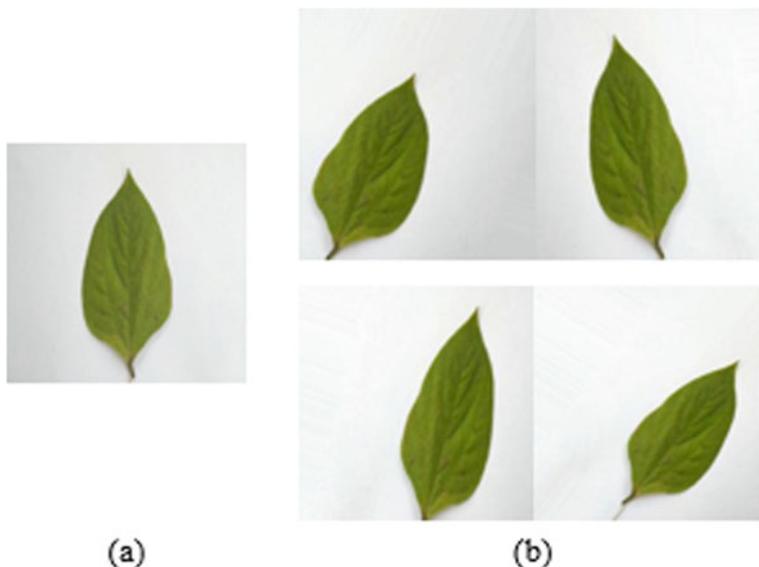


Fig. 4 **a** Original image, **b** augmented images

difficult tasks in image processing, as it determines the quality of the final result of the analysis [59]. During the segmentation phase, the image will be divided into several segments so that the analysis process becomes easy. In this study, edge detection is performed using the canny() edge detector and interactive foreground extraction is performed using Grebcut() algorithm. Figure 5 depicts the edge detection, and Fig. 6 depicts the foreground extraction process.

Steps for segmentation

The GrabCut algorithm segments object from the background in an image. The user has to mark a rectangular area as the primary input. The outer part of this rectangle is considered as background, and pixels in the outside area are considered as known background and inside are unknown background. A model is then created using this data to find out whether the unknown pixels are foreground or background. Figure 7 represents some of the segmented images.

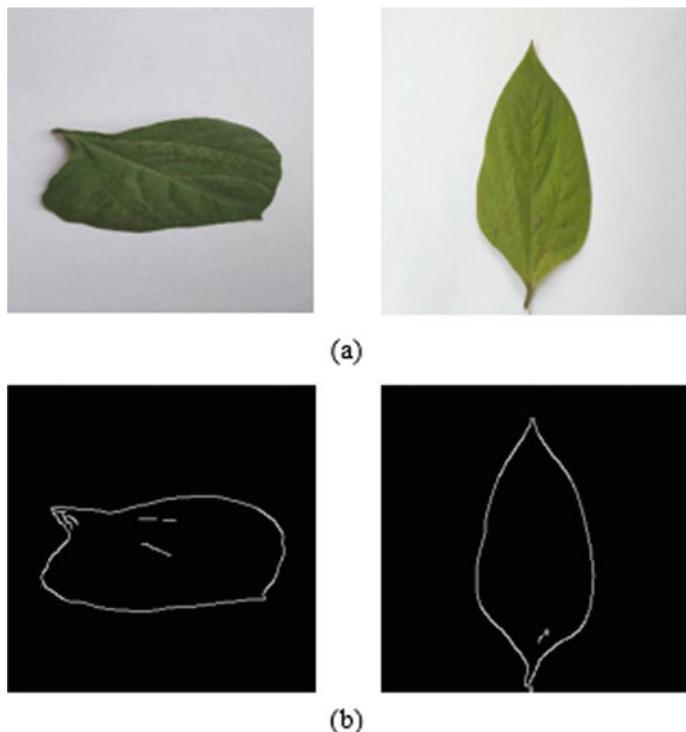


Fig. 5 **a** Original images, **b** extraction of boundary

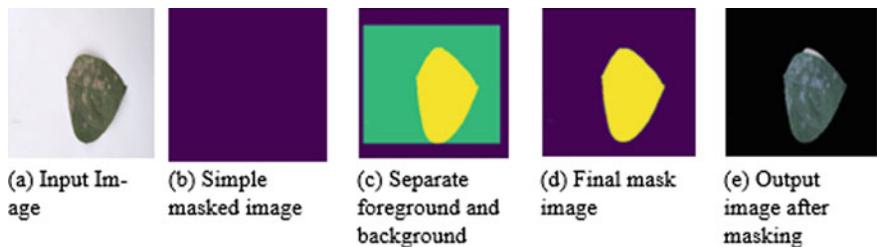


Fig. 6 Segmentation process

GrabCut is one of the extensively used algorithms for removing background in images. The automatic GrabCut technique was experimentally tested using a dataset of Mung bean leaf images as shown in Fig. 7. This work can be used in regions like plant leaf image classification and plant leaf disease detection from plant leaf images.

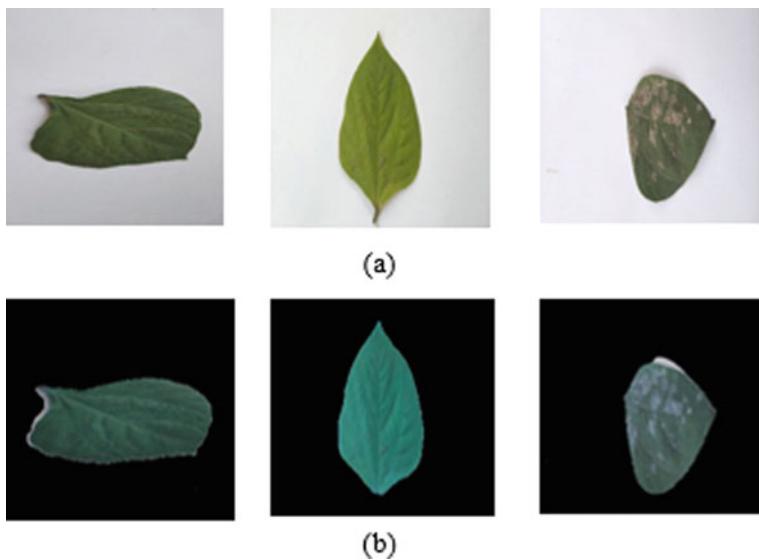


Fig. 7 **a** Original images, **b** segmented images

4 Conclusion

We considered the creation of the Mung bean plant organ image dataset. Dataset will be released to be used by researchers to save their time, efforts, and cost associated with dataset creation. Segmentation of the image will increase the accuracy of identification of healthy and diseased pixels.

References

1. Gowda CLL, Srinivasan S, Gaur PM, Saxena KB (2013) Enhancing the productivity and production of pulses in India. In: Shetty PK, Ayyappan S, Swaminathan MS (eds) Climate change and sustainable food security. National Institute of Advanced Studies, Bangalore and Indian Council of Agricultural Research, New Delhi, pp 63–76
2. Jacobs DR, Gallaher DD (2004) Whole-grain intake and cardiovascular disease: a review. *Current Atheroscler* 6:415–423
3. Gaston KJ, O'Neill MA (2004) Automated species identification: why not? *Phil Trans R Soc Lond B* 359:655–667. <https://doi.org/10.1098/rstb.2003.1442>
4. Aleixos N, Blasco J (2002) Multispectral inspection of citrus in real-time using machine vision and digital signal processors. *Comput Electron Agric* 33(2):121–137
5. Corkidi G, Balderas-Ruiz KA (2005) Assessing mango anthracnose using a new three-dimensional image-analysis technique to quantify lesions on fruit. *Plant Pathol* 55(2):250–257
6. López-García F, Andreu-García G (2010) Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach. *Comput Electron Agric* 71(2):189–197
7. Smith SE, Dickson S (1991) Quantification of active vascular-arbuscular mycorrhizal infection using image analysis and other techniques. *Aust J Plant Physiol* 18(6):637–648

8. Ahmad IS, Reid J (1999) Color classifier for symptomatic soybean seeds using image processing. *Plant Disease* 83(4):320–327
9. Ambatkar A, Bhandekar A, Tawale A, Vairagade C, Kotamkar K (2017) Leaf disease detection using image processing. In: Proceedings of international conference on recent trends in engineering science and technology, Vol 5, pp 333–336
10. Samnakki SS, Rajpurohit VS, Birje SJ (2012) Comparison of different leaf edge detection algorithms using fuzzy mathematical morphology. *Int J Innov Eng Technol* 1(2):15–21
11. Sabu A, Sreekumar K, Nair R (2017) Recognition of ayurvedic medicinal plants from leaves: a computer vision approach. In: Fourth international conference on image information processing, pp 574–578
12. Wang J, Jianlei He Y, Han CO, Li D (2013) An adaptive thresholding algorithm of field leaf image. *Comput Electron Agric* 96:23–39
13. Kumar MP, Surya CM, Gopi VP (2017) Identification of ayurvedic medicinal plants by image processing of leaf samples. In: Third international conference on research in computational intelligence and communication networks, pp 231–238
14. Dahigaonkar T, Kalyane R (2018) Identification of ayurvedic medicinal plants by image processing of leaf samples. *Int Res J Eng Technol* 5(5):351–355
15. Nisale SS, Bharambe CJ, More VN (2011) Detection and analysis of deficiencies in groundnut plant using geometric moments. *Proc World Acad Sci Eng Technol* 5:512–516
16. Arivazhagan S, Newlin Shebiah R, Ananthi S, Vishnu Varthini S (2013) Detection of unhealthy region of plant leaves and classification of plant diseases using texture features. *Agric Eng Int: CIGR J* 15(1):211–217
17. Venkataraman D, Mangayarkarasi N (2017) Support vector machine-based classification of medicinal plants using leaf features. In: International conference on advances in computing, communications, and informatics, pp 793–798
18. Aitwadkar PP, Deshpande SC, Savant AV (2018) Identification of Indian medicinal plant by using artificial neural network. *Int Res J Eng Technol* 5(4):1669–1671
19. Batvia V, Patel D, Vasant AR (2017) A survey on ayurvedic medicine classification using tensor flow. *Int J Comput Trends Technol* 53(2):68–70
20. Akbar Hidayatuloh MN (2018) Identification of tomato plant diseases by leaf image using SqueezeNet model. In: International conference on information technology systems and innovation (ICITSI) Bandung - Padang
21. Kawcher Ahmed TR (2019) Rice leaf disease detection using machine learning techniques. In: International conference on sustainable technologies for Industry 4.0 (STI). IEEE, Dhaka, pp 1–5
22. Le VN, Apopei B, Alameh K (2019) Effective plant discrimination based on the combination of local binary pattern operators and multi-class support vector machine methods. *Inf Process Agricult* 6:116–131
23. Sridhathan C, Senthil Kuma M (2018) Plant infection detection using image processing. *Int J Mod Eng Res (IJMER)* 8(7):13–16
24. Dhingra G, Kumar V (2019) A novel computer vision-based neutrosophic approach for leaf disease identification and classification. *Measurement* 135:782–794
25. Saleem G, Akhtar M (2019) Automated analysis of visual leaf shape features for plant classification. *Comput Electron Agric* 157:270–280
26. Y Sun, Jiang Z (2019) SLIC_SVM based leaf diseases saliency map extraction of the tea plant. *Comput Electr Agricult* 157:102–109
27. Sivasakthi S (2020) Plant leaf disease identification using image processing and SVM, ANN classifier methods. In: International conference on artificial intelligence and machine learning (J Anal Comput (JAC))
28. Kholis Majid YH (2013) I-PEDIA: mobile application for paddy disease identification using fuzzy entropy and probabilistic neural network. In: ICACSIS. IEEE, pp 403–406
29. Arvind KR, Raja P, Mukesh KV, Anirudh R, Ashwin R, Szczepanski C (2018) Disease classification in Maize crop using a bag of features and multiclass support vector machine. In: Proceedings of the second international conference on inventive systems and control (ICISC)

- 2018) IEEE Xplore Compliant—Part Number: CFP18J06-ART, ISBN: 978-1-5386-0807-4; DVD Part Number: CFP18J06DVD, ISBN: 978-1-5386-0806-7
- 30. Suryawati E, Sustika R, Sandra Yuwana R, Subekti A, Pardede HF (2018) Deep structured convolutional neural network for tomato diseases detection. In: ICACSIS 2018 978-1-7281-0135-4/18/S31.00 ©2018 IEEE
 - 31. Suresha M, Shreekanth KN, Thirumalesh B (2017) Recognition of diseases in paddy leaves using kNN classifier. In: 2nd International conference for convergence in technology (I2CT)
 - 32. Saradhambal G, Dhivya R, Latha S, Rajesh R (2018) Plant disease detection and its solution using image classification. Int J Pure Appl Math 119(14)
 - 33. Tucker CC, Chakraborty S (1997) Quantitative assessment of lesion characteristics and disease severity using digital image processing. J Phytopathol 145(7):273–278
 - 34. Zhang M, Meng Q (2011) Automatic citrus canker detection from leaf images captured in the field. Pattern Recog Lett 32(15):2036–2046
 - 35. Wang H, Li G, Ma Z, Li X (2012) Image recognition of plant diseases based on principal component analysis and neural networks. In: Proceedings of the IEEE International conference on Natural computation (ICNC), pp 246–251
 - 36. Zhang S, Wang Z (2016) Cucumber disease recognition based on Global-Local Singular value decomposition. Neurocomputing 205:341–348
 - 37. Qin F, Liu D, Sun B, Ruan L, Ma Z, Wang H (2016) Identification of alfalfa leaf diseases using image recognition technology. PLoS ONE 11(12):1–26
 - 38. Dey K, Sharma M, Meshram MR (2016) Image processing based leaf rot disease, detection of betel vine (*Piper betle* L.). In: Proceedings of the international conference on computational modeling and security (CMS), pp 748–754
 - 39. Es-saady Y, El Massi I, El Yassa M, Mammass D, Benazoun A (2016) Automatic recognition of plant leaf diseases based on a serial combination of two SVM classifiers In: 2016 international conference on electrical and information technologies (ICEIT), pp 561–566
 - 40. Ali H, Lali MI, Nawaz MZ, Sharif M, Saleem BA (2017) Symptom-based automated detection of citrus diseases using color histogram and textural descriptors. Comput Electron Agric 138:92–104
 - 41. Tippannavar S, Soma S (2017) A machine learning system for recognition of vegetable plant and classification of abnormality using leaf texture analysis. Int J Sci Eng Res 8(6):1558–1563
 - 42. Kaur P, Singla S, Singh S (2017) Detection and classification of leaf diseases using an integrated approach of support vector machine and particle swarm optimization. Int J Adv Appl Sci 4(8):79–83
 - 43. Mondal D, Kole DK, Roy K (2017) Gradation of yellow mosaic virus disease of okra and bitter gourd based on entropy-based binning and Naive Bayes classifier after identification of leaves. Comput Electron Agric 142:485–493
 - 44. Ma J, Du K, Zhang L, Zheng F, Chu J, Sun Z (2017) A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. Comput Electron Agric 142:110–117
 - 45. AL-Otaibi MB, Ashour AS, Dey N, Abdullah R, AL-Nufaei AA, Fuqian S (2017) Statistical image analysis based automated leaves classification. In: Proceedings of the 2nd International conference on information technology and intelligent transportation systems (ITITS), vol 296, p 469
 - 46. Manimegalai S, Sivakamasundari G (2017) Apple leaf diseases identification using support vector machine. In: Proceedings of the international conference on emerging trends in applications of computing (ICETAC), pp 1–4
 - 47. Chouhan SS, Kaul A, Singh UP, Jain S (2018) Bacterial foraging optimization based Radial Basis Function Neural Network (BRBFNN) for identification and classification of plant leaf diseases: an automatic approach towards plant pathology. IEEE Access 6:8852–8863
 - 48. Zhang S, Wang H, Huang W, You Z (2018) Plant diseased leaf segmentation and recognition by fusion of superpixel K-means and PHOG. Optik 157:866–872
 - 49. Alvarez-Gila PA, Seitz M, Ortiz-Barredo A, Echazarra J (2018) Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Comput Electron Agric

50. Junior JJDMS, Backes AR, Bruno OM (2018) Randomized neural network-based descriptors for shape classification. *Neurocomputing* 312:201–208
51. Sunny S, Gandhi MPI (2018) An efficient citrus canker detection method based on contrast limited adaptive histogram equalization enhancement. *Int J Appl Eng Res* 13(1):809–815
52. Nababa M, Laia Y, Sitanggang D, Sihombing O, Indra E, Siregar S, Purba W, Mancur R (2018) The diagnose of oil palm disease using Naive Bayes method based on expert system technology. *J Phys Conf Ser* 1007(1):1–5
53. Fuentes AF, Yoon S, Lee J, Park DS (2018) High-performance deep neural network-based tomato plant diseases and pests diagnosis system with a refinement filter bank. *Front Plant Sci* 9
54. Vijayashree T, Gopal A (2017) Leaf identification for the extraction of medicinal qualities using image processing algorithm. In: International conference on intelligent computing and control, Coimbatore, pp 1–4.<https://doi.org/10.1109/I2C2.2017.8321884>
55. Pushpa BR, Anand C, Nambiar Mithun P (2016) Ayurvedic plant species recognition using statistical parameters on leaf images. *Int J Appl Eng Res* 11(7):5142–5147
56. Kumar SE, Talasila V (2014) Leaf features-based approach for automated identification of medicinal plants. In: International conference on communication and signal processing, pp 210–214.<https://doi.org/10.1109/ICCS.2014.6949830>
57. Venkataraman D, Mangayarkarasi N (2016) Computer vision based feature extraction of leaves for identification of medicinal values of plants. In: IEEE International conference on computational intelligence and computing research, pp 1–5
58. Arun C, Christopher Durairaj D (2017) Identifying medicinal plant leaves using textures and optimal colour spaces channel. *Jurnal Ilmu Komputer dan Informasi* 10(1):19–28. <https://doi.org/10.21609/jiki.v10i1.405>
59. Hambarde SM, Jagtap SB (2014) Agricultural plant leaf disease detection and diagnosis using image processing based on morphological feature extraction. *IOSR J VLSI Sign Process (IOSR-JVSP)* 4(5)

Road Accident Analysis Using ML Classification Algorithms and Plotting Black Spot Areas on Map



Manu Tiwari, Piyush Nagar, Gautam Arya, and Surendra Singh Chauhan

Abstract This paper mainly deals with analyzing the road accident properly on each parameter/factor that causes accidents, i.e., speed of the vehicle, road's surface conditions, weather, age of the driver, type of road junctions, etc., at the city, district, state/UT's level and try to identify black spot areas and various flaws in road infrastructures. Road accidents are emerging as one of the major life-threatening causes in various countries. It was also claimed that roads are becoming deadlier than AIDS due to emerging technologies that are making vehicle average speed faster than ever before and also roads are one of the major causes of death of youngsters and thus affect the economic development of the country very badly. Our analysis showed us that Friday is the day in which most of the accidents occurred, accidents in which motorcycle is involved are more deadly than that of a light motor vehicle, and also, various results show us that men are mostly the victims of accidents than women. We did this analysis because there is a need to wake government agencies in order to remove these flaws also because saving lives is one of the most important duties of mankind, and we can also prevent million-dollar loss not only per individual but also at the national level. We can also give our next generation safer roads as it was predicted that by 2025 0.2 million deaths will be occurring yearly. Hence, this paper mainly brings light on the main factors causing accidents and areas of improvement by identifying accident-prone areas.

Keywords Road safety · Road accident · Fatality rate · Traffic volume · Severity analysis · Identification using ML

1 Introduction

We started this project, as there is a very worsening situation in road accident fatalities not only in India but also outside India. Also, there is a need to develop some software which cannot only help in detecting accident severity but also guides the government

M. Tiwari (✉) · P. Nagar · G. Arya · S. S. Chauhan
School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

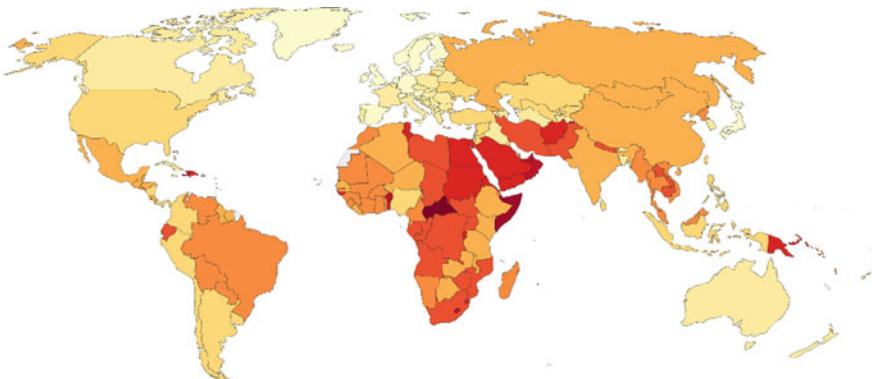


Fig. 1 Death rate from road accidents across the globe

about the flaws in the architecture of roads. The goal of this project is to investigate and determine what causes accidents and what attributes to their level of severity in the highly accident-prone areas. Through visualizations and machine learning algorithms, areas of concern will be highlighted, and the seriousness of accidents will be predicted as accurately as possible in the black spot areas. With the help of this project, we can figure out the different flaws in our roads also so that government can improve the infra and prevent accidents in the upcoming future and by doing so this project can save many lives.

Reference [1]. The figure (see Fig. 1) shows road accident death across the globe. (Legends-Red: Extremely high death date, Orange: Above average death rate, Yellow: Average death rate, Light Yellow: Less death rate).

2 Literature Survey

A review of existing literature provided insight into the dimensions of research done on road accidents across the globe. Literature extensive studies have been conducted relating to the type of peoples living in a particular area and act as aspects of road crashes and development of various basic needs to prevent an accident such as traffic signals. Literature showed us that countries which are well versed financially did the regular survey on road safety and genuine modification needed but on the other hand, countries which are still struggling with their financial/economic issues found it hard to spend money in this area and hence did very less work on road safety.

2.1 *Investigational Studies*

Jacobs along with Sayer [2] did an investigational study on various countries and concluded that traffic volume did not play a vital role in deadly accident as it was observed that both countries with very dense traffic and sparse traffic are facing accidents and also found that it is how a country is implementing its traffic violation rules and how much they are serious in developing the infrastructure to make their countries road network safer.

2.2 *Injury Related Studies*

McLeod et al. [3] stated in his research that accident is a result of various factors occurring simultaneously, and these factors include both violations of general traffic rules by taking an unnecessary risk, driving when physically unfit, and under the influence of various drugs, locality (the type of peoples present there) and this is stated as ‘adventurous risk’ and ‘health risk’.

2.3 *Demographic Studies on Road Accidents*

- Ray et al. [4] conducted a study and analyze that for the case where senior citizens involved in accidents were not due to alcohol or any drug and even the frequency of accident is not playing a vital role in their accidents.
- Stelmach and Nahom [5] reviewed literature and documented that with increment in age driving skills of the driver also declined as they are not enough active that they can react to a sudden uncertain obstacle.
- Chipman et al. [6] did a study on crash rates in Ontario, based on the type of drivers involved, driver-kilometers, and driver-days using survey estimates of time, distance driven, and the frequency of traffic crashes. Rates by age groups, the gender of the driver, and location were computed for almost all crashes and for accidents that result in severe injuries or fatality. Young male drivers and older women were found to be at higher risk under specific conditions.
- Yagil [7] stated that gender had been playing a vital role on roads as it was found that males are careless about traffic rules and are 100% more involved in accidents than the number of females involved, i.e., twice.
- Mayhew et al. [8] stated that senior citizens are not good at taking a left turn as they succumb to the right side of the way also, they are the one who majorly crashed at the junction.
- Lonczak et al. [9] shown that various mental situations (such as anger, attention/focus) are playing a major role in predicting the accident’s likelihood also

men are the ones who mainly become the victim of road accidents but their mental situation is same as that of the female who was involved in crashes.

2.4 Behavioral Studies on Road Accidents

- Hunt et al. [10] studied the effect of dementia on driving skills and discovered that it is strongly positively correlated with driving skills. He found that the recently passed candidate in driving test may not be friendly with learning/experience-based skills and how much attention is required to drive safely but an experienced driver can contain both the qualities.
- Turner et al. [11] reported youngsters lost their lives due to reckless driving nature.
- Factor et al. [12] examined the differences in road accident involvement, attitudes, and behaviors related to driving across gender, age, ethnicity, and socioeconomic groups.
- There was evidence that getting angry while driving affects the driving ability of a driver very badly and most of the accidents occurred under this mental situation, and also, there is a case that many such awful incidents prevented at the last moment, and if we count them as an accident, then this factor becomes one of the major cause of accidents [13]. As we all know about human behavior, they learn mostly by observing, and thus, it was found that most of the drivers learn various reflexes from their parents and relatives as stated by Refs. [14–16].
- Lee and Fazio [17] analyzed the factors that influenced the performance of emergency management personnel in the event of the occurrence of a freeway traffic crash using a Cox regression model.

2.5 Predictive Research on Road Accidents

- Abdel Wahab and Abdel-Aty [18] investigated the major effect of increasing in the count of light truck vehicles (LTV's) on fatalities that result from head-on collisions in the USA. Their forecast, using a time series model, was that by the year 2010, there will be an increase of 8% in the annual deaths due to head-on collisions in comparison with the data relating to the year 2000.
- Abdel Wahab and Abdel-Aty [19] also undertook a study on backside collisions involving LTV's using a transfer function time series model which predicted 5% increase in annual deaths of passengers in vehicles due to rear-end collisions by the year 2010.

2.6 *Advances in Road Accident Research*

- Harnen et al. [20] worked in the field of two-wheeler crashes he plotted the two-wheeler accidents with the type of junction where they occurred and found that two-wheelers are positively correlated with traffic density on a particular junction, i.e., more the vehicle on a particular place more the likelihood of accidents (two-wheeler), his whole analysis was based on Malaysian roads accidents.
- Kececi and Tao [21] in their analysis based on two types of dataset, i.e., one in which road surface conditions, are known and the other in which nothing is known about the road conditions in order to analyze the stability of LMV as well as a motorcycle.
- Chang and Wang [22] developed a special model called as classification and regression tree that is used to identify the similarities between an accident and environmental conditions during that period.

2.7 *Factors Associated with Road Accidents*

- Livneh and Hakkert [23] studied about dynamic changes in the factors which are causing road accidents so they started a survey in Israel in which they found that: (a) continuous increment in vehicles, (b) total distance traveled, (c) infrastructure/road conditions involved, (d) the average number of deaths per accident, (e) comparative increment of motorization with respect to severity, (f) causalities per vehicle per population of a particular area. Murray and Whiteing [24] in their thesis concluded about various methods by which we can prevent accidents as well as can reduce economical loss along with this main focus is also on how much expert a driver was and what are the necessary steps to be taken in order to reduce human error at driving seat. Red zone areas are the areas where likelihood of an accident is greater as stated by Ref. [25]. They also worked in this field in order to identify various factors which are still not counted as one of the major factors but they are playing a vital role in road crashes some of them are road build quality, climatic changes, age band of the driver, safety precautions are taken or not, etc.
- A transition condition, which alternates between free-flowing and congested conditions, contributes significantly to a road accident. However, there is a wide range of factors, as evident from other models [26] that contribute to road accidents such as human factors, geographical factors, infrastructure, information, and legal factors. Factors that pose a threat to road users are not available for Indian driving conditions [27]. There is a need to identify factors based on the vehicle user's rating of potential variables that cause accidents in a developing country. Motor vehicle users in a city of the USA rated the services at various distributed locations (where they were asked) [28].

2.8 Governmental Studies

UK Government. The government found that most of the road accidents are repeating themselves but with different vehicles and was found that they have some common conditions also and was said that near about 1.7 k peoples witnessed same causes of accident again and again. About 0.16 million deaths are reported in road accidents which are less than that of 2012.

USA Government. [29] USA government predicted and stated that by the end of 2020 more people will die due to road crashes than that of deadly diseases like AIDS. Analysis of age group most affected in the USA by road accident (see Fig. 2).

Indian's Analysis. The following researches were done on the nature of injuries and the severity levels mostly by the government of India but some others also did some work in that namely:

- Sahdev et al. [30] in Delhi, performed a study of road accident fatalities to determine about medical help in case of a road crash and found that most of the injured people can be cured but due to the lack of infrastructure and proper medical help they lost their lives.
- To make roads safer various researchers tried to figure out the relation between speed of vehicle and distance maintained from other vehicles (just before the crash occurred) and for Indian roads Gupta and Mandloi [31] did a thesis in India in order to identify black spots areas in India where likelihood of accident is high.

Indian Government.

- Reference [32] Road accidents are increasing in India exponentially, as compared to the year 1970, road accidents in India increased by about 10 times in the year 2013. Injuries along with fatalities increased and crossed 0.14 million (yearly).
- Since 2003, death rate has increased by about 5% and is about 3.6% higher than the total population increased.



Fig. 2 Most dangerous highways in the USA

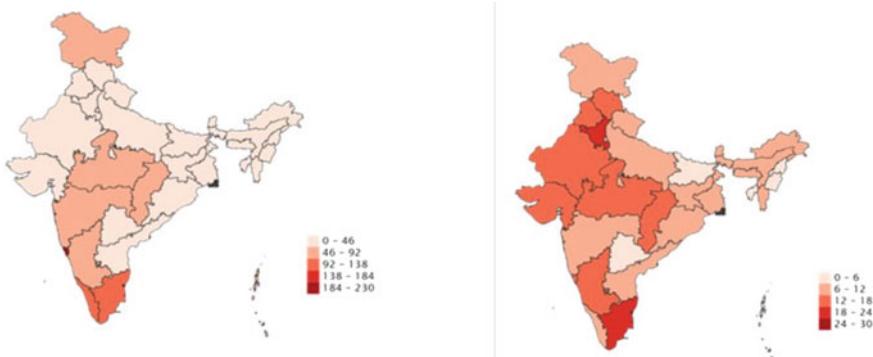


Fig. 3 Number of road accidents and number of persons killed in India (per lakh population)

- India is one of the major accidental hubs as in every 1 lakh people, 12 people died due to road accidents.
- Accidental deaths in India are 4 times higher than that of Western countries like the UK and about 2 times higher than that of Asian countries like Japan.
- In mid-2015, India reported 0.5 million road crashes which resulted in killing of about 0.05 million peoples and with the help of various agreements India is trying to reduce them by half in next couple of years.
- Also, in India infrastructure development rate is slower than that of traffic load increment this also causes congestion in the road and hence increases the likelihood of accidents.
- Highways in India are witnessing about 30% of the total accidents registered in the country.
- The below graph shows us that Goa and the southwest part of India are the most accident-prone areas (see Fig. 3).
- Mostly drivers are held responsible for accidents, and very few cases are due to violation of traffic rules.
- Near about three-fourths of accidents happened due to human error (because they are not well-trained). This includes over speeding, drunk driving along with hit and run cases.
- Near about 7.1% of road accidents caused due to second person's fault, which includes fault of cyclists, pedestrians and due to other vehicle drivers.
- It is daytime which witness most of the accidents, and youngster's age group is the one which is most vulnerable to road accidents. No matter the situation, most accidents were involved areas that were uncontrolled. One of the main areas where this happened was in the junction Detail T or staggered junction.
- Other areas of concern include accident locations that included mid-junction on roundabouts or main roads. No matter the location, details, or location of impact the common denominator seems to be a lack of signage or control in junction areas.

3 Proposed Model

In today's world, vast range of ML algorithms are available and in various categories such as (a) supervised and (b) unsupervised. In this research paper, we have worked on the analysis of the road accident, and to plot hot spot areas in order to do this we used classification algorithms of both above-mentioned types because firstly we want to analyze how data is related with each other also we have done the same on resampled data, in order to avoid bias. In this research paper, we are using K-modes (unsupervised) in order to handle mixed data and bagging classifier, AdaBoost classifier, random forest classifier (supervised).

3.1 Dataset Preparation

In order to achieve high accuracy, highly accurate dataset is required because if we work on inaccurate data, then the accuracy of our prediction will be affected, so we have prepared/processed data (see Fig. 4).

Data Cleaning. In this step, we have dropped various repetitive columns in order to reduce data redundancy and make data precise. Also, in this we have drop columns which are of no use along with manipulation in their data type format so that they can be useful for the analysis.

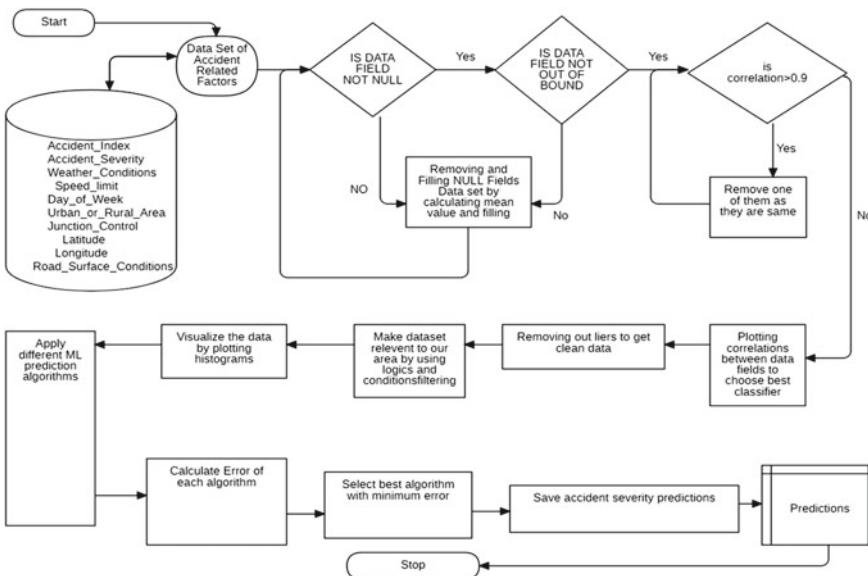


Fig. 4 Flow diagram of road accident analysis

Removal of nulls and outliers. The dataset contains columns, and some of its cells may be null which can affect our model. So, we have dropped those columns in which null values are more than 60% and process those columns with the help of statistics which are having nulls less than 60%. Also, we have plotted various correlations between various fields of dataset in order to find outliers also.

Feature manipulation/creation. In this, we have performed various mathematical/logical operations and sort out various useful concepts in our dataset which may not visible in our dataset primarily that we have collected from various sources, such as date and time are available in different columns so we have created a single column, namely ‘time_of_day,’ which contains clusters of different time divisions, created ‘age_band_of_drivers’ more condense groups for age band of driver in order to deal with some potential outliers.

3.2 Visualization

In this general visualization step, we are going to visualize the data, as per the above operations performed and try to find various correlations between them by plotting them in heatmaps, bar graphs, and try to analyze various outcomes.

3.3 Tests

Before analyzing seriousness of the accident, we want to find out which columns are mainly useful, and in order to achieve this, we have used various statistical methods/tests, such as chi-squared test; after performing this, we have taken out to perform visualization again but this time with respect to the accident seriousness.

3.4 Web Scraping

After visualizing data, we have to find some useful signs in order to find the relation between accident seriousness and various traffic rules/road conditions, we also used python library called ‘BeautifulSoap’ which helps us to scrap out only useful data from the Web sites.

3.5 Preprocessing

In this, we have gone through various encoding techniques, such as Label, One-Hot, etc., in order to filter our data. It also helps us to identify how data is responding;

i.e., it is imbalanced or balanced and makes it balanced, and also, we sort out various ML algorithms that will be effective.

3.6 Resampling/Undersampling

In this step, we have split our data, i.e., training dataset and testing dataset, with the help of inbuilt function, namely ‘train_test_split()’.

3.7 ML Algorithms

Unsupervised learning. Before we get into predictions, we are going to complete some machine learning in ordered to see how the data relates to each other. We are going to do this on the resampled data as well, in order to avoid bias. We will use two clusters which, in theory, represent the two variables for accident seriousness, not serious and serious. In this, we have used k-modes algorithm as follows as shown in Table 1. The k-modes clustering algorithm is a modified version of k-means algorithm which is one of the most popular unsupervised learning and is mainly used for partitional clustering algorithms. Huang, upgraded k-means clustering algorithm to k-modes clustering algorithm by using various groups of the categorical data [33, 34]. The advancement performed in the k-means is listed as follows: (i) usage of simple match techniques in order to analyze diversity in the objects present in data, (ii) secondly, calculate modes, and (iii) finally, update the values with recently calculated mode by using frequency models Say $y, y_{11}, y_{12}, \dots, y_{nm}$ (dataset) and we can represent it as after applying algorithm in order to group data by processing minimization in function (i.e., Eq. (1)). We can further evaluate Eq. (1) using Eqs. (2) and (3).

$$P(w, p) = \sum_{b=1}^k \sum_{a=1}^m w_{ab} d_{sim}(y_a, p_b) \quad (1)$$

where w_{ab} is an $M \times K$ matrix where each element belongs to either 0 or 1.

Also, we can define measure of dissimilarity as in equation.

$$d_{sim}(y_a, p_b) = \sum_{c=1}^n \delta(y_{ac}, z_{bc}) \quad (2)$$

Which can further evaluate by using Eq. (3).

$$\delta(y_{ac}, z_{bc}) = \begin{cases} 1 & \text{if } y_{ac} = z_{bc} \\ 0 & \text{if } y_{ac} \neq z_{bc} \end{cases} \quad (3)$$

The advantages of k-modes: (i) straightforward, (ii) efficient for handling big amount of dataset. The main disadvantages are as follows: (i) basic requirement of initialization and categories to be formed, (ii) it gives solution under some boundaries which may not be acceptable globally, and we got following outcomes on applying above algorithm as shown in Table 2.

Also, after performing k-modes we got the following visualizations after analyzing the results (see Fig. 5).

Looking at these graphs, we can see the patterns of how each category of each column pairs off with the clustering on ‘accident_seriousness’.

Supervised Learning. Now we have seen that how clusters are pairing off with data so we now try to use supervised learning and generate the outcomes, and also, we have used two types of strategies, namely resampled and undersampled. Resampled strategy can be defined as: First, we are going to run some standard classifier algorithms using the resampling method from above, gather the results of some scoring metrics (‘Accuracy,’ ‘Log Loss,’ ‘Cross-Validation,’ ‘Recall,’ ‘F1,’ ‘Error Rate’), and put those scores into a data frame, below are some ML algorithm used to predict accident severity, and also, we analyze how accurate are these:

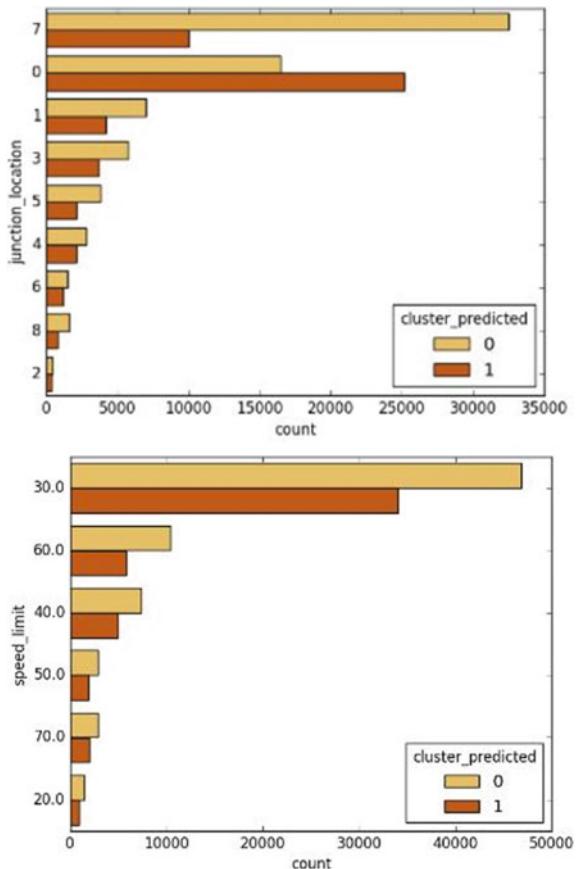
Table 1 Algorithm of k-modes

Step number	Input/output	Process performed
1	Dataset (Y) and No. of categories to be obtained (k)	Providing input to the model
2	Initialize modes for each dataset	Calculating no. of times an event occurred
3	Dissimilarities between the starting clusters and the dataset are generated	Applying Eq. (2) for evaluation
4	Analyzed correlations obtained	Applying Eq. (1) for analyzing
5	Allocate different targets of dataset to their closest centroid	Creating category of centroid using objects dissimilarities
6	New updated mode of each dataset	Just replace old mode value with the new one
7	Similarity index obtained	By comparing with new modes and repeat 4 and 5 steps
8	Required clusters are created	

Table 2 K-mode classifier outcomes

Latitude	Speed limit	Year	Month	Seriousness	Cluster predicted
52.647338	30.0	2012	3	1	0
51.395943	50.0	2015	6	1	0
51.530007	50.0	2012	10	1	0
50.723415	30.0	2014	9	1	0
51.568923	30.0	2011	1	1	1

Fig. 5 Visual analysis of k-modes algorithm outcomes



- *Bagging Classifier*. It is a classifier that uses metaestimators to construct results (predictions) by fitting the base (weak) classifiers on each randomized subsets of initial data, and by doing this, it creates a stronger estimator (which is more powerful estimator). It is generally used to reduce variance in the data.
- *AdaBoost Classifier*. AdaBoost is used to reduce the biasing in the dataset and is pronounced as adaptive boosting, adaptive means which has capability to modify itself, and hence, AdaBoost can create a stronger classifier which has greater capacity to analyze and predict more accurate results. It requires initialization of various categories with some weights as mentioned below:

$$\text{weight}(w_a) = \frac{1}{m} \quad (4)$$

where w_a = weight at ‘ a ’ training data set and m = Instance of the training going on.

Table 3 Supervised learning result analysis

Classifier	Accuracy	Log loss	Cross-validation	Recall	F1	Error rate
Bagging classifier	67.018	0.63	67.779	66.257	33.8	32.98
AdaBoost classifier	68.771	0.67	63.023	54.785	30.8	31.22
Random forest classifier	66.291	0.62	68.030	68.633	34.1	33.70

Table 4 Supervised learning with balanced data result analysis

Classifier	Accuracy	Log loss	Cross-validation	Recall	F1	Error rate
Balanced bagging classifier	77.115	0.52	76.889	51.340	36.3	33.70
Balanced random forest classifier	66.602	0.61	66.062	69.121	34.4	33.70

- *Random Forest Classifier.* It is a meta-estimator that uses k numbers of trees (decision trees) classifiers on different various small samples (or subsamples) of the collected dataset and use averages, in order to remove errors generated by oversampling it runs various decision trees parallelly and with the help of voting predicts the outcome.

After applying the above classifiers, we found the following visualizations/ results as shown in Table 3.

Now, we are applying balanced algorithms such as:

- Balanced Bagging Classifier
- Balanced Random Forest Classifier

After applying the above, we have found the following outcomes as shown in Table 4.

4 Result

Choice based on the visualizations is shown in Fig. 6, balanced bagging classifier from imblearn is the algorithm of choice for this data. While some of the scores may have been close, balanced bagging classifier had higher scores in accuracy, cross-validation, and specificity. The algorithm also had the lower error rate and false positive rates of the group. Balanced bagging classifier with ‘LightGBM’ performed best of among all the classifiers; however, we were not comfortable with how close its predictions were for serious accidents in the confusion matrix. Due to this, we



Fig. 6 Used classifiers performance graphs

decided to combine balanced bagging classifier with the second-highest performing algorithm, ‘LightGBM’ to see what results we would get (see Fig. 7).

After analyzing the above predictions, we have got below blackspots areas (see Fig. 8).

5 Conclusion and Future Scope

No matter the situation, most accidents were involved areas that were uncontrolled. One of the main areas where this happened was in the junction Detail T or staggered junction. Other areas of concern include accident locations that included mid-junction on roundabouts or main roads. No matter the location, details, or location of impact the common denominator seems to be a lack of signage or control in junction areas. The above analysis and resultant as shown in Table 5 show us that the following classifiers may perform best so the machine learning classifiers, and we applied are mentioned below along with their accuracy score as follows:

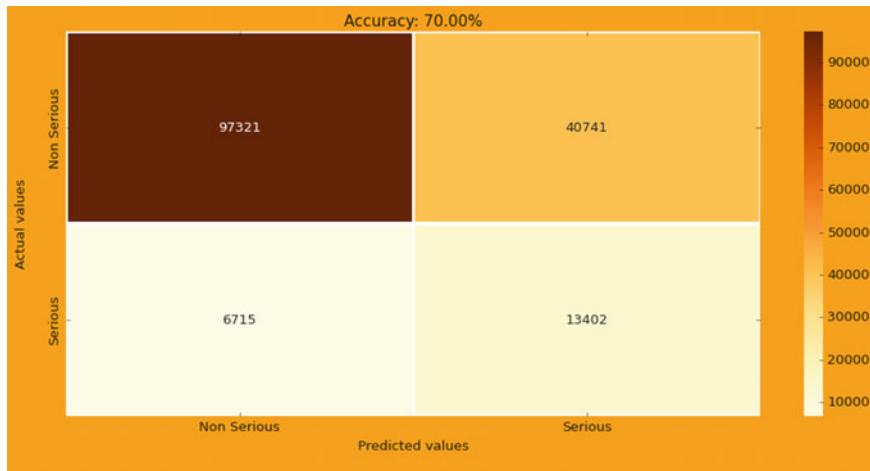


Fig. 7 Confusion matrix of balanced bagging classifier

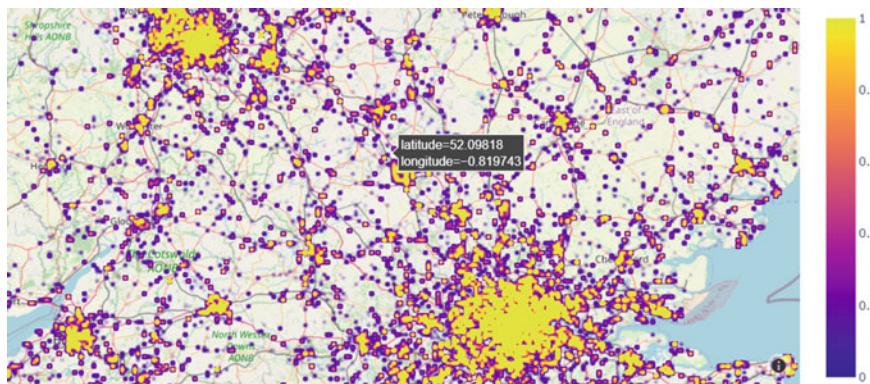


Fig. 8 Black spot areas

Table 5 Result analysis of all used classifiers

Classifier	Accuracy	Log loss	Cross-validation	Recall	Error rate
Bagging classifier	67.018	0.632	67.779	66.257	32.954
Ada boost classifier	68.771	0.677	63.023	54.785	31.229
Random forest classifier	66.291	0.622	68.030	68.633	33.709
Balance bagging classifier	77.115	0.522	76.889	51.340	33.709
Balanced random forest classifier	66.602	0.616	66.062	69.121	33.709
Balanced bagging classifier W/LGBM	70.00	0.582	69.570	68.560	30.00

- Random Forest Classifier (sklearn) with Accuracy of 66.291%.
- Balanced Bagging Classifier (imblearn) with Accuracy of 77.115%.
- Balanced Random Forest Classifier (imblearn) with Accuracy of 66.602%.

Now after applying the above classifiers, we found that balanced bagging is performing best with accuracy of near about 77.115%, so we used it and predict our result further and plot blackspot areas accordingly.

Future scope includes the following points as—with the help of visualizations and machine learning algorithms, areas of concern will be highlighted, and the seriousness of accidents will be predicted as accurately as possible in the black spot areas, and hence with the help of this project, we can reduce the fatality rates. With the help of this project, we can figure out the different flaws in our roads also by doing so, the government can improve the infra, prevent accidents in the upcoming future and can also spread awareness among their citizens about dos and donts and It can also be beneficial to insurance companies looking to change their rates in different areas, and it will also educate peoples on how to drive safely.

Acknowledgements We are very thankful to www.Kaggle.com because without it we do not get such a valuable dataset of the UK, i.e., vehicle and road separately that helped us analyze them separately.

References

1. Our World in Data-Death rate from road accidents in 2017, <https://ourworldindata.org/grapher/death-rates-road-incidents>. last accessed 24 April 2021
2. Jacobs, Sayer (1983) Road accidents in developing countries. *Accid Anal Prev* 15(5):337–353
3. McLeod R (2003) Influence of extrinsic and intrinsic risk factors on predicting probability of sustaining an injury. 35(1):71–80
4. Ray WA, Fought RL, Decker MD (1992) Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. *Am J Epidemiol* 136(7):873–883
5. Stelmach GE, Nahom A (1992) Cognitive-motor abilities of the elderly driver. *Hum Factors: J Hum Factors Ergon Soc* 34(1):53–65
6. Chipman ML, Smiley AM, Lee-Gosselin M (1993) The role of exposure on comparisons of accident risk with environment. 25(2):207–211
7. Yagil D (1998) Gender and age-related differences in attitude toward traffic laws and traffic violations transportation research part F. *Traffic Psychol Behav* 1(2):123–125
8. Mayhew (2006) Effects of age group along with experience of young driver accidents. *Traffic Injury Prev* 10(3):209–19
9. Lonczak HS (2006) Predicting risky and angry driving as a function of gender. 39(3):536–45
10. Hunt L, Morris JC, Edwards D, Wilson BS (1993) Driving performance in persons with mild senile dementia of the alzheimer type. *J Am Geriatr Soc* 41(7):747–753
11. Turner C, McClure R (2003) Age and gender differences in risk-taking behavior as an explanation for high incidence of motor vehicle crashes as a driver in young males. *Inj Control Saf Promot* 10(3):123–130
12. Roni Factor (2007) Inter-group differences in road-traffic crash involvement. 40(6):2000–2007
13. Joannes El, Ismini C, Charalambos D, Sofia G, Helene G, Myrsini C, Chliaoutaki (2002) Greek christian orthodox ecclesiastical lifestyle: could it become a pattern of health-related behavior? *Prevent Med* 34(4):428–435. <https://doi.org/10.1006/pmed.2001.1001>

14. Carlson WL, Klein D (1970) Familial versus institutional socialization of the young traffic offender. *J Safety Res* 2(1):13–25
15. Taubman O, Mario, B-A, Mikulincer Amit I (2004) A multi-factorial framework for understanding reckless driving—appraisal indicators and perceived environmental determinants. *Transp Res Part F: Traffic Psychol Behav* 7(6):333–349. <https://doi.org/10.1016/j.trf.2004.10.001>
16. David F, Preusser Allan F, Williams Adrian K, Lund (1985) Parental role in teenage driving. *J Youth Adolesc* 14(2):73–84. <https://doi.org/10.1007/BF02098648>
17. Lee J-T, Fazio J (2005) Influential factors in freeway crash response and clearance times by emergency management services in peak periods. *Traffic Inj Prev* 6(4):331–339
18. Abdelwahab, Abdel-Aty (2004) Investigating the effects of LTV percentages on head-on collisions. *Fatal Traffic Crashes* 130:429–437
19. Abdelwahab A-A (2004) Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure
20. Harnen S, Umar RSR, Wong SV, Hashim WIW (2003) Predictive model for motorcycle accidents at three-legged priority junctions. *Traffic Inj Prev* 4(4):363–369
21. Kececi EF, Tao G (2006) Adaptive vehicle skid control. *Mechatronics* 16(5):291–301
22. Chang L-Y, Wang H-W (2006) Analysis of traffic injury severity: an application of non-parametric classification tree techniques. *Accid Anal Prev* 38(5):1019–1027
23. Lvneh M, Hakket AS (1992) Some factors affecting the increase of road accidents in developing countries, with particular reference to Israel. *Accid Anal Prev* 4(2):117–33
24. Will MT, Whiteing (1995) Reducing commercial vehicle accidents through accident databases. *Logistics Inf Manage* 8(3) 22–29. <https://doi.org/10.1108/09576059510091643>
25. Sayed T, Abdelwahab W (Jan 1998) Comparison of fuzzy and neural classifiers for road accidents analysis. *J Comput Civil Eng* 12(1)
26. Aderamo AJ (2002) The structure of intra-urban road network development in Ilorin, Nigeria
27. Singh (2004) Road accident analysis: a case study of Patna City
28. Fang FC, Elefteriadou L, Pecheux KK, Pietrucha MT (2004) Using fuzzy clustering of user perception to define levels of service at signalized intersections. *J Transp Eng* 129:657–663
29. Most Dangerous Roads in America Infographic, <https://www.fleetowner.com/safety/article/21701577/25-most-deadly-highways-in-the-us>. last accessed 24 April 2021
30. Pradip SMJ, Bir L, Dogra STD (1994) Road traffic fatalities in Delhi: causes injury patterns and incidence of preventable deaths. *Accid Anal Prev* 26(3):377–384. [https://doi.org/10.1016/0001-4575\(94\)90011-6](https://doi.org/10.1016/0001-4575(94)90011-6)
31. Mandloi D, Gupta R (2003) Evaluation of accident black spots on roads using geographical information systems (GIS). Map India Conference. India
32. Overview of road accident in India, <https://prsindia.org/policy/vital-stats/overview-road-accidents-india>. last accessed 24 April 2021
33. Huang Z (1997) A clustering algorithm to cluster very large datasets. *Data Mining* 1–6
34. Huang Z (1998) Extensions to the K-means classification algo for clustering large datasets with categorical value. *2(3):283–304*

Detecting Hate Speech and Offensive Language Using Transformer Techniques



Mahin Bindra, Bhavya Sharma, and Nipun Bansal

Abstract Social media has become part of our everyday lives, and it has its pros and cons. Hate speech online is a new-born problem in our modern society that is growing at a steady pace, using the weaknesses of cohesive regimes that spill over several social media platforms. Due to the inconsistent and large size of social media, the detection of hate content has become a major concern to avoid conflicts and prevent unwanted activities. Hence, it is crucial to devise an automated system to detect the same. There have been many studies in the field of Artificial Intelligence to detect hate speech. Nonetheless, the efficiency of the models has not been up to the mark. In this study, we compare two different state-of-the-art models. The first is BERT-CNN model based on transformer, and the second is an LSTM model based on attention. Both these models work on a publicly available dataset containing messages categorized into different hate emotions. This study shows that the BERT-CNN model exhibits the highest accuracy, while outperforming the other.

Keywords Hate speech · Convolutional neural network · Transformers · Attention · Natural language processing

1 Introduction

Internet is easily accessible and with a surge in its utilization, people have started to misuse the internet. Cyberbullying has now become a common trend among these internet users as the bully tries to harass a person over social media, sending intimidating messages that possess a threatening nature of some sort [1]. People have started to do this under the impression that their identity is hidden and not accessible to the public. Hate speech is a form of insulting public speech directed at specific individuals or groups based on their gender, race, caste, sex, origin, etc. Although hate

M. Bindra (✉) · B. Sharma · N. Bansal

Department of Computer Science, Delhi Technological University, New Delhi, India

N. Bansal

e-mail: nipunbansal@dtu.ac.in

speech in many countries is considered a criminal offence, its usage is not declining. Instead, it is increasing at a very steady rate. There have been various studies that propose a link between hate speech and crimes related to it on social media platforms. With the increased use of hate speech over the internet, there has been an advent of hate-related crimes over these social platforms [2, 3]. Keeping in mind the drastic and traumatic effects hate speech can have on a person's brain, many big companies and individuals have come up with ways to detect and prevent this kind of behavior. Companies like Facebook, Google, and Twitter are trying their best to remove such content from their platform. Currently, these companies have employed human invigilators who manually red flag these messages and delete them [4]. Due to the extensive labor required to perform such a task, it is not always feasible. Hence, continuous advancements are being made in the field of machine learning. Machine learning algorithms, as we know, are categorized into two major divisions, the first one being the classical machine learning, and the second one is known as deep learning. Many studies have used either method to create models to detect hate speech [5]. The classical machine learning algorithms are dependent on feature engineering, which is a very tedious and extensive process. This process not only slows down text prediction but also decreases the impact on text representations. This downward impact can be semantic and syntactic [6]. Deep learning algorithms, on the other hand, are more efficient and perform a thorough training of the model. One of the most common is the Recurrent Neural Network (RNN) which is able to store text information sequentially over time, thereby executing contextual information with more accuracy, while classifying the same [7]. Developers have come up with techniques using Recurrent Neural Networks (RNN) and Long-Short Term Memory (LSTM), along with an attention mechanism to detect hate content. Recurrent neural networks forbid parallel processing and have a negative impact on the processing time, thus, developers are now focused on improving models that are based on attention mechanisms. This research by developers, has led to the introduction of transformers with the capability of self-attention [8]. Transformer algorithms are oriented such that they can capture long-term dependencies, with the help of positional embedding to remember the sequence of words in a particular order [8]. Hence, with the correct use of transformers, it is possible to devise models that can detect hate speech more efficiently than the traditional models.

2 Related Work

Models that use supervised machine learning techniques as the principal approach for detecting hate content are highly efficient [9]. The human language being open-ended, possesses many drawbacks, which causes difficulty during the detection or classification of text. Thus, existing models that detect hate messages fail to apprehend long-term dependencies, which further contributes to the loss of textual content, and this is crucial in interpretations of semantics [10]. To meet the challenge of long-term dependencies, RNN algorithms like the LSTM were devised [11]. Another

challenge these models face is the difficulty of differentiation of hate speech from other languages that may be offensive in nature. The lexical detection methods classify messages containing all sorts of offensive content into hate speech [12]. LSTM has a structure similar to the Vanilla RNN chain. The only difference is that it continuously includes more openings to dominate the amount of content allowed in all such situations. LSTM is very supportive in eliminating the vanishing gradient issue [13]. In conclusion, LSTM conserves long-term dependencies more effectively than Vanilla RNN, thus ensuring that the model apprehends more data. Without such advantages, LSTM is not able to do the above. Specifically, the LSTM performance decreases as the sequential length increases over a certain limit [14]. Additional studies aimed toward solving long-term dependencies have found a way to devise mechanisms based on attention [15]. Attention techniques work only assuming, that all the words in each sentence are pertinent [16]. This process allows context to be taken into account when separating crucial data such as hateful content. Algorithms based on attention have a history of successful progress in text detection and analysis. However, these models cannot operate on word sequences in parallel, as they use RNNs [17]. With the recent adaptation of the attention approach, there has been a gradual shift from RNNs to transformers [18]. Transformer instantly became a high-performance development structure of NLP, outperforming RNNs in NLG and NLU. There exist various types of transformer methods that are continuously succeeding in the field of ML and NLP. Bi-directional Encoder Representations from Transformers (BERT) [18] surpass previously achieved standards for general NLP functions [19]. BERT utilizes large, non-labeled information to create prototypes with adjustable variables as required for small controlled data for proper enhancement. The accomplishment of BERT has further motivated developers to devise algorithms dependent and inspired by it. Some of which are RoBERTa [20], DistilBERT [21] & XLNET [22]. RoBERTa, an advancement of BERT, which focuses on large databases to enhance the execution, DistilBERT on the other hand, adapts to a systematic type. XLNET is another standard default whose main role is to recreate the initial data taking into account different corrupt installations.

3 Materials and Management

3.1 Experimental Dataset

In this study, a multiclass hate speech dataset has been used for training the model. This dataset is developed by the Conversation AI team, established by Jigsaw and Google [23]. The dataset used contains approximately 1.5L text messages that have been categorized into “toxic”, “severe_toxic”, “obscene”, “threat”, “insult”, “identity_hate.”

3.2 Experimental Setup

All methods used the Python programming language. For the attention-based LSTM hate speech model, we used the Keras library. We, then implemented, the Bert-CNN model using the BERT-uncased transformer class, built on the PyTorch framework. The test was performed on a Windows operating system with an Intel Core i5-8250U CPU with GeForce GTX 1050 Ti, 16 GB RAM, and 1 TeraByte hard disk drive configuration.

3.3 Pre-Processing of Data

The messages exchanged on social media platforms are unpretentious, and the dataset we used was unstructured and contained a large amount of noise which could alter the model accuracy. Thus, pre-processing of data had to be done, which is a process used to enhance the performance while reducing processing time simultaneously. To pre-process the data, we followed the given steps:

- Text Cleaning/Normalization: To normalize data, we worked upon the construction words and expanded them into simpler words to avoid confusion for the model.
- Removal of Stop words: Words like “is”, “am”, “are”, Digits and Scape words are removed from the data as these words do not add or change the meaning of the sentence.
- Stemming: This process is used to reduce word repetitions by categorizing similar words and forming root/stem word indicating all words of the same class. We used the Snowball Stemmer for this process.
- Lemmatization: It is similar to stemming but a smarter way to categorize the words taking into account the tense of the word. We used the WordNetLemmatizer to carry out the process.
- Tokenization: Indexing of words in order of their occurrence, where the most occurred word is ranked the highest and least occurred is ranked the lowest.

3.4 Proposed Method

Attention-based LSTM Model

The attention mechanism attempts to implement actions on selectively relevant things by overlooking others in deep neural networks. The whole idea behind it is that each time the model gives an output, it only uses a part of the neuron where the most relevant information is present instead of using the whole input. In layman terms, it only pays attention to some of the words [24]. The Rectified Linear Activation function (ReLU) is a piecewise linear function that directly impacts the input. It is

the default activation function for most neural networks, as it is easier to use and yields better performance [25]. However, we have not used ReLU because it is not a self-normalizing activation function as it cannot give a negative value. So we have used SeLU. The equation for SeLU is shown below:

$$f(a, x) = \begin{cases} a(e^x - 1), & x < 0 \\ x, & x \geq 0 \end{cases} \quad (1)$$

Although the above formula appears to be like ReLU for values greater than zero, an extra parameter λ is convoluted here. This parameter represents the S (caled) in SeLU. If it is greater than one, the gradient is also greater than one, and the activation function can increase the variance. Therefore, SeLUs possess an outstanding quality of self-normalization, and they diminish the problem of vanishing gradients as well. Reasons why one should prefer SeLUs over ReLUs:

- Identical to ReLUs, SeLUs validate deep neural networks since they can handle the issue of vanishing gradients.
- As opposed to ReLUs, SeLUs do not die.
- SeLUs adapt faster and better than other activation functions, even after combining with batch normalization [26].

In this sequential model, our first layer was from the embedding matrix of maximum size length(150)*features. We used partial dropout to solve the problem of vanishing gradient. In the second layer, we used bi-directional LSTM with 0.5 dropout. We used attention as a hidden layer with other layers using the activation function SeLU. The sigmoid function was used in the output layer, and the model was compiled using binary cross entropy. Figure 1 depicts the architecture of this model.

BERT-CNN Model

The transformer model method uses the attention mechanism to improve the speed of training of these models. The advantage that the transformer has is that it can lend itself to parallelization. The transformer architecture, based solely on the attention mechanism, was first proposed by A. Vaswani et al. [8]. The pipeline of a basic transformer-based model includes processing of data, method application, and prediction analysis. Transformers are a better option as they inherit the feature of self-attention, a method used to bake the understanding of words relevant to the one being processed currently [27]. In this study, we use BERT base-uncased model, which is pre-trained on the English Language, the objective used, while pre-training is Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). The drawbacks of this model are that it is not able to differentiate between lowercase and uppercase. In addition to that, this model can have biased predictions depending on the type of data used [18]. The BERT model provides an edge over any other model in terms of pre-training, as it adapts itself to having a deep and profound understanding of the working of the language. In this feed-forward network, we

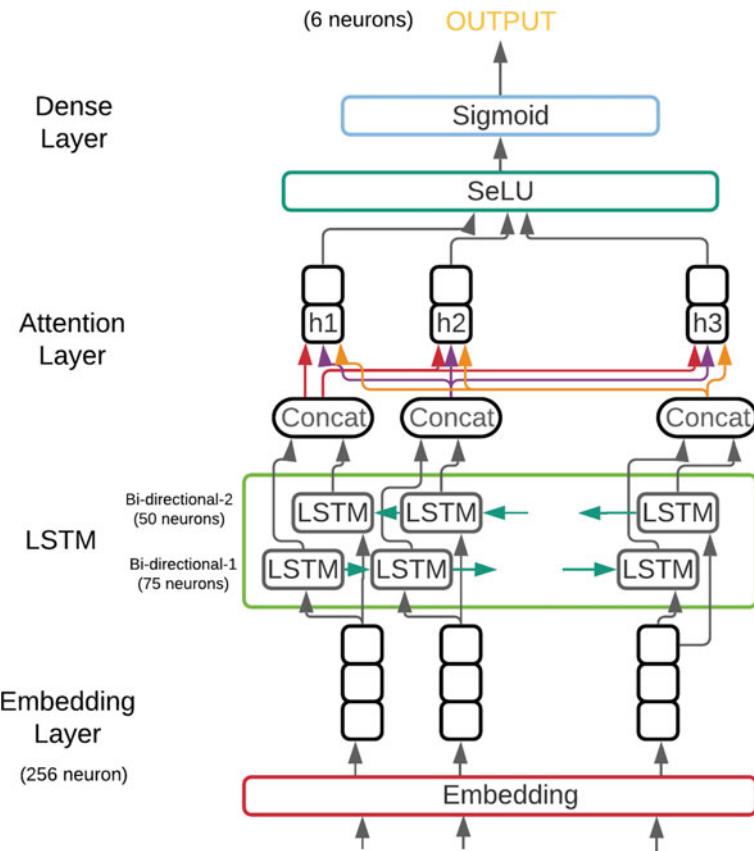


Fig. 1 Architecture for attention-based LSTM model

used BERT pre-trained embedding to create a vector representation of the words to use in our embedding matrix, just like we used GloVe in the attention-based LSTM model for the same purpose. We, then feed the embedding matrix into the text-based CNN mode. Here, we keep the kernel size to 2, 3, and 4. The size of the kernel will decide to extract the two highest weights of the sentences. In the first layer of the convolution network, we used ReLU as our activation function. Later, Max-pooling was used to select the maximum element from the region of the feature map. And this was followed by dropout. We used a classifier from Pytorch to create logits that confer the same output as the classifications of hate speech mentioned earlier. Figure 2 depicts the architecture of this model.

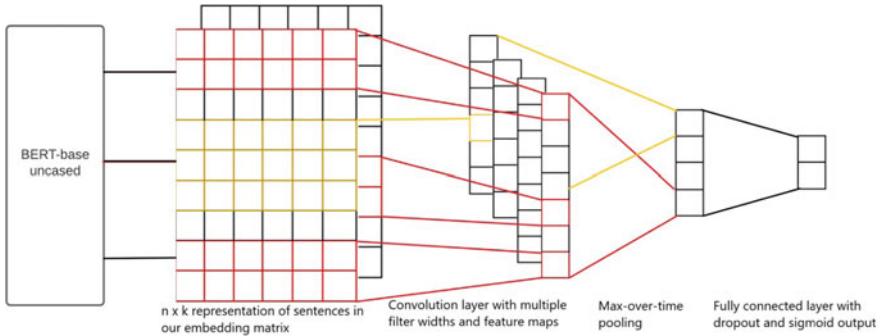


Fig. 2 Architecture for BERT-CNN model

3.5 Model Parameters

The tuning of hyper-parameters is crucial, while customization of the pre-trained model for it to perform specific tasks. We use BERT pre-trained embeddings from its dictionary, mapping them to our corpus to create an embedding matrix that feeds into our model. As depicted in Table 1, we fine-tuned the model to detect hate content by altering the sequence length, validation data split, batch size, and epoch count. The value of sequence length was 256 characters, and the ideal epoch count in this study was one. We used 10% of the data for validation purposes. The evaluation batch size value was eight, and the learning rate was 3e-5. We used the Adam Optimizer in the output layer. We can understand Adam as a combination of RMS prop and stochastic gradient descent.

Table 2 shows the hyper-parameters used to optimize the attention-based LSTM model for hate speech detection. In this model, we used a maximum sequence length of 150, and the batch size was 64 as the approach of this model is different than the BERT model. A bigger batch size yielded better results. The learning rate used in this model was 1e-5. In our model, binary cross-entropy function was used. The objective of this function is to return larger values for poor predictions and smaller values for fine predictions. In this function, if the probability associated with the True class is one, then the loss is minimum, or else the loss is maximum or high.

Table 1 Hyper-parameters of BERT-CNN model

Sequence length	Optimizer	Batch size	Epochs	Learning rate	Loss function
256	Adam	8	1	3e-5	BCE

Table 2 Hyper-parameters of attention-based LSTM model

Sequence length	Optimizer	Batch size	Epochs	Learning rate	Loss function
150	Adam	64	20	1e-5	Binary Cross Entropy

4 Results

We compared the results of the two proposed models, BERT and attention-based LSTM. We examined the following performance metrics in the two models: Accuracy, F-measure, Precision, Recall, and ROC (receiver operator characteristic) Curve. The dataset used was split in 90:10 ratio for training and testing, respectively.

4.1 Classification Report

The metrics mentioned below were used to correctly examine the model performance. This report is used to display the precision, sensitivity, f-measures, and accuracy values of a model. These metrics are illustrated on the basis of four parameters, True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP is the number of hate messages that our model predicts correctly, FP classifies the messages as hate despite them being neutral in nature. TN is the value of correctly predicted non-hate messages, and FN is incorrect prediction of hateful messages.

Accuracy: This metric is crucial as it gives information about the TP and TN . Accuracy is calculated as the ratio of summation of TP & TN to the total values of the confusion matrix. Accuracy helps to differentiate models from one another to pick the better one for classification. The efficiency of the model does not depend on accuracy solely, other parameters need to be considered.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

The accuracy observed from Fig. 8 for the attention-based LSTM model came out be 89%, whereas the accuracy observed from Fig. 7 for the BERT-CNN model was 95%. The BERT-CNN model outperformed the attention-based LSTM model. This is because of the sole reason that transformer-based algorithms have the capability to apprehend long-term dependencies better than attention-based LSTM models. Figure 3 depicts the plot of accuracy scores for attention-based LSTM model and BERT-CNN Model.

Fig. 3 Plot for accuracy scores of attention-based LSTM model and BERT-CNN model

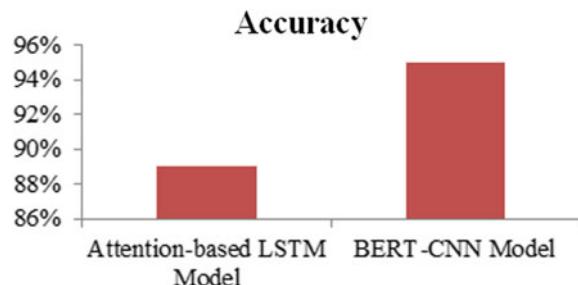


Fig. 4 Plot for precision scores of attention-based LSTM model and BERT-CNN model

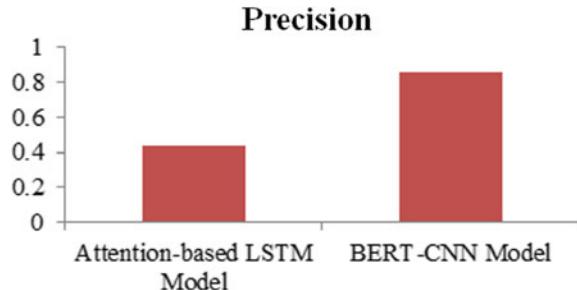
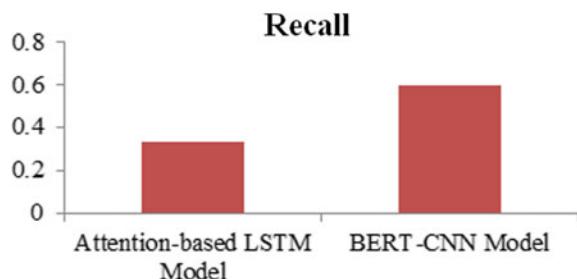


Fig. 5 Plot for Recall scores of attention-based LSTM model and BERT-CNN model



Precision: It is the ratio between correctly predicted hate messages and the total number of messages predicted as hate. A high precision value denotes low error rate.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

It is clear from Figs. 7 and 8 that the attention-based LSTM models show a precision of 0.44, whereas the BERT-CNN model shows a precision of 0.86. Figure 4, depicts the plot of Precision scores for attention-based LSTM model and BERT-CNN Model.

Recall (Sensitivity): The ratio of accurate positive predictions to the total messages that are hateful in nature is called Recall. The value of recall lies between 1.00 and 0.00, where 1.00 is the best and 0.00 is the worst value.

$$\text{Recall or Sensitivity} = \frac{TP}{TP + FN}. \quad (4)$$

We observe from Figs. 7 and 8, the recall value for the attention-based LSTM model was 0.33, whereas the recall value for the BERT-CNN model came out to be 0.60, which is preferable as it is above 0.5. Figure 5, depicts the plot of recall scores for attention-based LSTM model and BERT-CNN Model.

F-measure: F-measure is the weighted average of recall and precision. This parameter sometimes is considered as a better option to evaluate the model's performance as compared to accuracy, but this is so only in the case of unequal category

distribution.

$$F\text{-measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

Figures 7 and 8 give us an idea about the F-measure values of both the models. The F-measure of attention-based LSTM model is 0.38, whereas the F-measure of BERT-CNN model is 0.70. Figure 6 depicts the plot of F1-scores for attention-based LSTM model and BERT-CNN Model.

The attention-based LSTM model used a GloVe file, as there were certain word embeddings in the corpus not mapped correctly. As a result, the vector representations got hampered, which lead to low values of precision, recall, and F-measure values. The dataset used was imbalanced, and as a result, the performance metrics yielded low values for the attention-based LSTM model. Fixing the dataset could provide better

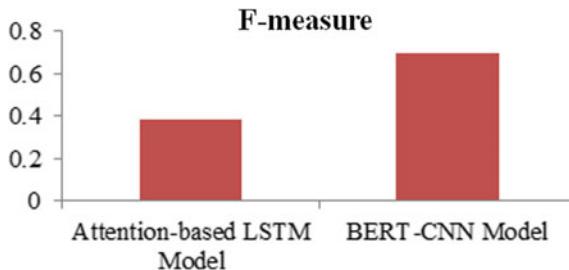


Fig. 6 Plot for F1-scores of attention-based LSTM model and BERT-CNN model

	precision	recall	f1-score	support
0.0	0.96	0.99	0.97	14377
1.0	0.86	0.60	0.70	1581
accuracy				0.95 15958

Fig. 7 Snapshot of precision, recall, f1-score, and support values for BERT-CNN model

	precision	recall	f1-score	support
0	0.93	0.95	0.94	14411
1	0.44	0.33	0.38	1547
accuracy				0.89 15958

Fig. 8 Snapshot of precision, recall, f1-score, and support values for attention-based LSTM model

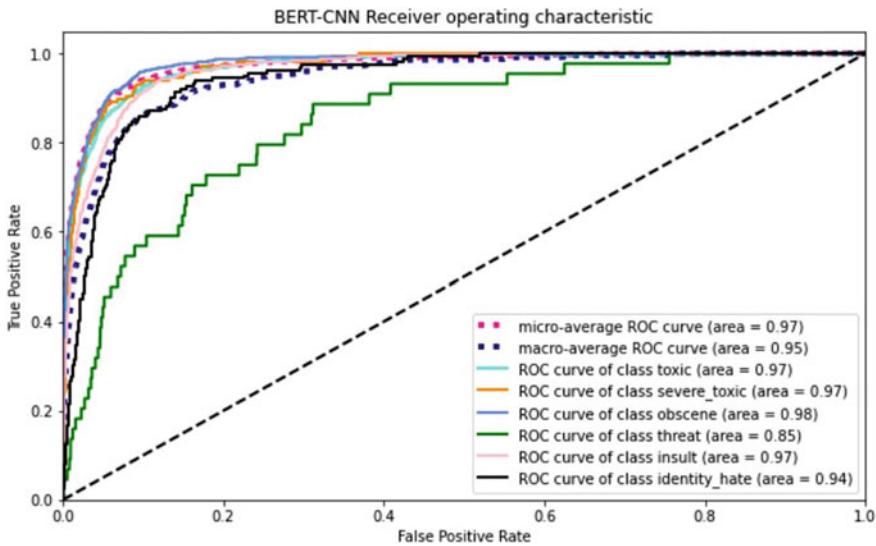


Fig. 9 ROC curve for BERT-CNN model

performance metric values for the attention-based LSTM model, but the BERT-CNN model yielded high values.

4.2 ROC Curve

The receiver operating characteristic plots the values between True Positive Rate (TPR) and False Positive Rate (FPR). This graph is used to represent the performance of the model at different classification thresholds. Figures 9 and 10 depict the ROC curve for the BERT-CNN model and the attention-based LSTM model, respectively. This was another metric used to compare how the two models performed in terms of classification. The AUROC value for the BERT-CNN model is 94.67%, whereas the AUROC value for the attention-based LSTM model is 84.16%. Clearly, the BERT-CNN model outperforms the attention-based LSTM model in terms of classification.

5 Conclusion and Future Work

Given the traumatizing effect hate speech can have on an individual, it is crucial to develop models that can differentiate whether a text contains any hate emotion/offensive language or not. Even after unwavering efforts from major organizations, the detection of hate speech remains a problem in today's time. In this

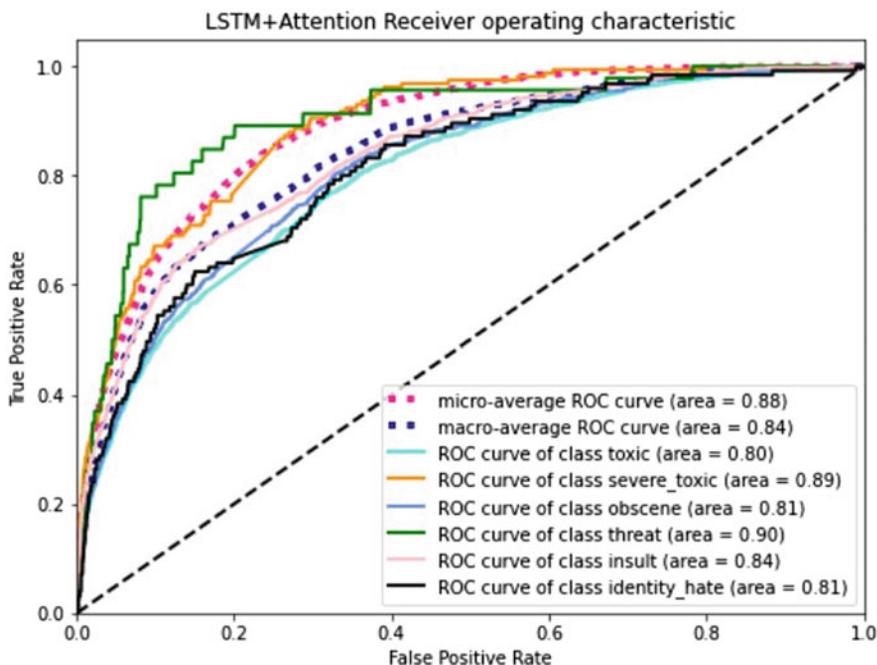


Fig. 10 ROC curve for attention-based LSTM model

study, we have tried to examine the working performance of the two state-of-the-art models in detecting hate content. The outcomes showed that the BERT-CNN model outperformed the attention-based LSTM model in hate speech detection on a multi-class dataset. The accuracy for the BERT-CNN model and attention-based LSTM was 95% and 89%, respectively. In future, we would like to work with a multilingual dataset for detecting hate speech. Transformers can only apply to limited training data, but hate over the internet can be expressed in many languages as people in many parts of the world speak multiple languages. The second part of future work can include detecting hate/offensive content using different format of data. In our study, we used textual data only. However, information on the internet is made available in many forms, such as pictures, video content, memes, etc., which may go undetected. Therefore, training the model using datasets containing hate speech in different formats can improve detection rates of hate content over the internet.

References

1. Atapattu T, Herath M, Zhang G, Falkner K (Dec 2020) Automated detection of cyberbullying against women and immigrants and cross-domain adaptability
2. Hosseini mardi H, Mattson SA, Rafiq RI, Han R, Lv Q, Mishra S (Mar 2015) Detection of

- cyberbullying incidents on the instagram social network
- 3. Zhang Z, Robinson D, Tepper J (2018) Detecting hate speech on twitter using a convolution-GRU based deep neural network. pp 745–760
 - 4. Waseem Z, Hovy D (2016) Hateful symbols or hateful people? Predictive features for hate speech detection on twitter. In: Proceedings of the NAACL student research workshop. pp 88–93
 - 5. Mutanga RT, Naicker N, Olugbara O (2020) Hate speech detection in twitter using transformer methods. *Int J Adv Comput Sci Appl* 11(9)
 - 6. Young T, Hazarika D, Poria S, Cambria E (Aug 2017) Recent trends in deep learning based natural language processing
 - 7. Sohangir S, Wang D, Pomeranets A, Khoshgoftaar TM (2018) Big data: deep learning for financial sentiment analysis. *J Big Data* 5(1):3
 - 8. Vaswani A, et al. (Jun 2017) Attention is all you need
 - 9. Schmidt A, Wiegand M (2017) A survey on hate speech detection using natural language processing. In: Proceedings of the fifth international workshop on natural language processing for social media. pp 1–10
 - 10. Wang Y, Huang M, Zhu X, Zhao L (2016) Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing. pp 606–615
 - 11. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
 - 12. Davidson T, Warmsley D, Macy M, Weber I (Mar 2017) Automated hate speech detection and the problem of offensive language
 - 13. Pascanu R, Mikolov T, Bengio Y (Nov 2012) On the difficulty of training recurrent neural networks
 - 14. Kowsari K, Meimandi KJ, Heidarysafa M, Mendu S, Barnes LE, Brown DE (Apr 2019) Text classification algorithms: a survey
 - 15. Bahdanau D, Cho K, Bengio Y (Sep 2014) Neural machine translation by jointly learning to align and translate
 - 16. Kim Y, Denton C, Hoang L, Rush AM (Feb 2017) Structured attention networks
 - 17. Wang X, Jiang W, Luo Z (2016) Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In: COLING
 - 18. Devlin J, Chang M-W, Lee K, Toutanova K (Oct 2018) BERT: pre-training of deep bidirectional transformers for language understanding
 - 19. Rajapakse T (2020) To distil or not to distil: BERT, RoBERTa, and XLNet. [Online]. Available: <https://towardsdatascience.com/to-distil-or-not-to-distil-bert-roberta-and-xlnet-c777ad92f8>. [Accessed: 10 Apr 2021]
 - 20. Liu Y, et al. (Jul 2019) RoBERTa: a robustly optimized BERT pre-training approach
 - 21. Sanh V, Debut L, Chaumond J, Wolf T (Oct 2019) DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter
 - 22. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV (2019) XLNet: generalized autoregressive pre-training for language understanding. In: Advances in neural information processing systems, vol 32
 - 23. Toxic Comment Classification Challenge Dataset (2018) [Online]. Available: <https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data>. [Accessed: 10 Apr 2021]
 - 24. Loginova K (2018) Attention in NLP. [Online]. Available: <https://medium.com/@edloginova/attention-in-nlp-734c6fa9d983>. [Accessed: 10 Apr 2021]
 - 25. Brownlee J (2019) A gentle introduction to the rectified linear unit (ReLU). [Online] Available: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/> [Accessed: 10 Apr 2021]
 - 26. Böhm T (2018) A first Introduction to SELUs and why you should start using them as your activation functions. [Online] Available: <https://towardsdatascience.com/gentle-introduction-to-selus-b19943068cd9>. [Accessed: 10 Apr 2021]
 - 27. Alamar J (2018) The illustrated transformer. [Online]. Available: <http://jalamar.github.io/illustrated-transformer/>. [Accessed 10 Apr 2021]

Enabling Multi-Factor Authentication and Verification in Searchable Encryption



Sai Deepika Panguluri, K. V. Lakshmy, and Chungath Srinivasan

Abstract Data management with third-party involvement is often considered a severe concern in the new normal. The cloud environment has become the current trend in the modern internet era, but anonymity, maintenance, and data secrecy have become more concerned. Trusting third-party that store confidential data on the cloud may cause entities irreparable damage. As a result, the data should be kept protected to sustain secrecy. Searchable Encryption (SE) solves the problem by browsing through encrypted data. Previous studies have mainly focused on encrypting, searching, and decrypting data using searchable encryption schemes. However, most of the schemes failed to include multi-factor authentication and verification, which adds an extra layer of security to data access with SE. We introduced a multi-factor authentication and verification scheme called the TOTPV-KASE scheme, i.e., time-based one-time password and verification in key-aggregate searchable encryption. We present a summary of various searchable encryption schemes and implement the time-based one-time password generated by the owner of the data and verify it using an Inter Planetary File System (IPFS).

Keywords Searchable Encryption (SE) · Data Owner (DO) · Data User (DU) · Cloud Service Provider (CSP) · Key-Aggregate Searchable Encryption (KASE) · Time-based One-Time Password (TOTP) · Inter Planetary File System (IPFS) · Time-based One-Time Password and Verification in Key-Aggregate Searchable Encryption (TOTPV-KASE)

S. D. Panguluri (✉) · K. V. Lakshmy · C. Srinivasan

TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

e-mail: cb.en.p2cys19010@cb.students.amrita.edu

K. V. Lakshmy

e-mail: kv_lakshmy@cb.amrita.edu

C. Srinivasan

e-mail: c_srinivasan@cb.amrita.edu

1 Introduction

Cloud computing refers to an architecture where a Cloud Service Provider (CSP) delivers services to customers based on their requirements. This lowers customer costs and is covered by a cloud policy known as pay-per-use. This aids in the installation, deployment, low maintenance, and scalability of the system. Customers prefer to use the cloud because of these benefits (Table 1).

1.1 *Cloud Computing*

Cloud computing has proven to be an ideal solution for providing customers with easy and on-demand access to their massive data. Cloud computing has become embedded in our everyday routine. It offers limitless tools and facilities, such as data storage and resource distribution on demand. Various sources produce a vast volume of confidential data every day from the medical sector, corporate industry, and government organizations. For the data owner, preserving this data has become a big issue. As a result, storing this massive data on individual devices in a privacy-preserving manner is challenging and also affecting its processing efficiency. So, cloud computing offers the best solution by providing massive storage and on-demand resource sharing by data access, while lowering infrastructure costs. Besides that storing our data in the cloud gives the cloud owner, an unreliable entity full control over our information. The question of privacy arises if the data stored is public. Encryption ensures data protection and confidentiality. Encryption is transforming plain text into unreadable data called ciphertext. The process of decryption will transform this ciphertext back into plain text. The data can only be decrypted by those who have been authorized. Since the data is encrypted and processing requests on the cloud takes a long time, the user's interaction with the data is rendered more difficult. As a result, searchable encryption techniques have been developed to address these issues. These techniques allow search operations on data stored on the cloud server in an encrypted format. Various searchable encryption systems have been proposed over the years. In SE schemes, the ciphertext is outsourced, and the operations are carried out in the cloud, with the results being forwarded to the users.

Table 1 Benefits of using cloud

File storage space	On-demand resource allocation
Accessibility at any time	Backups functionality
Cost associated with system infrastructure maintenance	Location independence
Computer independence	Pay-per-use
Easy installation, deployment	Scalability of the system

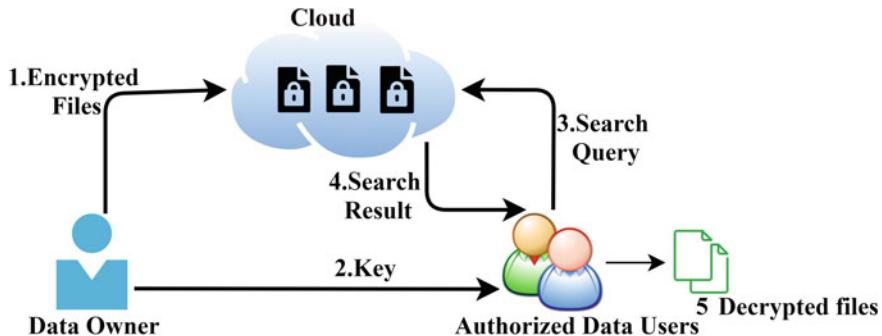


Fig. 1 Model of searchable encryption

Security Issues in cloud: Almost every company has migrated its data to the cloud to save space. Since public clouds are open, they face the greatest threat from the outside world among other cloud models. There are several barriers to cloud computing, the first of which will always be cloud security, which is a major concern for most companies. Data breaches, denial of service attacks, insider threats, vulnerable APIs, account hijacking, cyber-attacks, ransomware, phishing attacks, and regulatory enforcement concerns are just a few of them. There are many studies that tell about prevention of insider attacks [1]. According to the “Cloud Security Spotlight Report [2],” 90% of organizations have embraced cloud computing and are moderately concerned about public cloud security. Secure practices of cloud can be seen in [3].

1.2 *Searchable Encryption*

Wagner et al. [4] proposed Searchable Encryption in the year 2000. The concept is to perform a search operation on the ciphertext without decrypting it. For searching on encrypted data in an untrusted server, searchable encryption is used. Based on the number of keys involved in encryption, searchable encryption can be split into two groups. They are SSE (Searchable Symmetric Encryption) and SAE (Searchable Asymmetric Encryption). Public Key Encryption with Keyword Search (PEKS) is another name for searchable asymmetric encryption (SAE). SSE is accomplished through symmetric cryptography, while PEKS/SAE is accomplished through public key cryptography. Searchable encryption schemes mainly have three entities: 1. The data owner (DO) is the original owner of the data who wishes to outsource it to the cloud. They encrypts the file and upload in cloud. 2. Data user (DU) is an individual who queries the cloud server for the file. 3. The cloud service provider (CSP) stores and retrieves data in response to a query (Fig. 1).

Searchable Encryption Algorithm:

1. Data owner encrypts and uploads the encrypted file in the cloud.
2. DO shares the secret key to the valid users.
3. DU enters the search query and requests the cloud for data.
4. Cloud performs search operation and returns the result to the DU.
5. DU decrypts the files using the key.

Symmetric Searchable Encryption (SSE): SSE permits one party to transfer data to another while preserving the right to search the data using query. Except for the ciphertext, the query helps the server/other parties to learn nothing about the plaintext.

Public Key Encryption with Keyword Search (PEKS): The public key searchable encryption scheme, also known as PEKS, enables you to search encrypted data on an untrusted server without revealing any information. Only users with access to public keys are allowed to generate ciphertext in PEKS, and only users with access to private keys are allowed to create trapdoors for searching enciphers. The mechanism helps you to easily see if a keyword is present in the document without revealing any other information about it.

1.3 Organization of Paper

Section 2 discusses a review of the literature on searchable schemes that have been formulated over time. Section 3 discusses the existing KASE scheme. Section 4 explains the proposed scheme in which multi-factor authentication and verification come into picture. The paper comes to a close with Sect. 5.

2 Related Works

Song et al. [4] started the searchable encryption concept in which data owner is allowed to search over encrypted data on an email server. If data owner wish to search keywords, he/she submit the trapdoor to the server. The server searches over the encrypted data and retrieves the related documents. Boneh et al. [5] proposed the first-ever public key encryption with keyword search technique.

Symmetric Searchable Encryption (SSE): In symmetric searchable encryption, only the whole file or each word in a file is encrypted with a unique key. DES, AES are examples of symmetric encryption algorithms. Since the file's owner has access to the key, he can encrypt the word he's looking for and search the encrypted file with ease. The first searchable symmetric encryption scheme is proposed by Song et al. [4]. This scheme encrypts the file word by word. The user sends the keyword to the cloud with the same key to look for it. The disadvantage of this scheme is that it reveals the frequency of the word. Goh [6] attempted to address Song's flaw by creating a protected index table using pseudorandom functions and randomized bloom filters

for specific document identifiers. Bosch et al. [7] built on Goh et al. [6] definition by adding the concept of wild card searches. Bloom filters can introduce false positives, which is a disadvantage of this scheme. Each document is given its own index in Chang et al. [8] scheme. The scheme is more reliable than Goh's [6] because the number of words in a document is not revealed. This scheme has the disadvantage of being less effective and not supporting arbitrary updates of new terms. The Golle et al. [9] scheme helps you to check for several keywords with only one encrypted question. However, this strategy is impractical. Curtmola et al. [10] introduced the principle of symmetric searchable encryption (SSE), and later Kamara et al. [11] proposed a more advanced variant of SSE called dynamic SSE (DSSE), which allows for record addition and deletion in the index table. Curtmola's [10] and Kamara's schemes [11] are limited to keyword searches, with no support for range queries. Hacigumus et al. [12] suggested using the bucketization technique to provide SQL query support. The downside is that the scheme does not support scalability by using the bucketization technique. Hore et al. [13] suggested a method for bucketizing multidimensional data, while also allowing for range queries. The scalability constraint is the same as in the Hacigumus [12] system. Range queries are supported by Order Preserving Encryption (OPE)-based schemes. The disadvantage of OPE supporting schemes is that they expose the order relationship between ciphertext data. To secure databases, Popa et al. [14] suggested the CryptDB technique. As a database gateway, CryptDB employs a proxy server. For searching and computing on encrypted data, it employs range queries and homomorphic encryption methods such as the Paillier scheme. The disadvantage is that you have to trust the proxy server. Most of these research works in searchable encryption (e.g., [15–19]) were similar elements to Song et al. work. Many static and dynamic searchable symmetric encryption techniques are proposed [11, 20, 21].

Public key encryption using keyword search (PEKS): A public key with keyword search was proposed by Boneh et al. [5] in 2004. They were encrypting a message with keywords using the user's public key, and only the user with the proper private key could decrypt the message. It is based on a Computational Diffie-Hellman problem variant. In abstract, it uses two cyclic groups G_1, G_2 of prime order p , a bilinear map $e: G_1 \times G_1 \rightarrow G_2$. The map satisfies the following properties:

1. Computable: Given $a, b \in G_1$ there are polynomial-time algorithms to compute $e(a, b) \in G_2$.
2. Bilinear: For any integers $c, d \in [1, p]$ we have $e(g^c, g^d) = e(g, g)^{cd}$.
3. Non-degenerate: If g is a generator of G_1 then $e(g, g)$ is a generator of G_2 .
4. Two hash functions $H_1: \{0, 1\}^* \rightarrow G_1, H_2: G_2 \rightarrow \{0, 1\}^{log p}$ and the security parameter $\{G_1, G_2, e, H_1, H_2, a, b\}$.

PEKS is made up of five probabilistic polynomial-time algorithms. However, the computational cost of public key encryption will be heavy, and it will only be available to a small number of keyword searches. Park et al. [22] and Boneh et al. [5] generalized the PEKS scheme (Conjunctive search [5, 22], subset and collection [5] queries) on encrypted data to perform an efficient query search. Besides that, according to Boneh et al. [5], the scheme would need to create a stable channel

(such as SSL) to keep trapdoors out of the correspondence. It is expensive to set up a stable channel. Baek et al. [23] suggested a new PEKS called “Safe Channel Free—Public Key Encryption with Keyword Search (SCFPEKS)” to solve this issue. PEKS was created mostly for data sharing situations [5, 24, 25]. Deng et al. [26] designed a multiple-user searchable encryption scheme with keyword authorization (MSESKA) using asymmetric searchable encryption and six polynomial-time algorithms. To protect the search scheme, Yin et al. [27] developed a secure index methodology that combined Decisional Diffie–Hellman (DDH) and Bilinear Diffie–Hellman (BDH) assumptions with the bloom filter technique. Wang et al. [28] created a stable hybrid indexes scheme based on PEKS with the aim of reducing search difficulty. Ibraimi et al. [29] introduced third-party delegation using Public Key Encryption with Delegated Search (PKEDS), which enables a third-party to search a text. Two forms of public key encryption used for keyword search. They are deterministic SE [30] and plaintext-checkable Encryption [31]. On an ElGamal-based technique, Canard et al. [31] created an effective plaintext-checkable encryption scheme. Dynamic aggregate key sharing schemes were also proposed in which the key is dynamically generated based on request, and the old key is not used anymore [32].

3 KASE Scheme

The key-aggregate searchable encryption (KASE) for group data sharing through cloud storage [33], in which the owner of the file has to share a single key for a collection of documents and the user who needs the data has to submit a single request to the cloud.

3.1 *Functionality of KASE Scheme*

Algorithms in KASE Scheme: There are seven algorithms in KASE [33] scheme.

1. Set $(1^\beta, x)$: This will be used by CSP to set up system parameters where β is the security parameter which is an integer and G is the cyclic group.
 - (a) This generates a bilinear map group system $A = (y, G, G_1, e(., .))$, where y is the order of G and $2^\beta \leq y \leq 2^{\beta+1}$
 - (b) Maximum number of files that belong to Data owner is set to x . Picks a random generator $r \in G$ and random $\lambda \in Z_p$ and computes $r_i = r^{(\lambda,i)} \in G$ for $i = \{1, 2, \dots, x, x + 2, \dots, 2x\}$. Selects one way hash function $H: \{0, 1\}^* \rightarrow G$.
 - (c) Finally, the server produces the system parameters.
 $sysparams = (A, Pub, H)$ where $Pub = (g, g_1, g_2, \dots, g_x, g_{x+2}, \dots, g_{2x})$

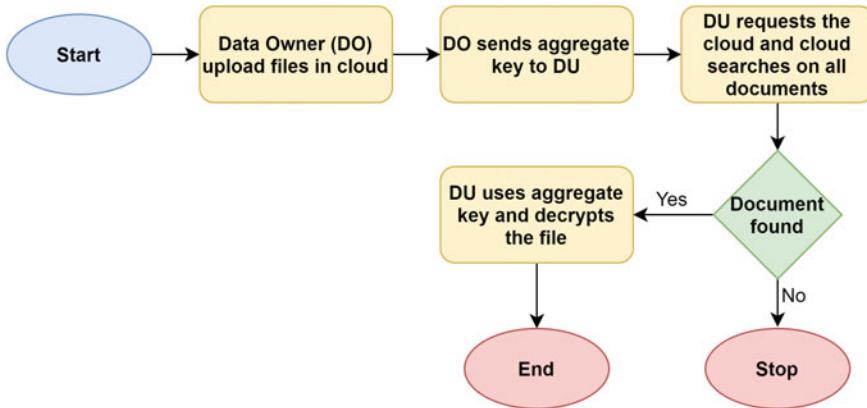


Fig. 2 Functionality of existing KASE scheme

2. Key Generation: This Key Generation algorithm is used by the DO to produce $(pubk, mk)$. Picks a random $\eta \in Z_p$.
 - (a) $pubk = \rho = g^\eta, mk = \eta$
3. Encrypt $(pubk, i)$: When uploading i^{th} document, the DO use this encrypt algorithm for encrypting the file and also to generate the ciphertexts of keywords.
 - (a) To produce the keyword ciphertexts, this takes as input file index $i \in \{1, 2, \dots, x\}$ and selects a random $a \in Z_p$ as document's searchable encryption key k_i .
 - (b) Generates $c_1 = g^a$ and $c_2 = (\rho * g_i)^a$. For keyword d , ciphertext c_d as:
$$c_d = e(g, H(d))^a / e(g_1, g_x)^a$$
 - (c) Note that c_1, c_2 are public and c_1, c_2 are stored in cloud server.
4. Extracting (mk, S) : This extracting algorithm is used by the DO to produce an aggregate key.
 - (a) Aggregate key is calculated as follows: $AggKey = \prod_{j \in S} g_{x+1-j}^\eta$.
 - (b) Data owner sends this $AggKey$ and set S to the user.
5. BuildTrapdoor $(AggKey, d)$: This algorithm is used by the user to build a trapdoor to perform a search on document through the keyword.
 - (a) $T = AggKey * H(d)$
 - (b) User sends (T, S) to cloud server.
6. Adjusting $(sysparams, i, S, T)$: This adjusting algorithm will be used by the server to generate correct trapdoor, which can be determined as follows:
 - (a) $T_i = T * \prod_{j \in S, j \neq i} g_{x+1-j+i}$

- (b) To complete the keyword search, the cloud server employs the Testing algorithm.
- 7. Testing (T_i, i): This is used by cloud server to conduct a keyword search on the i th document and returns true or false results by judging:
 - (a) If $c_d = e(T_i, c_1) / e(pub, c_2)$ where $pub = \prod_{j \in S} g_{x+1-j}$

Drawback of existing scheme: The aggregate key is shared before the user's request, which is the main disadvantage of KASE scheme. Even after refusing user access, this may lead to key misuse.

4 Proposed Scheme

The design of our TOTPV-KASE scheme was inspired by the key-aggregate searchable encryption (KASE) for Group Data Sharing via Cloud Storage [33]. Rather, utilizing many individual keys, we adapt the idea proposed in KASE to generate aggregate searchable encryption environment.

Why TOTPV-KASE scheme: We have improved KASE scheme by implementing an extra layer of protection to cloud data with a time-based one-time password. Furthermore, to check the decrypted file's integrity, we introduced an interplanetary file system (IPFS) that matches the original file's hash value to the decrypted file. We have developed an application to facilitate all the features.

4.1 Functionality of Our Proposed Scheme

1. Every user who wishes to use the cloud services, i.e., to store their files or access other files, must register to the cloud with their Name, Username, Password, Mail ID, contact number, and Group Information.
2. Consider a few files owned by User-1, who is the owner of the data. User-2 is assumed to have the requirement of User-1 files. User-1 is considered Data Owner, and User-2 is considered Data User.
3. To store files in the cloud, we have adopted the same encryption scheme used in KASE, and a detailed encryption algorithm is mentioned in the above Algorithms in the KASE scheme section. User-1 (DO) encrypts the file using the algorithm Encrypt () .
4. User-2 will enter the corresponding keywords of the file and submit a request to the cloud.
5. Cloud server now searches for the keywords in all the files present in it. If the keywords are matched, the server sends a request to all the Data Owners (including User-1) with the respective File ID and the name of the data user (User2).

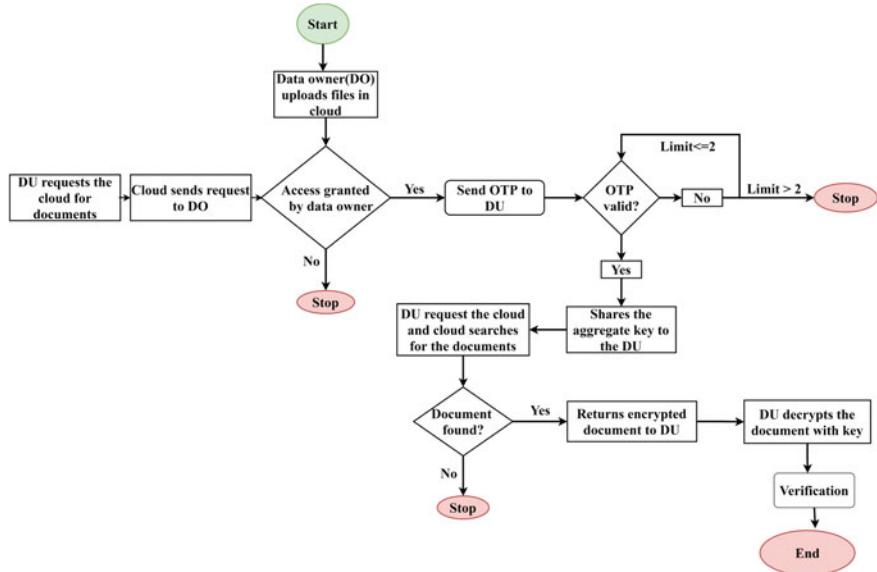


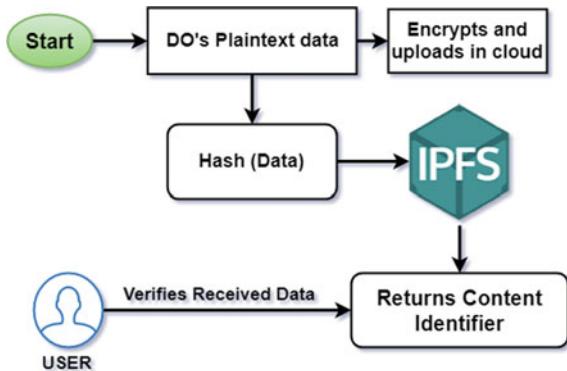
Fig. 3 Flow of TOTPV-KASE scheme

6. Once User-1 (DO) approves the permissions to User-2 (DU), a time-based one-time password (TOTP—valid for 10 min) is sent to the registered Mail ID of User-2.
7. If User-2 enters a valid TOTP within the constrained time, User-1 sends the aggregate key of the file to User-2.
8. A trapdoor is created with regard to the aggregate key and sent to the cloud again, and the encrypted file with appropriate permission is retrieved by User-2.
9. User-2 decrypts the file with the aggregate key, followed by the verification process mentioned in [Sect. 4.2].

4.2 Verification Process

- Inter Planetary File System (IPFS) is a distributed storage system that takes data and returns a Content Identifier (CID).
- In our case, User (DO) calculates the SHA-256 hash of the original data and stores it in IPFS before starting the encryption process. IPFS will return the CID of the original data, and the value is stored in the cloud. Our application will internally calculate the SHA-256 hash of decrypted data from User-2 and redirect to the IPFS page corresponding to the CID stored and checks for SHA-256, which on the proper match will display a message saying "Data is not Tampered" (Fig. 4).

Fig. 4 Integrity check using IPFS



Algorithm 1 Verification algorithm

```

1: data  $\leftarrow$  OriginalFilefromDO
2: H (data)  $\leftarrow$  SHA - 256(data)
3: ipfs add H (data)
4: CID  $\leftarrow$  IPFS
5: CID is shared with DU
6: if H (Decrypted Data) == Content(CID) then
7:   "Data is not tampered"
8: else
9:   "Data is tampered"
10: end if
  
```

We can also check this by using public IPFS by giving the respective CID of the file in the URL.

<https://ipfs.io/ipfs/CID>—Gives the hash of file that is associated with the CID which was added by the data owner (Table 2).

5 Conclusion and Future Work

In this paper, we have introduced a new scheme called TOTP-V-KASE, which provides an additional layer of security offered in searchable encryption schemes for cloud services. We have introduced TOTP-V for validating access to appropriate users and IPFS for integrity checking of data over the cloud. We have now developed our model for a single-data owner and user. In future, we would like to offer our model as a plug-and-use service by adding any number of data owners and data users.

Table 2 ID associated with hash of files

File ID	SHA-256 hash	CID
1	05E25B8EE44E4BD54F085206A3925E172 AF91D20CBF02EC686F06DBD843AA1E0	QmUCgDvYpALSD4yRmG AGCGhBNwzFPv3jRm9J4 LxrSNz8WQ
2	6F572527F1BCCE9CE49D8BF213BAB57C B8CDEF200A31065A0BA7C6FB1D63FCC6	QmS88xqc439mjLTac3r3tN hmuNG3d7AK9DPUJCmk cjHhyZ
3	9DEFF08B1AC4DBF4E8F2450CDEE30827 51D708B345D68CF0EC97101B351E5946	QmUn3idVDjijqoPiz5K3YT znYVGLZiJxMA5J15sKy iUMJpd
4	16F41FFF07E708496CB28D397CC30202F3 5BEF459DC9B0E515F3F220DA263125	QmNyQgFGQRMpbgLZ69 HCp33W7iA8Cim54jbzQ39 Mo3gtUd
5	7D18704866817A40B39AECD4B6221CF06 596F8192A813F466AD5F6DA5DD2CAA9	QmartcwjFM2epJeZq1ogV MixvMpsbDfNpHfkooSJ 7ZPX21

References

- Sundararajan S, Narayanan H, Pavithran V, Vorungati K, Achuthan K (2011) Preventing insider attacks in the cloud. In: International conference on advances in computing and communications. Springer, pp 488–500
- Schulze H. Cloud security spotlight report. https://iaas-blog.it-grad.ru/wp-content/uploads/cloud-security-spotlight-report_28381.pdf
- Harini N, Shyamala CK, Padmanabhan TR (2011) Securing cloud environment. In: Cyber security, cyber crime and cyber forensics: applications and perspectives. IGI Global, pp 115–123
- Song DX, Wagner D, Perrig A (2000) Practical techniques for searches on encrypted data. In Proceeding 2000 IEEE symposium on security and privacy. S&P 2000. IEEE, pp 44–55
- Boneh D, Di Crescenzo G, Ostrovsky R, Persiano G (2004) Public key encryption with keyword search. In: International conference on the theory and applications of cryptographic techniques. Springer, pp 506–522
- Goh E-J (2003) Secure indexes for efficient searching on encrypted compressed data. Technical report, technical report 2003/216, cryptology ePrint Archive, 2003. <http://eprint...>
- Bösch C, Tang Q, Hartel P, Jonker W (2012) Selective document retrieval from encrypted database. In: International conference on information security. Springer, pp 224–241
- Chang Y-C, Mitzenmacher M (2005) Privacy preserving keyword searches on remote encrypted data. In International conference on applied cryptography and network security. Springer, pp 442–455
- Golle P, Staddon J, Waters B (2004) Secure conjunctive keyword search over encrypted data. In International conference on applied cryptography and network security. Springer, pp 31–45
- Curtmola R, Garay J, Kamara S, Ostrovsky R (2011) Searchable symmetric encryption: improved definitions and efficient constructions. J Comput Secur 19(5):895–934
- Kamara S, Papamanthou C, Roeder T (2012) Dynamic searchable symmetric encryption. In: Proceedings of the 2012 ACM conference on computer and communications security. pp 965–976
- Hacıgümüş H, Iyer B, Li C, Mehrotra S (2002) Executing SQL over encrypted data in the database-service-provider model. In: Proceedings of the 2002 ACM SIGMOD international conference on management of data. pp 216–227
- Hore B, Mehrotra S, Canim M, Kantarcioğlu M (2012) Secure multidimensional range queries over outsourced data. VLDB J 21(3):333–358

14. Popa RA, Redfield CMS, Zeldovich N, Balakrishnan H (2011) CryptDB: protecting confidentiality with encrypted query processing. In: Proceedings of the twenty-third ACM symposium on operating systems principles. pp 85–100
15. Moataz T, Shikfa A, Cuppens-Boulahia N, Cuppens F (2013) Semantic search over encrypted data. In: ICT 2013. IEEE, pp 1–5
16. Saleem M, Warsi MR, Khan NS (2017) Secure metadata based search over encrypted cloud data supporting similarity ranking. *Int J Comput Sci Inf Sec* 15(3):353
17. Salehi MA, Caldwell T, Fernandez A, Mickiewicz E, Rozier E, Zonouz S, Redberg D (2014) Reseed: regular expression search over encrypted data in the cloud. In: 2014 IEEE 7th international conference on cloud computing. IEEE, pp 673–680
18. Sun X, Zhu Y, Xia Z, Chen L (2014) Privacy preserving keyword based semantic search over encrypted cloud data. *Int J Secur Appl* 8(3):9–20
19. Xia Z, Zhu Y, Sun X, Chen L (2014) Secure semantic expansion based search over encrypted cloud data supporting similarity ranking. *J Cloud Comput* 3(1):1–11
20. Kamara S, Papamanthou C (2013) Parallel and dynamic searchable symmetric encryption. In: International conference on financial cryptography and data security. Springer, pp 258–274
21. Van Liesdonk P, Sedghi S, Doumen J, Hartel P, Jonker W (2010) Computationally efficient searchable symmetric encryption. In: Workshop on secure data management. Springer, pp 87–100
22. Park DJ, Kim K, Lee PJ (2004) Public key encryption with conjunctive field keyword search. In: International workshop on information security applications. Springer, pp 73–86
23. Baek J, Safavi-Naini R, Susilo W (2008) Public key encryption with keyword search revisited. In: International conference on computational science and its applications. Springer, pp 1249–1259
24. Bösch C, Hartel P, Jonker W, Peter A (2014) A survey of provably secure searchable encryption. *ACM Comput Surv (CSUR)* 47(2):1–51
25. Han F, Qin J, Jiankun H (2016) Secure searches in the cloud: a survey. *Futur Gener Comput Syst* 62:66–75
26. Deng Z, Li K, Li K, Zhou J (2017) A multi-user searchable encryption scheme with keyword authorization in a cloud storage. *Futur Gener Comput Syst* 72:208–218
27. Yin H, Zheng Q, LuOu, Li K (2017) A query privacy-enhanced and secure search scheme over encrypted data in cloud computing. *J Comput Syst Sci* 90:14–27
28. Wang W, Xu P, Li H, Yang LT (2016) Secure hybrid-indexed search for high efficiency over keyword searchable ciphertexts. *Future Gener Comput Syst* 55:353–361
29. Ibraimi L, Nikova S, Hartel P, Jonker W (2011) Public-key encryption with delegated search. In: International conference on applied cryptography and network security. Springer, pp 532–549
30. Bellare M, Boldyreva A, O'Neill A (2007) Deterministic and efficiently searchable encryption. In: Annual international cryptology conference. Springer, pp 535–552
31. Dunkelman O (2012) Topics in cryptology-CT-RSA 2012: the cryptographers' track at the RSA conference 2012, San Francisco, CA, USA, February 27–March 2, 2012, Proceedings, vol 7178. Springer Science & Business Media
32. James M, Srinivasan C, Lakshmy KV, Sethumadhavan M (2016) Decrypting shared encrypted data files stored in a cloud using dynamic key aggregation. In: Computational intelligence, cyber security and computational models. Springer, pp 385–392
33. Cui B, Liu Z, Wang L (2015) Key-aggregate searchable encryption(kase) for group data sharing via cloud storage. *IEEE Trans Comput* 65(8):2374–2385

Music Genre Classification Using CNN and RNN-LSTM



Rohan Gupta, Shivam Ashish, Himanshu Shekhar,
and MS. Deepica S. Dominic

Abstract “The beautiful thing about music is that it is similar to a time machine. Only one song is needed to take a person back to a moment in time and nothing else. Music is an adhesive that holds people together.” The objective is to find a better machine learning (ML) algorithm which can predict the genre of songs. MIR is one of the greatest techniques used nowadays to get useful information from the music (audio signal). In this, we use CNN and RNN to classify the music clips. In this, we use CNN and RNN to classify the music clips. These are one of the deep learning architectures which has been immensely utilized for pattern recognition in the past. The neutral central network has a large scope to capture the informative characteristics of the music model variable with minimal prior knowledge provided. We will compare the performance of all models and record their outcomes in terms of predictive accuracy.

Keywords Machine learning algorithm · Music genre classification · Music information retrieval (MIR) · CNN · RNN

1 Introduction

In today’s digital world, there would hardly be any person who is not accessing Internet in their day-to-day lives. People are addicted to various Web sites and apps which have become an inseparable part of their lives. One such kind of apps is the music streaming apps such as YouTube Music, Spotify, Itunes, Wynk, and many more. Since the music database of such apps is really huge and ever-increasing, there is always a need to organize and categorize the music which is prevailing inside this database. One of the most prominent solutions to this problem is classification of

R. Gupta (✉) · S. Ashish · H. Shekhar · MS. D. S. Dominic
School of Computing Science and Engineering, Galgotias University, Uttar Pradesh, Greater Noida, India

MS. D. S. Dominic
e-mail: deepica.dominic@galgotiasuniversity.edu.in

music genre. While listening to online music using these apps, sometimes the songs are not categorized correctly as per their music genre collection. This problem affects the mood of the user if they are listening to songs of a specific genre, which are being automatically categorized by the app. This problem can be resolved by utilizing algorithms of machine learning. Some of the possible ML algorithms are traditional classification methods, HMM, neural network, etc. In this research, we are planning to use CNN and RNN to enhance the accuracy of our genre classification prediction model.

2 Literature Review

According to human mindset, categories of music generally represents a string of songs which expresses an overall impact and frame of mind. In this sense, special musical characteristics are expected to exist which are responsible for classifying music. Hence, overall tone of the song can be referred as harmonious, nostalgic, or vigorous. However, the determination of characteristics which can be used as an expression of a music genre is the most challenging task. Therefore, extraction of features plays a quite significant role in classification of genres. In this scenario, the different kinds of music genres should be identifiable by the extracted key musical features. The selection of the key musical feature is the cardinal factor for the success of classification task. The classifier's performance is also of utmost importance as it gives better feature selectivity.

The key features which are utilized so far for the classification are pitch, rhythm, harmony, and tone. In the past, several authors have strived to assemble these characteristics in accordance with these above-mentioned features, degree of abstraction, and degree of locality.

As stated in “Music Genre Classification using Machine Learning Techniques” by Harsh Bahuleyan which was published in [1], there are two methods for selection of key characteristics. The first method suggests creation of a spectrogram of sound wave signal. This spectrogram then has to be treated as an image. After this, an image-based classifier has to be trained on those images to distinguish the music genre. VGG-16 which is a CNN-based image classifier can be used in this method. The second method suggests extraction of time and frequency domain features from a sound wave signal. After this, a feature-based classifier has to be trained on those features to distinguish the music genre. XGBoost which is a traditional machine learning classifier can be used in this method.

As mentioned in “A Study on Music Genre Classification Based on Universal Acoustic Models” by Jeremy Reed and Chin-Hui Lee which was published in [2], a set of computer instructions can also lead to similar results as the ones received by the earlier solutions of the genre classification problem. These instructions were based on HMM modeling. According to their research if any classification is identical to language identification, then modeling of HMMs would be able to classify a single

HMM for complete language. HMMs are widely used by most of the speech applications on the phonetic level. Hence, it can be utilized to improve the performance of classification. It was also proved by this research that a similar approach can be taken for classifying music.

As proposed in the “On music genre classification via compressive sampling” by Kaichun K. Chang, Jyh-Shing Roger Jang and Costas S. Iliopoulos which was published in [3], a CS-based classifier can be used for music genre classification. They also claimed to verify the performance of this classifier on a general dataset. Usage of several feature sets was also tried to enhance the performance of classifier. These experiments confirmed that the combination of a CS-based classifier along with multiple feature sets can surpass the results and efficiency of all the older approaches.

As suggested in “Music Genre Classification Using Locality Preserving Non-Negative Tensor Factorization and Sparse Representations” by Yannis Panagakis and Constantine Kotropoulos which was published in [4], a robust framework which makes use of scattered depiction can be used for genre classification. A technique named LPNTF which assimilates the primary spatial structure of the scattered representation in accordance with the music genre into NFT was the main theme used in this framework.

3 Dataset

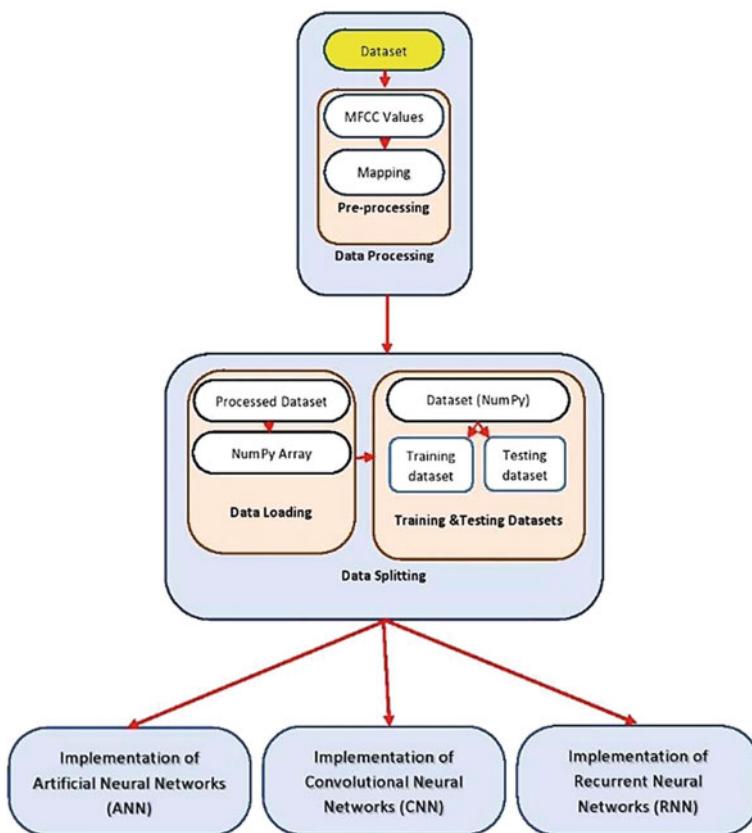
GTZAN dataset was utilized to conduct this study. This dataset primarily consists of audio files pertaining to ten music genres. There are 100 audio files of each music genre. The duration of each audio file is 30 s. The visual depiction of each audio file is also provided in the dataset. One of the approaches to distinguish music genres in this dataset is through neural networks. For this approach to work, first the audio signals has to be transformed into MEL spectrograms. Once these spectrograms are created, then these images can work as an input to neural network. This dataset also consists of two csv files containing the features of audio files. The first csv file contains the value of mean and variance which has been computed over various features of each song (30 s duration). The second csv file consists a similar pattern as the first csv file but with one difference. Before calculating the mean and variance for each song, every song has been split into ten audio files, each of 3 s duration. In this research, we would utilize the audio files which are belonging to below mentioned ten genre tags only:

|Blues – 100|
|Classical – 100|
|Disco—–100|
|Country—–100|

```
|Hip - hop--100|  
|Jazz--099|  
|Pop--100|  
|Rock--100|  
|Metal--100|  
|Reggae--100|
```

There are two csv files: features_30_sec.csv-size 1.06 MB features_3_sec.csv-size 10.56 MB.

4 Architecture



5 Methodology

In this area, the particulars of data preprocessing steps are explained. This section also contains the different approaches to this classification problem.

5.1 Data Processing

It provides a preprocessed dataset for our machine learning model. We are using labeled dataset, i.e., [5] “GTZAN” by marsyas.info. So, here we are trying to find the Mel-frequency cepstrum coefficients (MFCCs) of our data. MFCC captures timbral or textural aspects of sound. It is a frequency domain feature and works almost like a human auditory system. It resulted in a bunch of coefficient vectors, and we can calculate all of these coefficients at each frame, so that we can check how the MFCCs are evolving over time. After this, we mapped these values with their genre categories.

Spectrogram Generation [6]. Spectrogram is a 2D pictorial depiction of a signal where time is represented on x-axis and frequency is represented on y-axis . To evaluate the value of a given frequency within a provided time window, a color map is used. During this research, all the audio signals are converted into multiple MEL spectrograms. These spectrograms consist of Mel-frequency bins which are represented on y-axis. The below mentioned factors were taken into consideration while generating power spectrogram using STFT:

- Sampling rate (SRate) = 22,050
- Frame or window size (fez) = 2048
- Time advance b/w frames = 512
- Freq. Scale: Mel
- No. of Mel: 96
- Frequency(max) = srate/2.

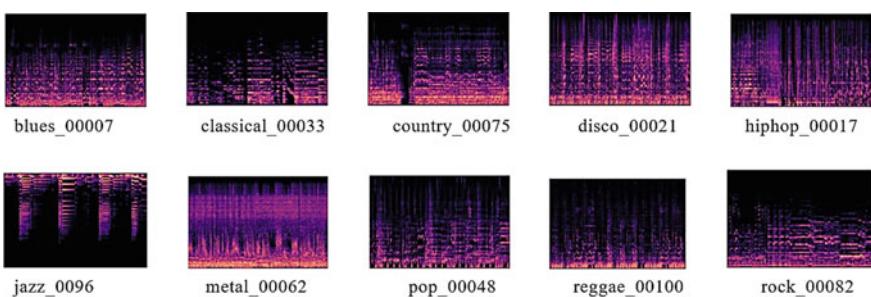


Fig. 1 Spectrograms for a sample audio data

Mel-Frequency Cepstrum Coefficients (MFCCs). The MFCC feature extraction technique basically includes windowing the signal, applying the discrete Fourier transform (DFT), taking the log of the magnitude, and then warping the frequencies on a Mel scale, followed by applying the inverse discrete cosine transform [7] (DCT).

Discrete Fourier transform (DFT). The DFT converts a finite sequence of equally spaced samples of a function into a same-length sequence of equally spaced samples of the discrete-time Fourier transform (DTFT), which is a complex-valued function of frequency.

Discrete cosine transform (DCT). A discrete cosine transform (DCT) is the sum of cosine functions expressed form sequence of data points oscillating at different frequencies.

5.2 *Data Splitting*

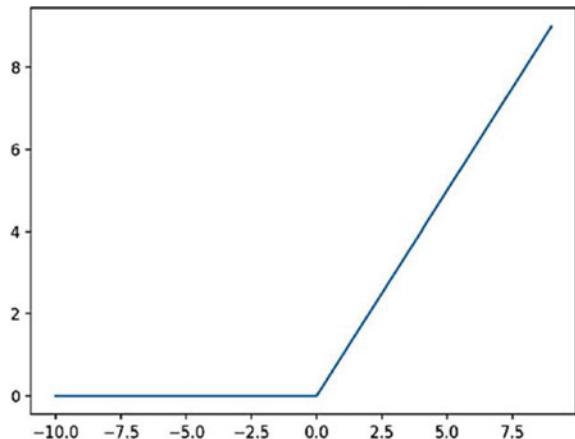
In this step, the preprocessed dataset is split up into two parts—training dataset and testing dataset. Firstly, we had to load the data into the model, so we convert it into NumPy array. This provides ease to access the data and some extra details too. Then, we split the dataset, it should be randomized data for better model training. For this, we had to import some modules from Scikit-learn (or sklearn) to split the dataset. It automatically took random data form dataset and create training and testing dataset.

Scikit-learn. Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It consists of several models/algorithms for both supervised as well as unsupervised training along with DBSCAN algorithm. It is devised to interact with Python libraries such as NumPy and SciPy.

5.3 *Artificial Neural Network (ANN)*

As we implemented multi-class classification which is opposite to binary classification, we select the multilayer perceptron which is a simple neural network. But the question arises why do we need a neural network? Is not an artificial neuron good enough? So, the answer to this question is that if we want to work on quite complex problems like real-world problems, then we need to scale up our network. A bigger network of neurons is required. Since a single neuron is highly efficient in dealing with simple problems which are associated with linearity but in this situation, we are dealing with more complex problem which is associated to nonlinearity. Therefore, we need a network of neurons which work together. These networks can produce highly nonlinear functions which can help us in solving highly nonlinear problems.

Fig. 2 Rectified linear (ReLU) activation function



Neural Architecture

Sequence Model. Sequence models are the machine learning models that input or output sequences of data. Sequential data includes text streams, audio clips, video clips, time series data, and etc. So here we used an input layer, three hidden layers, and an output layer. The input layer that we used here is flatten. Hidden layers are the simple dense layers.

Activation function. In hidden layer, we generally use the Sigmoid function, but this time we used a new type of activation function that is called ReLU. ReLU function is highly effective in deep learning. It enables us to train the network way faster than the sigmoid function. As a result, we get better convergence of the network. This is because ReLU function reduces the probabilities of vanishing gradient. In output layer, the activation function used is softmax (Fig. 2).

$$\text{ReLU}(h) = \begin{cases} 0, & \text{if } h < 0 \\ h, & \text{if } h \geq 0 \end{cases} \quad (1)$$

Adam Optimizer. Adam is an optimizer that is basically like a variation of a stochastic gradient descent for training deep learning models. The best properties of AdaGrad and RMSProp algorithms are combined by this optimizer to provide a new algorithm. This new algorithm can handle sparse gradients on noisy problems. One more advantage of using this optimizer is its ease of configuration as most of the default configuration parameters work on most problems.

Batching. Batching is the process of identifying the number of samples to be used from the training dataset. These samples then would be used to calculate gradient. There are different types of batching which is used to retrain a keras network.

- *Stochastic Gradient Descent.* To calculate the gradient in this method, only one sample segment of the dataset is considered. So, a feedforward is performed which is followed by a back propagation, and hence, the gradient is calculated. After

that the weights are updated directly. This method is quite quick to perform, but it gives highly inaccurate results. It is due to the presence of huge amount of noise which is equivalent to a scenario where batch size is equal to one.

- *Full batching.* To calculate the gradient in this method, the complete training dataset is used. So, the full training dataset is passed through the model and the gradient is calculated. After that, the weights are updated on the whole training set. The biggest disadvantage of this method is that usually datasets are really huge in size which makes this complete process super slow and highly memory intensive. It is almost impractical. The biggest advantage of this method is that results are highly accurate because we are calculating the gradient on all the samples in the training set.
- *Mini-batch.* To calculate the gradient in this method, a subset of training set is used. This subset is a set of 16–128 samples which are selected from the training dataset. Hence, this method is called a mini-batch. Once the subset is finalized, the gradient is calculated for all the chosen samples and weights are updated afterward. The no. of samples to be selected in the subset depends on the type of problem that is being dealt with. This method is basically a middle ground of both the above-mentioned methods. So, the process is relatively quick and less memory-intensive than full batch method. Also, the results are quite accurate when compared to stochastic gradient method. Hence, we use this method of batching.

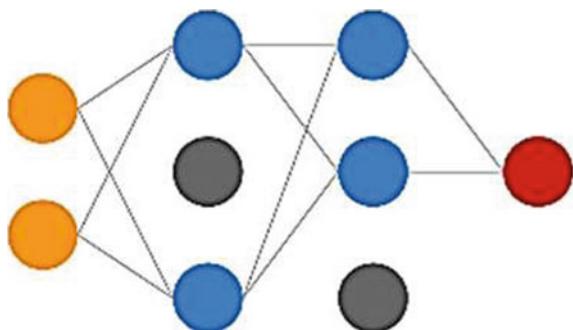
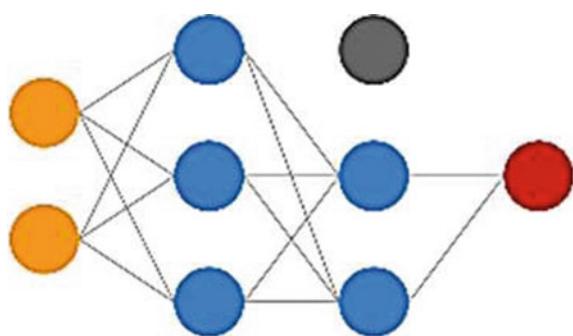
5.4 Overfitting

The result that we achieved on training dataset from neural network was highly accurate, i.e., 97%, whereas the same network did not give such accurate results when tested with full dataset. There was a huge difference of approximately 40%. Hence, we observed that it is a case of overfitting. Basically, it means that the model was performing quite well on the training set, but it is facing issues with data it had never seen before. Here, we are using two methods to remove overfitting which are as follows.

Dropout. Dropout is a technique that enables us to randomly drop neurons while training the model which in turn increases the network robustness. For example, we have our first batch of data and we decide to randomly drop certain neurons. In Fig. 3, we dropped two neurons which are shown in gray. So, all the connections will not work for these neurons, and hence, the training just happens with the remaining parts of the network.

Similarly, with the second batch of data, we can just change the dropped neurons, and this would be done randomly. We have a certain probability of dropping neurons in a layer. As shown in Fig. 4, we drop this gray neuron in this second hidden layer, and we restore the previous neurons.

Hence, we are increasing the network robustness. Since the network cannot rely too much on specific neurons, so all the neurons have to take some responsibility of

Fig. 3 Dropping of neurons**Fig. 4** Change in a dropped neuron

the prediction process. It is similar to reshaping of the network which then redistribute responsibilities to all of the neurons, so that none of them is indispensable.

Regularization. It is a technique that adds penalty to the error function. The main idea here is to punish large weights. The larger the weights, the higher would be the penalty given to the error function. There are two types of regularization that are generally used in deep learning. These are L1 and L2 regularization. In L2 regularization, we minimize the squared value of the weights as shown in formula. As a consequence, this method is way less robust to outliers, but the advantage of this method is that it can learn quite complex (Fig. 5).

$$E(p, y) = \frac{1}{2}(p - y)^2 + \lambda \sum W_i^2 \quad (2)$$

5.5 Convolutional Neural Networks (CNN)

As per methodology, anyone can recognize that different genres have different characteristic patterns. These patterns are visible in the spectrograms of these audio signals. Therefore, spectrograms are basically images which work as input to a CNN.

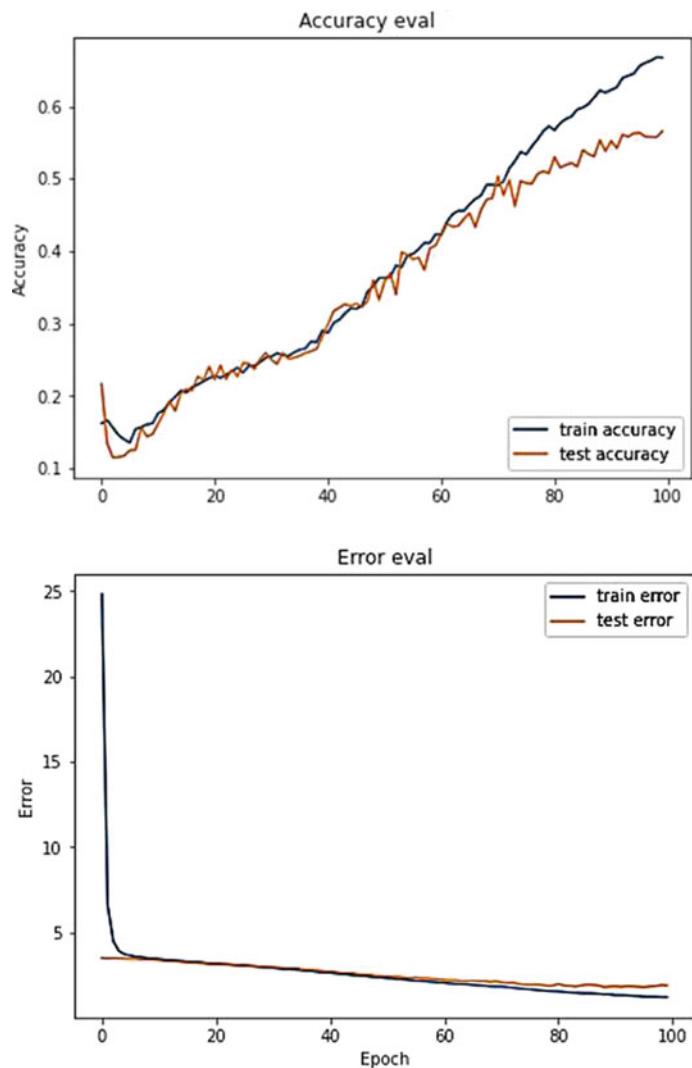


Fig. 5 ANN learning curve

CNN are the advanced type of neural networks, they are mainly used for processing images, and all the time we have found out that they perform well with images than the equivalents, for example, multilayer perception architecture. The good thing about this architecture is that they require very less parameters than dense layers.

Image data is like structured data, and pixels are not randomly positioned. There are certain emergent structures like edges, shapes, invariance to translation, and scale invariance. CNN is trying to emulate human vision system. We extract basic features

and CNN try to learn those different types of features. All of this process relies on a couple of components, i.e., convolution and pooling.

Convolution. At this stage, a matrix filter (Kernel) is slide over the input image. On the image, kernel is placed first. Then, we compute an Hadamard product of the kernel and the overspread segment of the image. The summation of these values gives us a feature value. During the training, we use many such kernels which are learned via backpropagation.

$$\sum_{i=1}^P image.K \quad (2)$$

Kernel, Kernel or a filter is a feature detector which different kernels detect different features. For example, in Fig. 6, these kernels used to identify oblique lines and vertical lines.

We have different types of architectural decision which helps to decide the type of convolution. Since, we were using spectrogram as images (colored images) hence, ‘depth’ is the key feature in this. In grayscale image, the depth is equal to one, but if you have a colored image, i.e., in RGB, each pixel has three values one for the red, one for the green, and one for the blue color. In terms of kernel, it is divided into three parts or three grids where one grid is for red, one for green, and one for the blue channel. As these grids are independent, we are getting three-dimensional array where first two dimensions represent the width and the height of the kernel and the third dimension represent the ‘depth’ (Fig. 7).

Pooling. This is the process of dimension reduction into feature map by convolution steps. This process is called as down sampling. It works in similar manner as convolutional, i.e., overlay a grid on top of an image. Generally, for deep learning, max-pooling is used with 2×2 matrix. Out of the four elements inside the matrix, the maximum value is picked up. We keep moving this window across the feature map with a predefined stride [8] (Fig. 8).

Implementation of convolution and pooling. In case of spectrograms, the time and frequency can be considered as x and y indexes for the pixels of an image. Also, amplitude can be considered as the value associated to each pixel. This concept is basically comparing a spectrogram with a 2D array. For example, we have 13 MFCCs, 512 samples as hop length, and 51,200 samples in an audio file, then the expected data shape which would be in our CNN is $100 \times 13 \times 1$. Here, 100 denotes the time windows at which we take 13 values. The 13 denotes the no. of MFCCs and 1 denotes the depth (Fig. 9).

Fig. 6 Kernel for detecting different lines

Oblique line detector	Vertical line detector
1 0 0	0 1 0
0 1 0	0 1 0
0 0 1	0 1 0

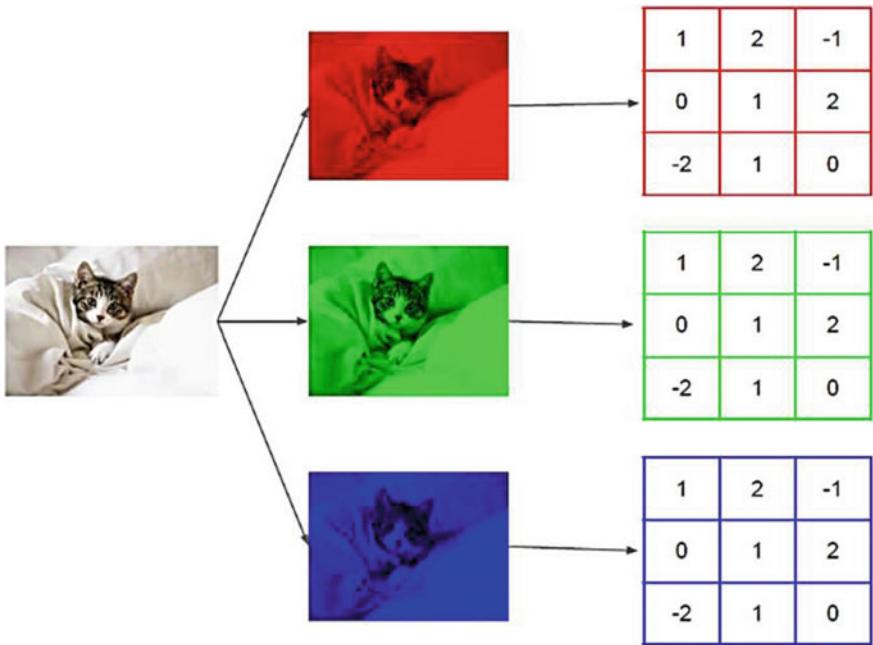


Fig. 7 Depth of a RGB image

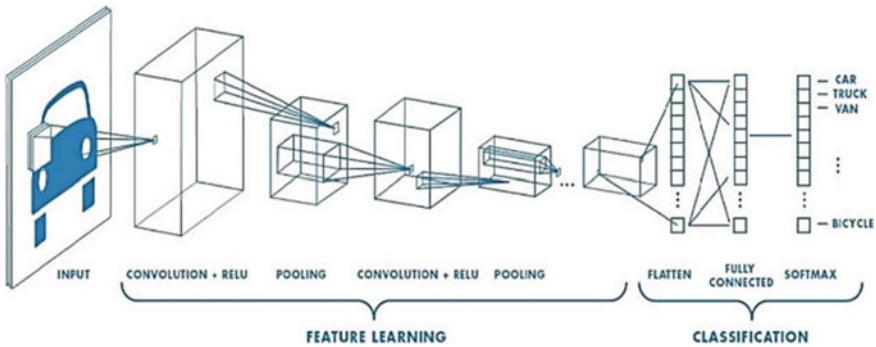


Fig. 8 Convolutional neural networks (CNN) architecture

5.6 Recurrent Neural Networks (RNN)

RNN is a process which is used for the dataset where order is extremely important. Some examples of such dataset are time series or measures of different values that are taken at fixed time intervals. The dataset can have variable amount of words/length like musical notes. Hence, this method is able to process sequential data in such

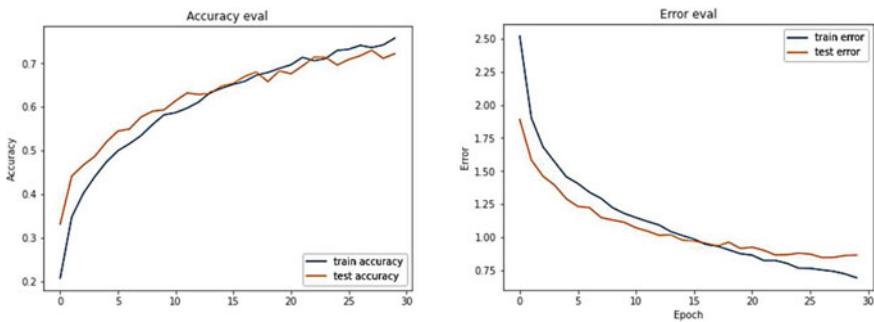


Fig. 9 CNN learning curve

a way that each data point is processed in context. It is an ideal method for audio processing.

As we know that audio files are in the format of waveform so, we can consider these as a univariate time series. It denotes that we have only one value or one measure that is taken at each interval. Contrary to this dataset, there are another type of dataset such as MFCCs which are multivariate time series. Multivariate time series are the ones where we take more measurements for each interval and the relative dimension is given by the overall number of intervals. For a time series dataset, we take input data points one at a time. Our goal is to predict next step on the basis of previous data points which helps the model in understanding context.

In RNN, recurrent layer is the layer that is able to process sequential data in a sequential way. The cell is the one that is responsible for processing this sequential data at each step. We give input data that is represented by this X_t then the cell does some processing and it gives output as Y_t and H_t . H_t is called as a state vector or a hidden state which keeps memory of the cell at a certain point in time, whereas Y_t is the actual output. The whole point of the recurrence here is given by the fact that H_t or the state vector is going to be reused at a later point of time at the next step. In this way, we can have information about the context (Fig. 10).

Memory cell is a very simple neural network, and here, we use dense layer, and the input is given by the combination of the state vector and the input data. The state

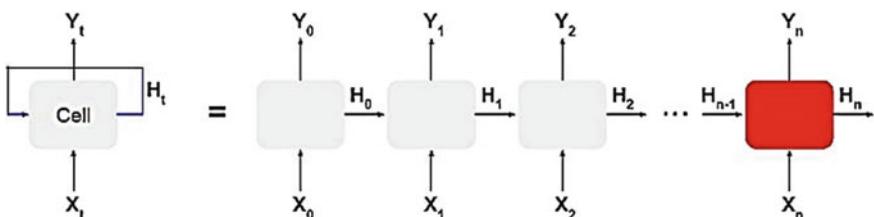


Fig. 10 Unrolling a recurrent layer

vector is used at the previous timestamp, and the input data is used at the current timestamp. The activation function that we use is the hyperbolic tangent ($\tanh H$).

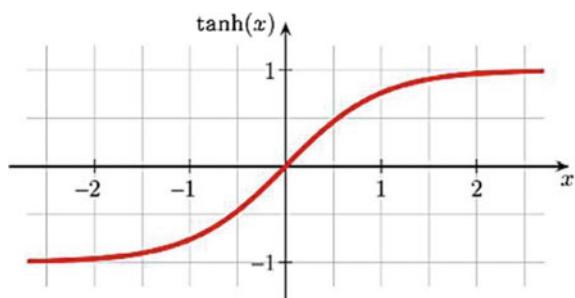
We could have used other activation functions, but it turns out that the training with RNN is quite difficult because of a number of reasons. Mainly vanishing gradients where the gradients tend to disappear, and at the same time, we have the issue of exploding gradients where the gradients grow bigger. If we use ReLU, then CNNs revenue is not bounded, but in RNN, ReLU can explode, whereas $\tanh H$ constrains the values between -1 and 1 . To train the network, we back propagate the error through time. So, RNN is unrolled and treated as feedforward network (Fig. 11).

But there are few issues with simple RNN. One of the problem is that they do not have a long-term memory. So, the network cannot use information from the distant past. Also, one of the problems is that they cannot learn patterns with long dependencies. In order to avoid these issues, a number of variants have been devised, and a specific one which is very successful is called long short-term memory network (LSTM).

Long short-term memory network (LSTM). As a solution to the shortcomings of normal RNNs, long shorts and memory networks were introduced. LSTM is a special type of recurrence neural networks where we have memory cells that enable us to learn long-term patterns. It is quite effective in detecting pattern which is up to 100 steps but it struggled with 100 s/1000 s of steps. In RNNs, the memory cell is a simple dense layer with hyperbolic tangent (\tanh) as an activation function, but in LSTM, the cell structure has been modified. The standard formulation of a single LSTM cell can be given by following equations [9] (Fig. 12).

where i , f , o , C , and \tilde{C} are the input gate, forget gate, output gate, memory cell content, and new memory cell content, respectively. σ is the sigmoid function, and \tanh is the hyperbolic tangent function. The sigmoid function is used to form three gates in the memory cell. On the other hand, the \tanh function is used to scale up the output of a particular memory cell. The sigmoid function ranks the output between zero and one. Hence, the values which are closer to zero will be forgotten in the cell state, whereas the values which are closer to one will be kept. The \tanh function is used to scale up the output of a particular memory cell (Fig. 13).

Fig. 11 Hyperbolic tangent ($\tanh H$) activation function



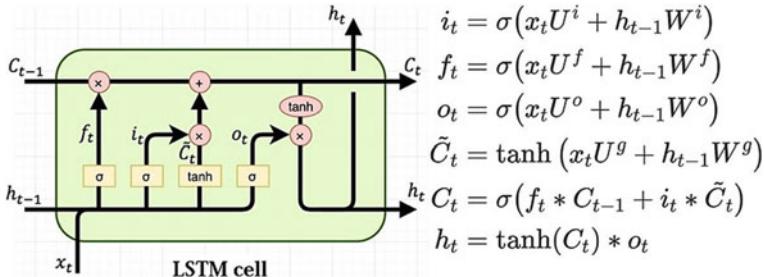


Fig. 12 LSTM cell and equations

6 Result

We decided to use supervised learning dataset for our research. So, we utilized the same dataset to train all three models—ANN, CNN, and RNN. By following this approach, we were able to calculate the accuracy percentage. Once evaluated, we compared this percentage among all the models. As a result, we found that accuracy percentage is highest in RNN model.

7 Conclusion

This research is carried out on an audio dataset named “GTZAN” provided by marsyas.info. Here, spectrograms are generated for all the audio signals in a dataset which are then considered as images. Later, MFCCs are calculated from these images to identify the music genres of these audio files. The MFCCs are evaluated by extraction of time domain and frequency domain characteristics from these images. We used two different approaches—CNN and RNN. We observed that the results obtained from CNN-based deep learning models had better accuracy than the results of feature engineered models.

We also noticed some key features of CNN such as it requires very few parameters as compared to dense layer. Also the use of Kernels which works as a filter helps in easier detection of features of an image. Similarly, we observed that the RNN-based deep learning models are capable of predicting the next step based on the previous data points which helps the model in understanding context. However, simple RNN models do not have long-term memories due to which these models are unable to understand patterns with long dependencies. To mitigate this problem, long short-term memory (LSTM) is used which is quite effective in detecting pattern up to 100 steps. The final results of ANN, CNN, and RNN are shown in Figs. 5, 9, and 12, respectively. Hence, we can conclude that using RNN-based deep learning model along with LSTM is the better approach to classify music genres in an audio dataset.

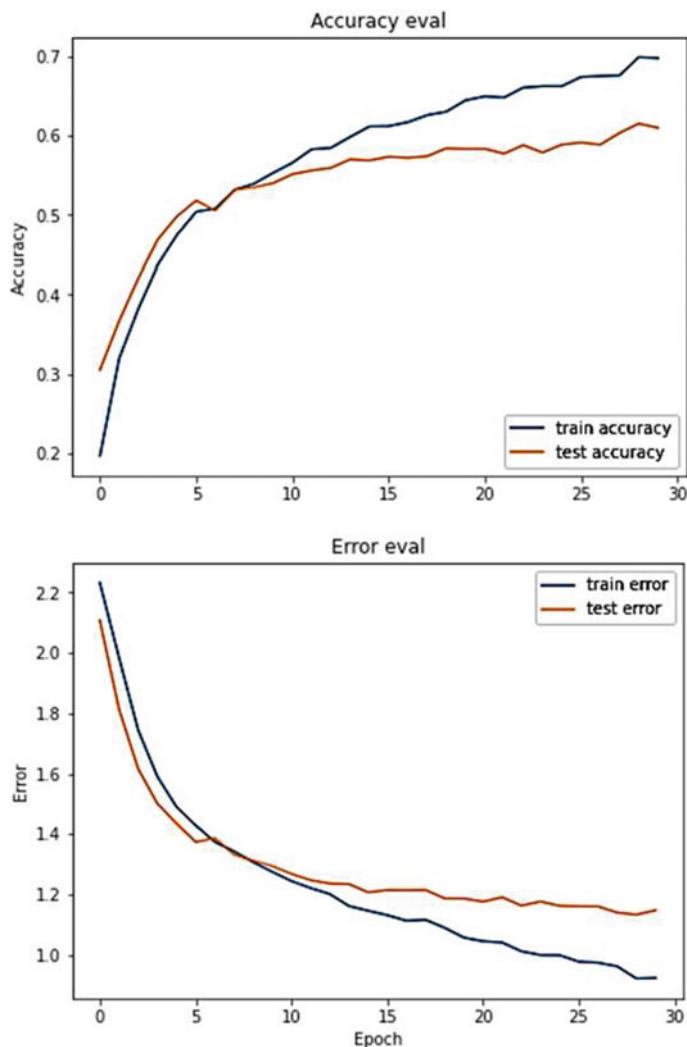


Fig. 13 RNN–LSTM learning curve

References

1. Bahuleyan H (2018) Music genre classification using machine learning techniques, Sound (cs.SD); Audio and speech processing (eess.AS)
2. Reed J, Lee C-H (2006) A study on music genre classification based on universal acoustic models. ISMIR
3. Chang KK, Jang J-SR, Iliopoulos CS (2013) On music genre classification via compressive sampling. 2013 IEEE international conference on multimedia and Expo (ICME)
4. Panagakis Y, Kotropoulos C, Kotropoulos C (2009) Music genre classification using locality preserving non-negative tensor factorization and sparse representations. ISMIR

5. Music Genre Classification. <https://www.kaggle.com/c/music-genre-classification/datasets>
6. Bahuleyan H (2018) Music genre classification using machine learning techniques. arXiv.org
7. Lachmish M (2018) Music genre classification. <https://medium.com/@matanlachmish/music-genre-classification-470aac983d>
8. Saha S (2018) A comprehensive guide to convolutional neural networks. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
9. Shewalkar A, Nyavanandi D, Ludwig SA (2019) Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM AND GRU. JAISCR 9(4):235–245

House Prices Using Machine Learning Algorithms



Vishal, Amit Singh, and Chirag Chaudhary

Abstract Most people's bucket list has one thing in common, i.e., having a dream house. People consider buying houses on various factors such as surroundings and locality. Budget is the main consideration when people lookout for buying their dream house. Housing prices were projected in this project using numerous variables covering fourteen characteristics relating to residential homes, the landscape, and their surroundings. The purpose of this project is to construct a model that is ready to estimate the value of the house accurately based upon features. Three machine learning algorithms are used to compared, i.e., linear regression, random forest regression, and Python decision tree regression using Python libraries such as pandas and matplotlib. Predicting home prices using machine learning algorithms is the fastest way to forecast the price of homes. As independent variables for forecasting house prices, this study uses 14 attributes or characteristics. It uses characteristics such as crime rate, population density, and residential land proportion.

Keywords House price prediction · Machine learning · Python

1 Introduction

Predictive models are used to achieve results at the heart of technological advances, which is data. The methodology most commonly used in this method is machine learning. Machine learning means having a reliable dataset and further supporting predictions that the machine itself learns the proportion of significance of a particular event that the whole system supports its pre-loaded data and predicts the outcome accordingly.

This system's numerous modern applications include predicting stock values, predicting the possibility of an earthquake, and predicting company profits, and thus, the list has endless possibilities which are numerous modern applications of this scheme. We have considered Boston as our primary location for our experimental

Vishal (✉) · A. Singh · C. Chaudhary
School of Computer Science and Engineering, Galgotias University, Uttar Pradesh, Greater Noida, India

research and are forecasting real-time house prices for several locations in and around Boston. Parameters such as “crime rate in each capital,” proportion of non-retail company acres, have been used, in order to give correct results for all circumstances, we have taken a validated dataset with diversity into consideration.

We used the ‘Boston House Prices’ dataset in this project, fetched and uploaded by Dr. Jason to Kaggle [1]. We may verify the output of each individual approach by applying multiple techniques on this dataset. The paper is structured as follows: Sect. 2 outlines the literature review; Sect. 3 outlines the functioning of the proposed model; Sect. 4 outlines the module implementation; Sect. 5 defines the merits of the proposed system; and Sect. 6 draws a conclusion, which is accompanied by acknowledgment at the end.

2 Literature Review

In 2019, Data Science Challenge competition was held at Engineering Education (EEF) [2]. For the competition, the training and testing data was to be generated from more than twelve million housing advertisements which were collected from Web sites of Brazil. Dataset consists attributes generated from 4 years of housing advertisements [2015–2018], and every home was available, none for the rent. The competition includes predicting prices using ML models.

Statistical models are a standard research approach which over an extended period of time forecast property costs. In a report to explain the difference in housing prices, the impact of place choices on the cost of land was studied [2]. In Metropolis County, USA, Benny Goodman and Thibodeau used stratified linear models to quantify housing prices, mistreatment data extracted between 1995 and 1997 from twenty-eight thousand single-family transactions. Benny Goodman and Thibodeau claim that the housing market is segmental by the standard of public education, which is not properly a property attribute, among the contributions from this report.

Gupta and Das utilized Spatial Bayesian VARs (BVARs) in 2010 to foresee the downturn within the rate of development in genuine domain costs for the twenty biggest US states [3]. To estimate the lull, it was generally centered on month to month real house cost development rates. They know that the unit of BVAR models is well-equipped to state the long-term direction of genuine house costs, and they enormously think little of the diminish.

There is an immersive writing of literature taking into account U.S. house values. In 2007, Rapach Strauss used an ARDL concept system and observed that a benchmark AR model could be beaten by ARDL models. With real loading value growth, the auto-regressive dispersed slack model holds twenty-five determinants.

3 Methodology

Machine learning is of mainly three types—supervised, unsupervised, and reinforcement learning (reward-based learning). This project is based on supervised machine learning, and the model is given labeled/structured input data as well as expected outputs. This data is then divided into training set and test set in the ratio of 80:20. This model is trained until it can accurately predict or identify something. Accuracy of this model is calculated by using it on test set. It yields to good results when provided with data that is never seen before by the model.

Classification and regression are the two subcategories of supervised machine learning [4]. Regression is a static approach for modeling the relationship between a dependent or independent variable and one or more dependent or independent variables. It aids in the prediction of continuous variables.

e.g., Weather forecast.

We have used three regression-based models:

Linear regression.

Decision tree regression.

Random forest regression.

3.1 Architecture Diagram

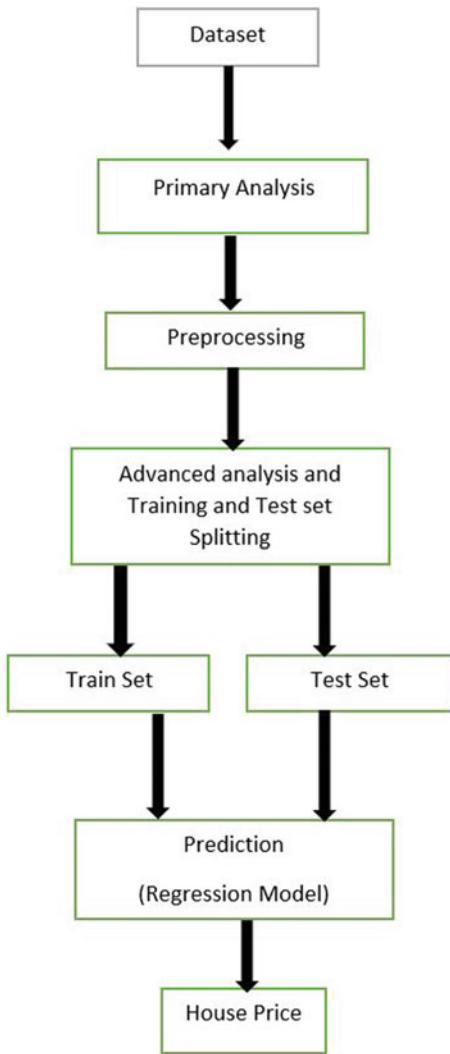
See Fig. 1.

3.2 Working

For the working of this project, we have taken a dataset having information related to attributes. Dataset is then preprocessed to make it efficient for the purpose. Data is then divided into train set and test set. Test set is kept untouched and is used for checking the accuracy later. Relations are gathered between different attributes. Also, different attribute combinations are tried in order to predict more precisely. Finding of missing attributes is done. After everything is done, different models are tested on the test set. Evaluation of models is done, and a desired model is selected that can be used for predicting the prices.

3.3 Selecting a Model

Based on the results of these three models, random forest regression model is considered best suitable for predicting the price of houses due to its least mean and median

Fig. 1 Architecture diagram

values among the selected models. A single decision tree has high variance, but when we combine multiple decision trees (random forest) resultant variance becomes low and yields to more precise values.

4 Module Implementation

In this project, Python is used for programming purpose as it has a support of various libraries such as NumPy, pandas, matplotlib, and sklearn. The software used for this

Table 1 Description of attributes

S. No.	Code	Description
1	CRIM	Crime rate
2	ZN	Residential land zone population
3	INDUS	Land occupied for business proportion
4	NOX	Nitrogen oxide index
5	CHAS	Close to river
6	RM	No. of rooms
7	AGE	Property occupied by owners
8	DIS	Weighted distance from Govt. offices
9	RAD	Distance from highway
10	TAX	Tax rate per \$1000
11	PTRATIO	Student teacher ratio
12	LSTAT	Population status
13	B	Proportion of blacks
14	MEDV	Owner occupied homes median value

project is Google Colaboratory, and the operating system used is windows. Below is the list of 14 attributes used in this machine learning project (Table 1).

4.1 Data Preprocessing

The preprocessing of data is the most critical activity. The data provided to the model is not always ready to use. There may be some missing values or attributes. It plays an important role in accuracy. This is because information is mostly chaotic and has some incomplete or incorrect values often. The knowledge can be in several different ways. Machines only understand binary, i.e., the 1 s and 0 s, so the free text, image, or video is not understood by them.

Rows in train set: 404.

Rows in test set: 102.

4.2 Data Splitting

This module split the preprocessed dataset into training and testing dataset. Firstly, we had to load the dataset into the model, so we convert it into NumPy array. This provides ease to access the information and a few extra details too. Then, we split the dataset, it should be randomized data for better model training. For this, we had to import some modules from Scikit-Learn (or sklearn) to separate the dataset. It

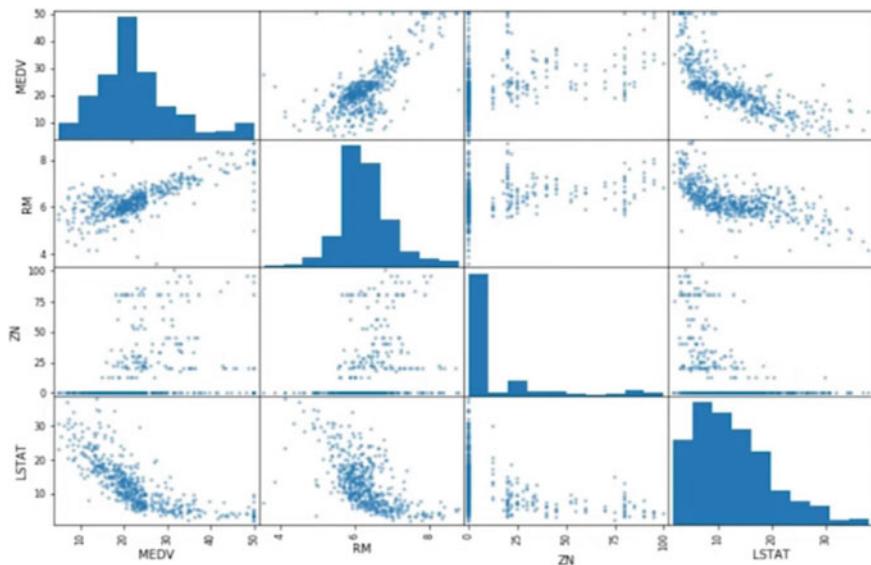


Fig. 2 Plotting of attributes

automatically took random data form dataset and create training and testing dataset [5]. In this project, we split our datasets for training and testing in the ratio 4:1.

To use the dataset, some Python libraries are used, such as:

Numpy. NumPy also called as Numerical Python is a Python programming language library that provides a large set of functions for working with multidimensional arrays and matrices.

Scikit-learn. It is a machine learning library for Python programming. It provides many supervised and unsupervised learning algorithms and can conveniently communicate with the NumPy and SciPy Python libraries [6].

Pandas. It is an often-used Python-based fast, powerful, and easy to use data analysis and manipulation tool. It is mainly used to read, sort, and filter files from csv/excel.

4.3 Finding Co-Relations

Finding co-relation can be defined as comparing attributes using histograms and graphs. It helps in finding strong co-relation between attributes, and it can be both positive or negative. It helps in better understanding of dataset. Outlet values can be removed. It helps in increasing the accuracy of model and helps in predicting more precise values (Fig. 3).

Trying different attributes combinations can also results in enhancing the accuracy of model. Below is the graph plot of MEDV and TAXRM (Fig. 4).

Fig. 3 Strongly positive co-relation between MEDV and RM

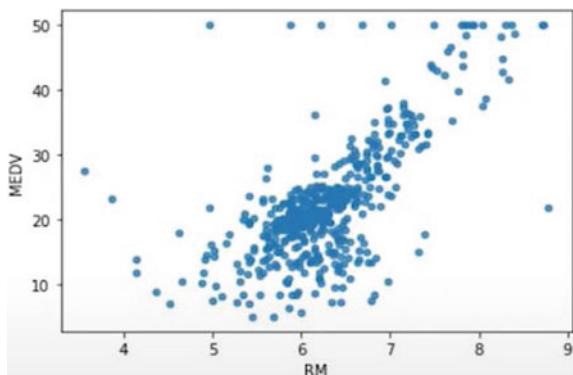
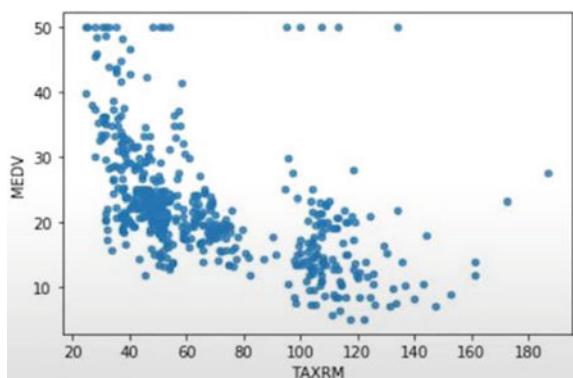


Fig. 4 Strongly negative co-relation between MEDV and TAXRM



Here, TAXRM is a combination of TAX and RM attribute and results in strong negative co-relation.

4.4 Finding Missing Attributes

Data is usually not clean, and you may always have attributes that are corrupt or lacking. In order to induce the easiest results, it is important to acknowledge, label, and manage missing data while designing machine learning models.

To take care of missing attributes, you have three options:

Get rid of the missing data points.

Get rid of the whole attribute.

Set the value to some value (0, mean or median).

4.5 Feature Scaling

ML is like producing a fruit crush that is blended. We want to combine all fruits not consistent with their size, but supported their proper proportion, if we would like to urge the simplest mixed juice. We just must note that grapes and oranges do not seem to be the identical unless in some context we make them the image of compare their attributes. Similarly, in many machine learning algorithms, we wish to undertake and do scaling so on place all features to the identical place so a substantial number does not influence the model only due to their large magnitude.

During the preprocessing of data before implementing a machine learning model, feature scaling in machine learning is one altogether the foremost important steps. Scaling may make a distinction between a poor model of machine learning and the next one.

Primarily, two types of feature scaling methods are used:

Min–max scaling

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

Standardization

$$X' = \frac{X - \mu}{\sigma} \quad (2)$$

4.6 Performance Measure

The root mean square deviation (RMSD) is used for the standard regression problem output metric. The output metric for regression tasks is usually RMSD. So, for this unique question, we choose it. Other metrics for results include mean absolute error, Manhattan Norm, etc. We are going to use RMSD for this problem, though.

$$RMSD = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}} \quad (3)$$

4.7 Overfitting

The result which we achieved using decision tree was highly accurate, i.e., 99% approx. for the training dataset but gives a difference of approx. 37% while tested on full dataset. This means that the data is overfitted. Overfitting basically refers that

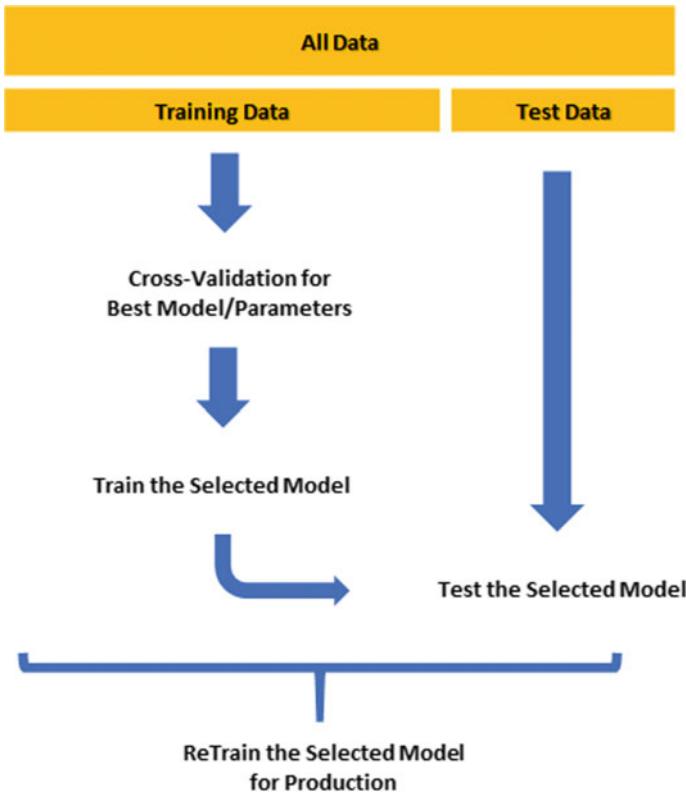


Fig. 5 Cross-validation technique

the model is performing quite well on training dataset but faces issue with data that is never seen before.

Hence, we have used cross-validation technique to prevent overfitting.

Cross-Validation. This technique is used to evaluate machine learning models by grouping them in subsets and then evaluating. For example, data is divided into three subsets. Firstly, first and second subset are taken for training and third for testing, and then, first and third for training and second for testing and so on. It helps in preventing overfitting of data (Fig. 5).

4.8 Feature Scaling

Below are the three regression models we have used.

Linear Regression. Linear regression is used to identify linear relation between dependent and independent variables [5]. Dependent variables are continuous in nature.

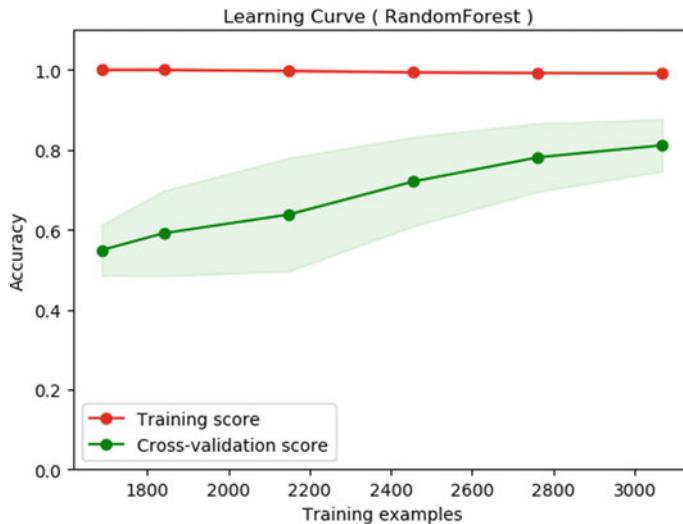


Fig. 6 Random forest learning curve

Decision Tree Regression. Decision tree regression forms a hierachal structure based on decisions, and root nodes are used in determining values. Decision tree generally has high variance which leads to overfitting of train data.

Random Forest Regression. Random forest regression can be described as a bunch of decision trees formed by row sampling and feature sampling of dataset also known as Bagging [7].

Single decision tree generally has high variance, but when a bunch of decision tree is created, variance becomes low.

For classification purpose, we consider majority output as the final output, and for regression problem, we use mean/median as the final output in random forest (Fig. 6).

4.9 Selecting a Model

Based on the results of these three models, random forest regression model is considered best suitable for predicting the price of houses due to its least mean and median values among the selected models. A single decision tree has high variance, but when we combine multiple decision trees (random forest), resultant variance becomes low and yields to more precise values (Table 2).

Table 2 Comparative study

Models	Linear regression	Decision tree	Random forest
Mean error	4.221894675406022	4.189504502474483	3.494650261111624
Standard deviation	0.7520304927151625	0.848096620323756	0.762041223886678

5 Merits of Proposed System

Fast—Better for real-time processing.

Reliable—The machine learning model selected is reliable enough.

Precise—The machine learning model selected is very precise and can be trained end-to-end in order to gain accuracy.

6 Conclusion and Further Scope

Nowadays, everything is shifting from manual to automated systems. Doing work manually is difficult, takes more time and still not very accurate, and the amount of knowledge we gain is less with respect to time. Whether you take searching or you are researching on something it will take time to put your hands-on keyboard and type each word and mean while you will lose interest on that topic. So here main issue is to create machine that will efficiently provide much accurate results and keep things interesting and more knowledge in less time. By using machine learning algorithms predicting prices of houses become reliable, convenient, more precise and even very much faster than manually predicting. Also, accuracy of these models can be enhanced by using large datasets.

A real-estate Web application can be made from this model which can help users find house in their budget with ease. It will also work as attracting more buyers on a real-estate Web site.

Acknowledgements The author would really like to precise a deep sense of gratitude and thank our mentor Mr. Surendra Kumar, Assistant Professor, Galgotias University, Greater Noida, for his useful advice and input, without whose permission, wise counsel and able guidance, it would have not been possible to hold out our research during this manner. Finally, I express my indebtedness to any or all who have directly or indirectly contributing to form our research successful.

References

1. Boston house price prediction. <https://www.kaggle.com/shreayan98c/boston-house-price-prediction>
2. Sousa S, Dihanster W, Klaus B Housing prices prediction with a deep learning and random forest ensemble

3. Satish GN, Raghavendran CV, Sugnana Rao MD, Srinivasulu C (2019) Int J Innov Technol Exploring Eng (IJITEE) 8(9):2278–3075
4. Real estate price prediction with regression and classification, CS 229 Autumn 2016 Project final report
5. Montgomery DC, Peck EA, Vining GG (2015) Introduction to linear regression analysis
6. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
7. Breiman L. Random forests. SpringerLink. <https://doi.org/10.1023/A:1010933404324>
8. Fan C, Cui Z, Zhong X. With machine learning algorithms, house prices prediction. Proceedings of the 10th international machine learning and computing conference 2018—ICMLC 2018. 3195106.3195133. 10.1145

HEVC Encoding and Decoding using Fast Algorithm for Intra Frame Partition-Prediction Using DTCWT



V. Madhurima and K. Padmapriya

Abstract Video compression techniques need to be robust, flexible and provide higher coding efficiency in order to meet the increasing demand of bit rate to support transmission and storage of high resolution videos. High efficiency video coding (HEVC) is significantly increasing the coding efficiency of its ancestor H.264/Advance Video Coding. Evaluation of partition-prediction algorithm proposed in this work is carried out by comparing the MSE and PSNR results with two other methods. The direct partition-prediction method and DWT-based partition-prediction method results are compared with the proposed results. Four different data sets are considered for evaluation, and the prediction is obtained for all 33 modes. Dual-tree complex wavelet (DTCWT)-based partition-prediction algorithm output is able to reconstruct all features as compared with DWT-based algorithm. Thus, the proposed partition-prediction algorithm is able to retrieve all the required features during decomposition and increases limited redundancies in the DTCWT coefficients, thus enhancing the coding efficiency in HEVC.

Keywords Dual-tree complex wavelet · Intra-frame prediction · Coding efficiency · HEVC · Kingsbury filters

1 Introduction

Video compression techniques need to be robust, flexible and provide higher coding efficiency in order to meet the increasing demand of bit rate to support transmission and storage of high resolution videos. The joint activity of Joint Collaborative Team on Video Coding (JCT-VC), ITU-T and ISO/IEC JTC has resulted in HEVC standards. Coding efficiency and flexibility in HEVC standards have also improved compression performance in the range of 50% bit-rate reduction with perceptual video quality [1]. In HEVC, one of the tools is the intra-prediction process. Data prediction within the frame spatially from region to region is carried out to improve

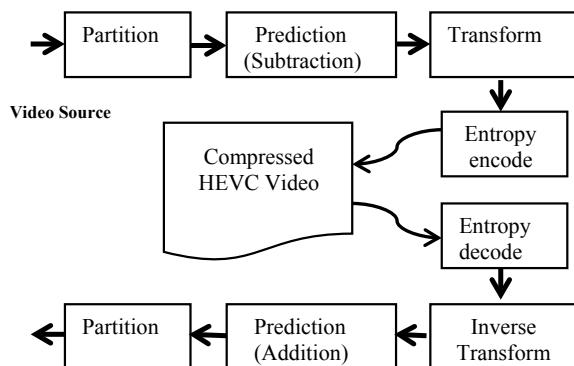
V. Madhurima (✉) · K. Padmapriya
Department of ECE, UCE, JNTUK, Kakinada, Andhra Pradesh, India

compression efficiency [2]. Encoding time in HEVC is complex due to partitioning and prediction of both intra-frame and inter-frame to improve coding efficiency. Prediction algorithm in spatial domain needs to be faster, and the work carried out by Pan presents fast algorithms that are based on histogram analysis of the edge map patterns in the image. Fangwen has improvised Pans algorithm by computing the border map considering only the probable modes of prediction. The algorithm presented by Hsien is demonstrated to be faster at it considers the symmetry in adjacent angles of directional predictions. There are also algorithms that operate in frequency band rather than spatial domain. Zhenyu has developed algorithm that performs partition and prediction considering dominant edges in the sub-bands computed using Haar transforms. Processing sub-bands in the wavelet domain is advantageous as the sub-bands localize the patterns in four different sub-bands, and processing of data becomes easier. Limitations in terms of computing time, complexity and coding efficiency are a trade-off in HEVC encoding. Improving efficiency in terms of coding gain is achieved by replacing DWT using complex wavelets. Complex wavelets are effectively used for image processing and signal processing applications, and in this paper for the first time, complex wavelets are used for intra-prediction coding process.

2 Intra-Prediction Coding

High efficiency video coding (HEVC) standards supporting all existing H.264/MPEG-4 are more flexible in defining partition size, predicting modes and transform block sizes. Use of interpolation filters, daglocking filters, motion vector estimation algorithms and parallel processing architecture is the major advances in HEVC [3]. Video coding algorithm based on HEVC standard [4] is presented in Fig. 1. The video coding model incorporates both spatial and temporal correlation between pixels, frames and objects within frames and between frames.

Fig. 1 Video coding system



M	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				

Fig. 2 Prediction of 4×4 macro-block

The input frame is partitioned into multiple sub-units, and each unit is predicted by the intra-frame prediction module, and the predicted sub-unit is subtracted from the actual sub-unit [2]. The residual of the subtraction is transformed using discrete cosine transform, and the transformed output is quantized. The transformed output, predicted units and the mode information along with the corresponding header are entropy encoded to complete the compression of the input data. The compressed data is decoded at the receiver by the HEVC decoder that performs the inverse process. Intra-frame prediction is the process of finding the predicting the pixels considering the neighbouring pixels [5]. Intra-frame prediction is carried out by predicting not individual pixels but by predicting group of pixels or macro-blocks which are defined with sizes of 16×16 , 8×8 and 4×4 . Considering the 4×4 block shown in Fig. 2, intra-frame prediction is discussed [6].

The 4×4 macro-block pixels represented by a to p are predicted considering the neighbouring pixels A to M that are along the rows and columns of the 4×4 macro-block. There are nine modes of prediction defined as 0 to 8, and the nine modes are presented in Table 1, [7]. These nine modes are also considered of the macro-block is of 8×8 size. If the macro-block is set to 16×16 , then there are four prediction modes.

In addition to three sizes of macro-blocks and eight angular prediction modes, the macro-block of size 32×32 and angular predictions in 33 directions (shown in Fig. 3) are considered in HEVC making it more efficient coding standard compared with AVC coding. In HEVC considering block size of 4×4 to 32×32 , there are four macro-blocks, and for each macro-block, the angular predictions are in 33 directions which gives rise to 132 combinations of prediction modes, and this also increases computation complexity. There needs to be a trade-off between coding efficiency and encoding complexity.

In addition to three sizes of macro-blocks and eight angular prediction modes, the macro-block of size 32×32 and angular predictions in 33 directions (shown in Fig. 3) are considered in HEVC making it more efficient coding standard compared with AVC coding. In HEVC considering block size of 4×4 to 32×32 , there are four macro-blocks, and for each macro-block, the angular predictions are in 33 directions which gives rise to 132 combinations of prediction modes, and this also increases computation complexity. There needs to be a trade-off between coding efficiency and encoding complexity.

Table 1 Eight modes of angular predictions

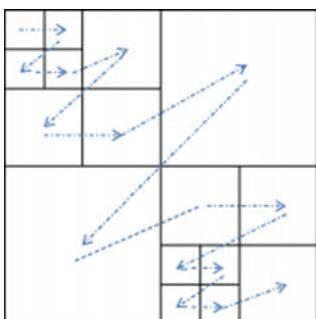
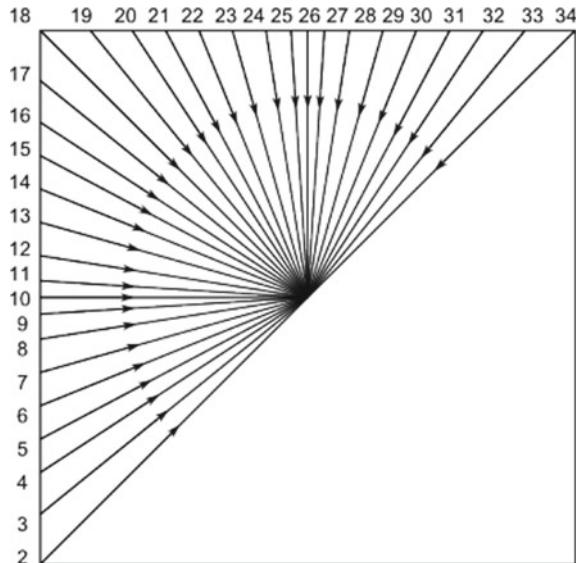
Mode 0 (vertical)	Mode 1 (horizontal)	Mode 2 (DC)
<ul style="list-style-type: none"> • a, e, i, m = A • b, f, j, n = B • c, g, k, o = C • d, h, l, p = D 	<ul style="list-style-type: none"> • a, b, c, d, e = I • e, f, g, h = J • i, j, k, l = K • M, n, o, p = L 	<ul style="list-style-type: none"> • All the predicted pixels are equal to mean value of all references(A, B...L)
Mode 3(diagonal down-left)	Mode 4(diagonal down-right)	Mode 5(vertical-right)
<ul style="list-style-type: none"> • a = (A + 2B + C)/4 • b, e = (B + 2C + D + 2)/4 • c, f, i = (C + 2D + E + 2)/4 • d, g, j, m = (D + 2E + F + 2)/4 • h, k, n = (E + 2F + G + 2)/4 • l, o = (F + 2G + H + 2)/4 • p = (G + 3H + 2)/4 	<ul style="list-style-type: none"> • d = (B + 2C + D + 2)/4 • c, h = (A + 2B + C + 2)/4 • b, g, l = (Q + 2A + B + 2)/4 • a, f, k, p = (A + 2Q + I + 2)/4 • e, j, o = (Q + 2I + J + 2)/4 • i, n = (I + 2J + K + 2)/4 • m = (J + 2K + L + 2)/4 	<ul style="list-style-type: none"> • a, j = (E + A + I)/2 • b, k = (A + B + I)/2 • c, l = (B + C + I)/2 • d = (C + D + I)/2 • e, n = (I + 2E + A + 2)/4 • f, o = (E + 2A + B + 2)/4 • g, p = (A + 2B + C + 2)/4 • h = (B + 2C + D + 2)/4 • i = (E + 2I + J + 1)/4 • m = (I + 2J + K + 2)/4
Mode 6(horizontal-down)	Mode 7(vertical-left)	Mode 8(horizontal-up)
<ul style="list-style-type: none"> • a, g = (E + I + l)/2 • b, h = (I + 2E + A + 2)/4 • c = (E + 2A + B + 2)/4 • d = (A + 2B + C + 2)/4 • e, k = (I + J + l)/2 • f, l = (E + 2I + J + 2)/4 • o, i = (J + K + l)/2 • j, p = (I + 2J + K + 2)/4 • m = (K + L + l)/2 • n = (J + 2K + L + 2)/4 	<ul style="list-style-type: none"> • a = (A + B + l)/2 • b, i = (B + C + l)/2 • c, j = (C + D + l)/2 • d, k = (D + E + l)/2 • e = (A + 2B + C + 2)/4 • f, m = (B + 2C + D + 2)/4 • g, n = (C + 2D + E + 2)/4 • h, o = (D + 2E + F + 2)/4 • i = (E + F + l)/2 • p = (E + 2F + G + 2)/4 	<ul style="list-style-type: none"> • a = (I + J + l)/4 • b = (I + 2J + K + 2)/4 • c, e = (J + K + l)/2 • d, f = (J + 2K + L + 2)/4 • g, i = (K + L + l)/2 • h, j = (K + 2L + L + i)/4 • k, m, l, n, o, p = L

HEVC recommends use of high efficient and flexible blocks defined as coding tree unit (CTU), coding unit (CU), prediction unit (PU) and tree unit (TU). Encoding of a frame is carried out by first dividing the frame into CTUs that contains coding tree blocks for both luma and chroma components. Figure 4 presents one of the possibilities of partition process for a 64×64 CTU and 8×8 CU.

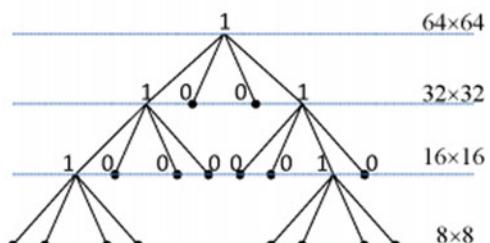
Flexible partitioning of CU is advantageous in HEVC coding. When there is homogeneity in the frame, larger CU can be used to represent the region in the frame using smaller number of symbols instead of using smaller block sizes. Flexibility in CTU sizes enables HEVC codec to be optimized for different content, application and devices [8]. Figure 5 presents the CTU sizes for different image resolutions as per HEVC standards [8].

In HEVC, the distinction between macro-block and sub-macro-block is eliminated and is expressed as CU only thus providing flexibility in formulating the multilevel hierarchical quadrature structure. In intra-frame coding with CTB partitioning with PU and TU, there are 3000 combinations, and all of these combinations need to be checked to choose optimal one which leads to complexity and encoding time. It

Fig. 3 Angular prediction modes in 33 directions



(a)



(b)

Fig. 4 Partition of frame into CTU and minimum CU

is required to speed up the process, and there are various schemes that have been reported in literature [9].

From statistical observations of natural images, it is observed that the horizontal and vertical edges or patterns occur more frequently than any other form of patterns. Prediction of these patterns is easier as the intra-pixel displacement both in vertical and horizontal directions helps in precise prediction. The patterns in the diagonal directions that occur very less frequent are difficult to predict as the displacement parameters gets larger. HEVC addresses these predictions accurately with 132 angular possibilities. Rate distortion (RDO) is a technique introduced in HEVC to achieve high efficiency in intra-frame prediction that helps in providing information

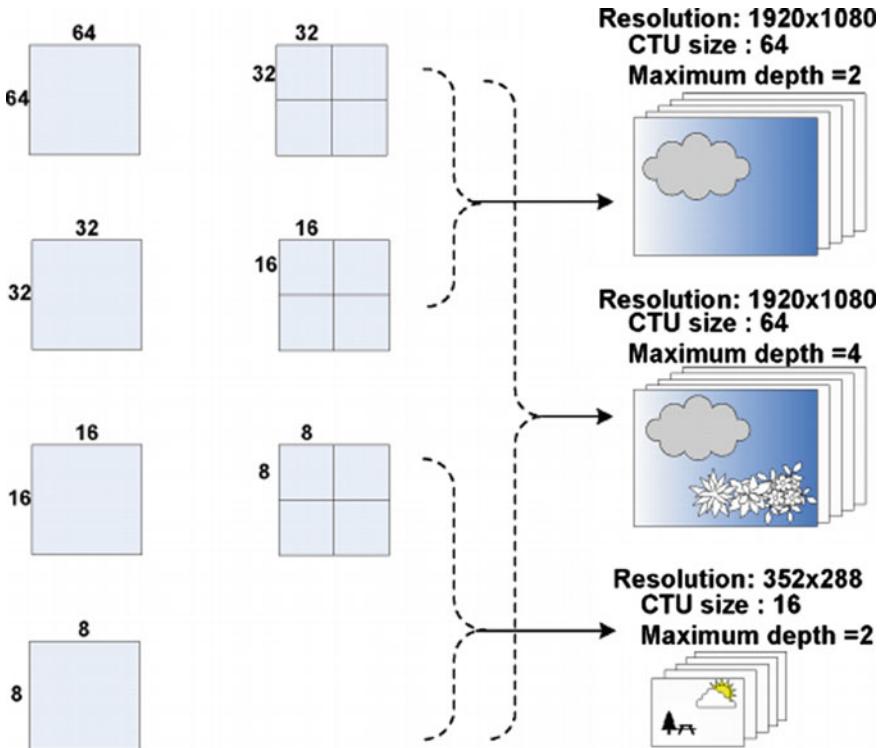


Fig. 5 CTU size with corresponding CU for different resolution images

on the appropriate macro-block size that need to be considered for prediction and the angular direction. RDO computation is also expensive due to large number of prediction operations. Damian Ruiz et al. have presented fast algorithm for intra-prediction based on discrete wavelet transform (DWT). Performing DWT on the input image and decomposing it to multi-resolution sub-bands information on homogeneity is computed to quickly identify optimum partition modes that are faster than the RDO technique. DWT sub-bands have limited directional patterns captured in sub-bands and are shift variant. Over the last few years, DWT limitations are addressed by use of dual-tree complex wavelet transform (DTCWT). DTCWT generates both real and imaginary sub-bands, with six directional features. The intra-prediction algorithm presented in this work uses DTCWT in place of DWT to compute optimum macro-block size and angular prediction modes.

3 Improved Partition-Prediction for Intra-Frame Encoding

Kingsbury introduced DTCWT to achieve shift invariance by increasing sampling rate by 2 in DWT at each level of both real and imaginary tree [6]. In DWT, down sampling by two introduces aliasing, which is addressed by having two trees (a and b) processing input data in parallel by four filters La, Ha and Lb, Hb such that there is one sample offset between these filter pairs. Two mother wavelets that are Hilbert transform pair are considered for each tree. Figure 6 presents level 1 DTCWT structure for input image X. The first stage is the row processing, and the second stage is column processing. The first stage generates four outputs denoted as y_1 , y_2 , y_3 and y_4 that are further processed to generate eight output complex sub-bands denoted as {LLR1, LLC1, LHR1, LHC1, HLR1, HLC1, HHR1 and HHCl}. There will be four low pass and twelve high pass real sub-bands. The low pass complex sub-bands LLR1 and LLC1 are combined into only two sub-bands considering the magnitude of the sub-bands. Similarly, the 12 high sub-bands are combined into 6 sub-bands

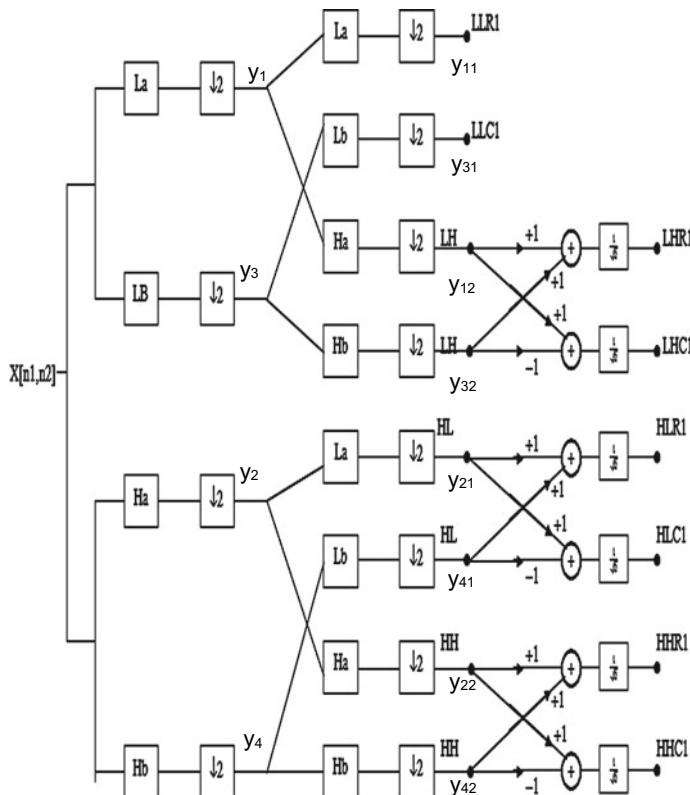


Fig. 6 Level 1 2D DTCWT structure

considering the magnitude of sub-bands. With tree level DTCWT decomposition, there will be two low pass bands and six high pass band generated every level.

3.1 Prediction Algorithm

Intra-frame prediction is demonstrated to achieve good performance if the partition is carried out considering 16×16 macro-block size. It is required to compute the homogeneity level of the macro-block, and an improved algorithm based on DTCWT is presented. The improved algorithm is of three steps:

Step1: The input frame is decomposed into complex wavelet sub-bands (μ) used considering Kingsbury 10-tap Q shift filter is for decomposition.

Step2: Decomposition is carried out for three levels and $\lambda = 3$ is set.

Step3: Homogeneity (HG) analysis is carried out considering $+ 64 \times 64$ macro-block (MB).

Step4: If there are homogeneous patterns, the partition is set to 64×64 and the prediction are carried out.

Step5: If not homogenous, then the homogeneity analysis considering 32×32 MB ($\lambda = 2$).

Step6: If there are homogeneous patterns, the partition is set to 32×32 and the prediction are carried out.

Step7: If not homogenous, the analysis considering 16×16 MB ($\lambda = 1$).

Step8: If there are homogeneous patterns, the partition is set to 16×16 and the carried prediction is carried out if not prediction is out for 8×8 .

The selection of sub-bands is also another parameter for cost function in intra-prediction coding algorithm proposed in this work. Homogeneity evaluation is carried out by computing directionality homogeneity index (DHI) obtained from each of the sub-bands that are obtained at level 2 and level 3. The HDI is a ratio of energy (C) in each sub-band to the variance (σ^2) of the sub-band as in Eq. (1),

$$\Gamma(\mu, \lambda) = C_{\mu\lambda}(k, I)/\sigma_{\mu\lambda}^2 \quad (1)$$

The DHI is compared with the threshold for each of the six sub-bands that are set using empirical methods considering both level 2 and level 3 sub-bands. In the three steps method during the second and third step, the DHI is compared with threshold level, and global homogeneity condition is considered if the DHI is less than threshold level.

The proposed prediction algorithm is modelled in MATLAB and is evaluated for its performances considering MSE and PSNR as evaluation metrics using Eqs. (2) and (3).

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \{[I(i, j) - K(i, j)]\}^2 \quad (2)$$

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (3)$$

4 Results and Discussion

Evaluation of partition-prediction algorithm proposed in this work is carried out by comparing the MSE and PSNR results with two other methods. The direct partition-prediction method and DWT-based partition-prediction method results are compared with the proposed results. Four different data sets are considered for evaluation, and the prediction is obtained for all 33 modes. Figure 7 presents the partition-prediction results for the proposed algorithm developed based on DTCWT considering all 34 modes. Figure 8 presents the original image and reconstructed image after partition-prediction and the corresponding residual image are presented in Fig. 9 for two different modes. Figure 10 presents the MSE performances of all three methods considered in this work. The MSE for the proposed method is the smallest compared with all three methods. PSNR comparisons are presented in Fig. 11, and the PSNR results for all 34 modes the proposed algorithm achieve highest PSNR Results.

Table 2 compares the intra-frame partition-prediction results for second set of image data. The proposed algorithm achieves lowest MSE and better PSNR compared with DWT-based methods. Table 2 results states that the proposed algorithm can attain higher compression rates at low modes. To evaluate the image reconstruction performance, the DWT and DTCWT filters for the selected bands are used to reconstruct the image by considering all the reconstruction features. The residual images for the two algorithms are presented in Fig. 12.

Figures 13 and 14 present the comparison of third and fourth image data set.

Figure 8 shows results compared with both DWT-based algorithm and DTCWT-based partition-prediction output; DTCWT is capable of reconstructing almost all the features, and also, the proposed partition-prediction algorithm is capable of retrieving all required features throughout decomposition and increases redundancies in the DTCWT coefficients, consequently enhancing the coding efficiency in HEVC. At the same time, it requires to design appropriate filters that satisfy for image reconstruction to improve PSNR.



Fig. 7 Partition-prediction results for 34 modes

5 Conclusion

HEVC standard can provide a significant amount of increased coding efficiency compared to previous standards, including H.264/MPEG-4 AVC. Over the last few years, DWT limitations are addressed by use of dual-tree complex wavelet transform



Fig. 8 Original image and reconstructed image

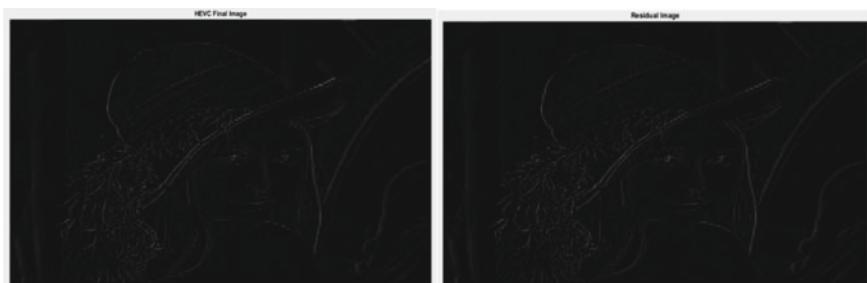


Fig. 9 Residual images for two different modes (mode 0 and mode 15)

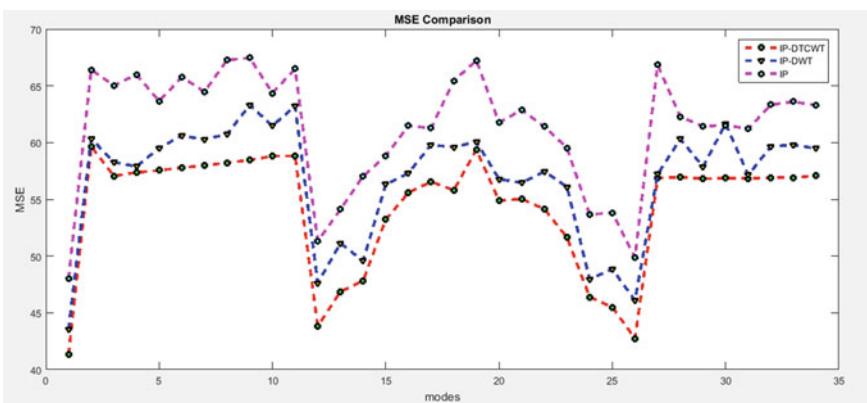


Fig. 10 MSE results comparison

(DTCWT). DTCWT generates both real and imaginary sub-bands, with six directional features. The intra-prediction algorithm presented in this work uses DTCWT in place of DWT to compute optimum macro-block size and angular prediction modes. Intra-frame prediction is demonstrated to achieve good performance if the partition

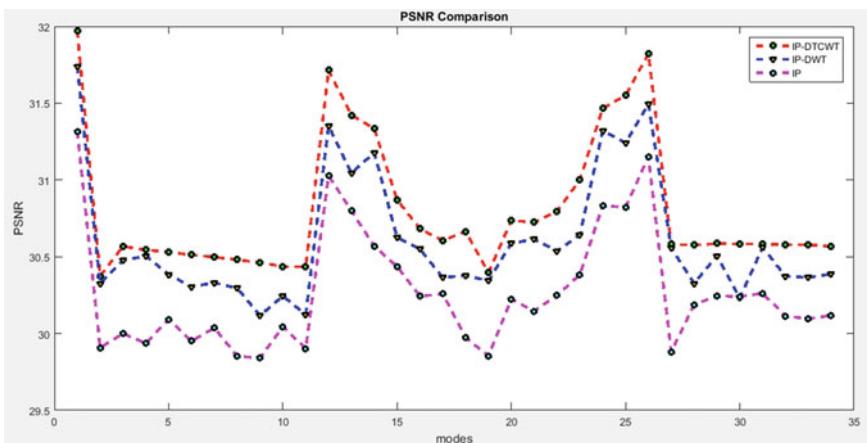


Fig. 11 PSNR results comparison

Table 2 Comparison of intra-frame prediction results

Modes	DWT-based intra-frame prediction		DTCWT-based intra-frame prediction	
	MSE	PSNR	MSE	PSNR
0	0.3988	51.1751	0.3963	52.1504
2	0.3981	51.1747	0.3963	52.1504
6	0.3892	50.7761	0.3884	52.2376
10	0.7219	49.1782	0.7100	49.6183
14	1.411	47.1800	1.2016	47.3331
20	1.734	45.0818	1.6759	45.8884
22	2.2911	44.1838	2.1839	44.7385
25	2.797	43.1857	2.6852	43.8410
28	3.4241	42.1888	3.2247	43.0459
30	3.896	41.9201	3.7596	42.3794
31	4.8521	41.1947	4.3247	41.7712
34	5.325	40.8977	4.9308	41.2017

is carried out considering 16×16 macro-block size. It is required to compute the homogeneity level of the macro-block, and an improved algorithm based on DTCWT is presented. Proposed scheme of fast DTCWT-based intra-prediction using decomposition methods. It is observed that the proposed algorithm can achieve greater compression in low modes. To evaluate the performance of image reconstruction with all features, the reconstruction is carried out with DWT and DTCWT filters for the selected bands. The proposed algorithm achieves the lowest MSE and better PSNR compared to DWT-based methods.

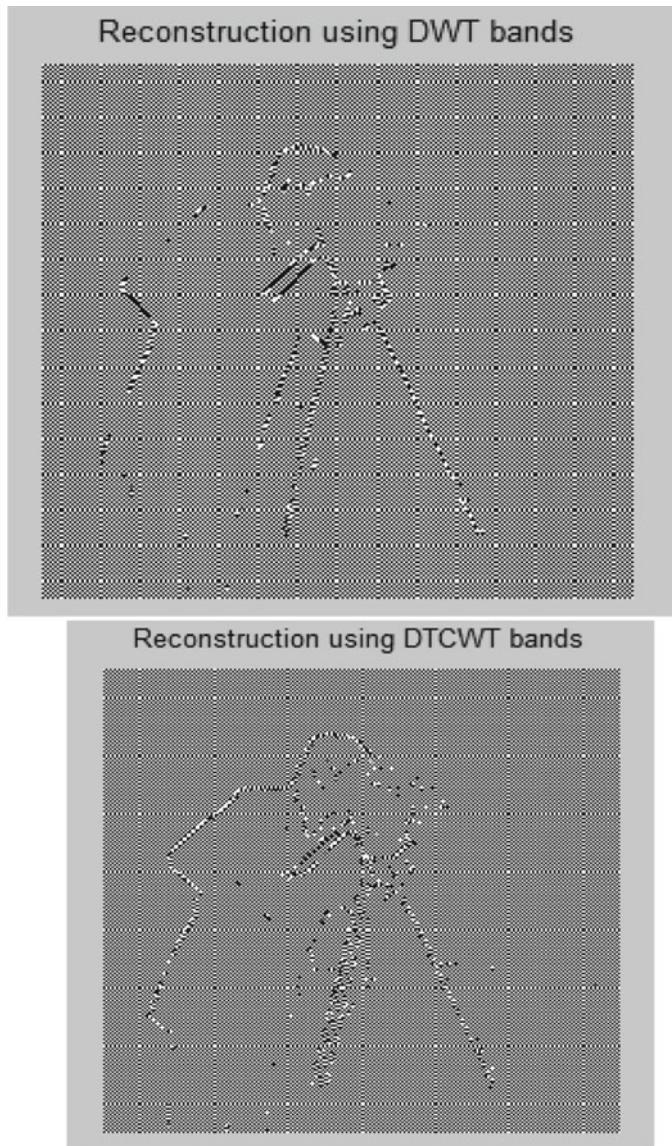


Fig. 12 Residual image comparisons for second data set

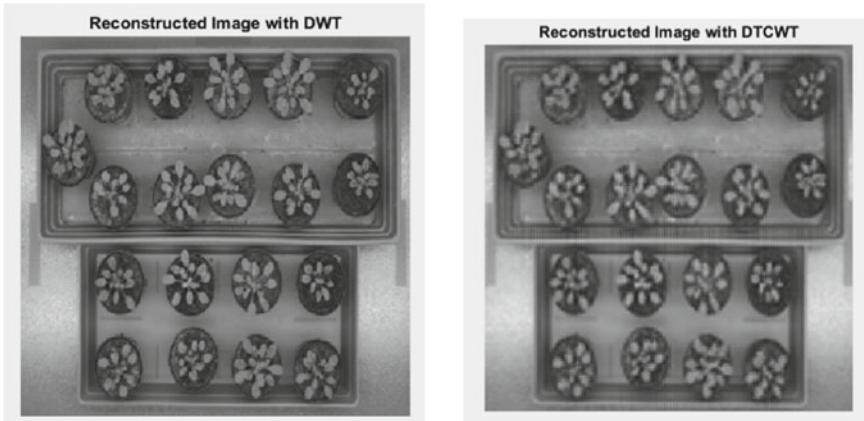


Fig. 13 Reconstructed images for third data set

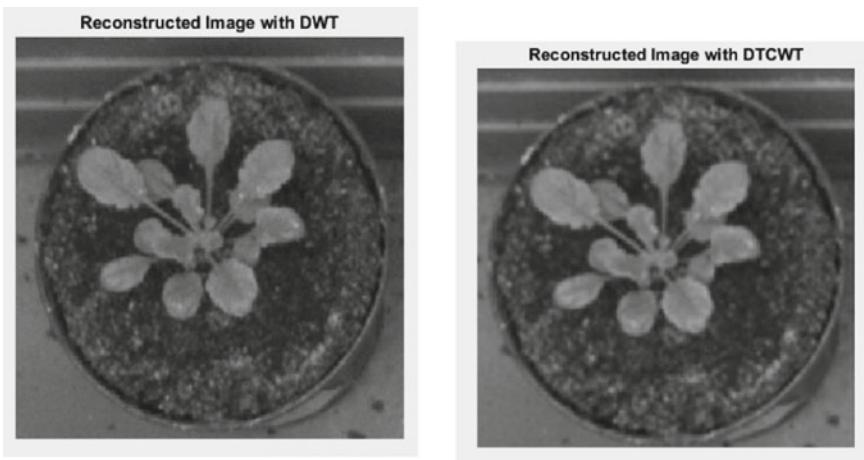


Fig. 14 Reconstructed images for fourth data set

References

1. Richardson IE (2013) HEVC: an introduction to high efficiency video coding. <http://www.vcodex.com/h265.html>
2. Kim IK, Min J, Lee T, Han W-J, Park J (2012) Block partitioning structure in the HEVC standard. *IEEE Trans Circuits Syst Video Technol* 22(12):1697–1706
3. Wang Y, Ostermann J, Zhang Y-Q (2002) Video processing and communications. Prentice Hall
4. Sullivan GJ, Ohm J-R, Han W-J, Wiegand T (Dec 2012) Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans Circ Syst Video Technol* 22(12)
5. Richardson IE (2010) The H.264 advanced video compression standard, 2nd edn. Wiley
6. Zhao J, Segall CA, Yiping S (13 Oct 2011) Methods and systems for intra prediction, United States patent application publication, Pub. No.: US 2011/0249741 A1

7. Lainema J, Bossen F, Han W-J, Min J, Ugur K (Dec 2012) Intra coding of the HEVC standard. *IEEE Trans Circ Syst Video Technol* 22(12)
8. McCann K, Han W-J, Kim I-K, Min JH, Alshina E, Alshin A, Lee T, Chen J, Seregin V, Lee S, Hong YM, Cheon MS, Shlyakhov N (Apr 2010) Samsung's response to the call for proposals on video compression technology, JCT-VC document A124, 1st JCT-VC meeting
9. Xiong J, Li H, Wu Q, Meng F (2014) A fast HEVC inter CU selection method based on pyramid motion divergence. *IEEE Trans Multimedia* 16(2):559–564

A Perspective Toward 6G Connecting Technology



Neha Katiyar, Jyoti Srivastava, and Kushall Pal Singh

Abstract 5G communication is the most trending technology and nowadays commercialized to the whole world. Still, now is the time to look forward beyond this technology that could be more forward than 6G. It is the improved technology for future applications and also for emerging space and terrestrial equipment. It is a unique opportunity for people to progress in various sectors like education, health, environment, business, and many more. The 6G is the sixth generation of mobile and telecommunication systems. The six genesis flagship program is the large-scale set up of eight-year research initiatives that implement the test and critical enabling technology for 6G. In this paper, the authors reviewed the technology advancement in the 6G network. This study shows the comparative analysis of efficient, cost-effective, and specific and aggregate efforts toward breakthrough innovations.

Keywords 6G · 5G · Mobile technology connectivity · Quantum technology · WCDMA

1 Introduction

Research and innovation is an intelligent method that should strongly emerge in recent years and has a strong perspective in the advancement of classical computing method and provides tremendous response quantum computing related to quantum technology. Quantum-related computing method realizes that they need an intelligent communication network for the next generation beyond the 5G network is 6G. It is an emerging network that is increasing the interconnectivity between humans and machines. The 6G networks contain all the features of the network, which 5G

N. Katiyar (✉)

Indian Institute of Management, Rohtak, India

J. Srivastava

Madan Mohan Malviya University of Technology, Gorakhpur, India

K. P. Singh

Malaviya National Institute of Technology, Jaipur, India

have like (EMBB) enhanced mobile broadband and (MMTC) massive machine-type communications or (ULLC) ultra-reliable low latency communications. The 6G technology is the most remarkable improvement in comparison to 5G technology. It has the feature of fast multi-mode adaption, flexible utilization of multiple bands, and easy adaption of network architecture. It increases the performance and maximizes the data throughput or TOPS. This technology provides security and protects the system. A chain technology mechanism has been used to meet the objectives of 6G. This technology has ultra-low latency and superfast broadband connectivity on the wireless systems, which is helpful in IoT applications. Technological advancement includes the ubiquitous sensing techniques for the intelligent cyber-system physically installed in smart societies.

The main contribution of this paper is to review recent state-of-the-art techniques in 6G communication and their applications. To achieve this, the authors have exhaustively surveyed the recent three-year work related to limited scope with popular applications of 6G.

2 Why is 6G Faster?

The industrial revolution 4.0 was started by using 5G. The digital platform, digital application, automation of home and organization, and cyber-computation ethics always maintain a technology needed that could be faster than the beyond all this application is 6G.

In the 1980s, a communication network was established that should be used for analog communication. It provides services with a data rate of 2.4 kbps. It is used to transmit information through analog signals but has lots of congestion in the communication network with no security system. It is also called IG. It used the technology at that time as frequency division multiple access (FDMA). After some advancement in the network, the second-generation network arrives at 2G. The communication network enables two technology within it here: TDMA and CDMA. The most important feature of this technology is SMS which is also used in short message services nowadays. This technology provides a boom in the communication sector, and the use of cell phones and communication devices increases. The third generation of the mobile network is much faster than the 2G network. It enables the technology (WCDMA) and CDMA and having a data rate up to 2 to 100 Mbps. Due to this 3G network, the boom that the 2G network should take is rapidly increased and should define the technical work should be continued with the help of the mobile system. After then 4G has evolved, which is the handiest technology, in terms of cost-effectiveness as compared to other communication technologies. Many software applications and platforms are deployed with the help of 4G, the main reason being cost-effectiveness. It gave birth to the smart environment and smart people of society. It has a data rate up to 20 Gbps. The 5G is technology is the advancements of all the technology. It can have a data rate of up to 20 Gbps, compatible, reliable,

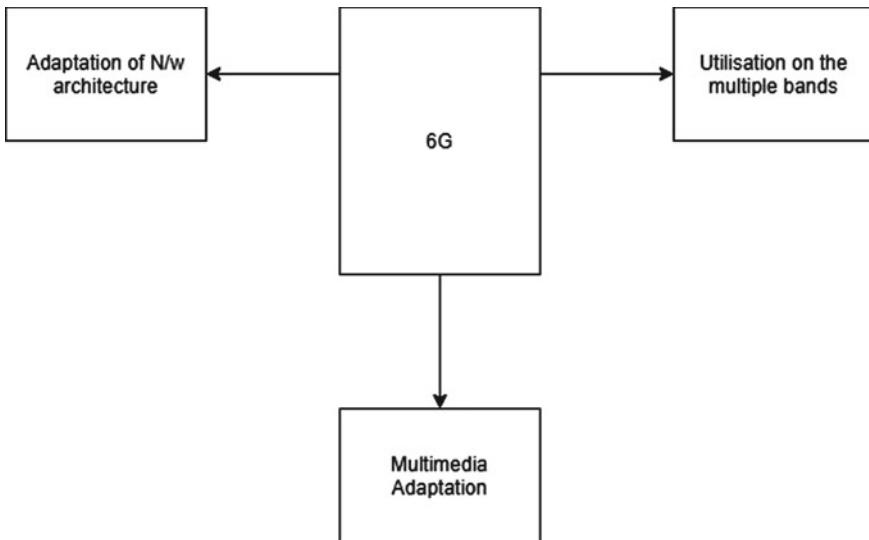


Fig. 1 6G features

and is secured for operating any intelligent system. It can enable the multiple input or multiple output (MIMO) techniques massive machine-type communication.

The 6G sixth-generation network technology is the most modern technology. The researchers and innovators have researched this network. The six genesis flagship program is mainly developed to test all enabling technology on 6G. This technology data rate measured 1 tbps (terra bits per seconds). It is the revolution of mobile technology changement.

It increases the performance and maximizes the data throughput or JOPS. It is also secured technology and provides protection to the systems. It is used in content-driven routing and blockchain technology. Its ultra-low latency and superfast broadband connectivity on the wireless system have increased the performance and maximized the data throughput. It also includes the ubiquitous sensing techniques for the intelligent cyber-system, which is physically installed in smart cities—some features of 6g, as shown in Fig. 1.

3 Related Work

In this modern era, 6G is the superfast connecting technology in this world. Many researchers doing research on this technology to make it eco-friendly for people, and users can use it in an efficient manner; some of the research work and the network area of 6G is discussed in Table 1; this table has the complete description of research

Table 1 Description of work in 6G technology

S. No.	Table column head	Author	Area	Year	Description
1	Macro Giordani		Mobile communications and network	2020	It focuses on a full-stack system-level perspective on 6G networks. It introduced the new paradigm of 6G networks. It focused on six network features like novel disruptive communication technology and its innovative network architectures
2	Tonggy Huang		Green 6G network	2019	It focused on the evaluation of the 6G network. It characterizes 6G network. It focused on that how the 6G network is useful for AI potential technologies. It integrates 6G network blockchain technique for better enhancement in future. It focused how the network evaluate from 1 to 3G, 3G to 4G, 4G to 5G, and 5G to 6G
3	Faisal Tariq		Haptic communication	2019	It focused on emerging technology 6G. It provides comparison between 5 and 6G technology. It provides individual data rate of 6G is 100 Gbps. It focuses on the vision of 6G network. It also provides information of Haptic communication and how it is helpful for remote surgery and simulated realities

(continued)

Table 1 (continued)

S. No.	Table column head	Author	Area	Year	Description
4	Yimming Huo	Multi-functional technology		2019	It focused on that how 5G removes the barriers of network technology. The 5G collaborates with different technologies Global Navigation Satellite System (GNSS) massive MIMO, and it also collaborates with technology DPA-MIMO like distributed phased arrays-based MIMO. It focused on 5G network also its scenarios of classification and 5G and 6G-GHz communication
5	Zubaida Khan	Wireless innovative systems		2018	It focused on the wireless communication of the 6G network without any network interruption. 6G is the only technology that provides superfast—streaming without any buffering. It also focused on 6G cutting-edge features. It provides details about Wireless Innovative System for Dynamic Operating Mega-communication (WISDOM)

(continued)

Table 1 (continued)

S. No.	Table column head	Author	Area	Year	Description
6	Macoz Kutz		Genesis program for smart society	2018	It focused on the communication system with a mobile network of 6G. It focuses on a research program of 6G is 6G–Genisis Flagship program. For building the smart society and ecosystem. It focused on network densification with mobility infrastructure
7	Yang Lu		6G challenges	2020	6G explores the new communication techniques without any restriction. 6G having most or the connectivity feature, deep connectivity, intelligent connectivity, and ubiquitous connectivity. Due to the advanced feature of connectivity, it has multidimensional coverage connections. It focused on the current development of 6G technology. It focused on the sparse theory (compressed sensing). It also focused on the performance of 6G network. It focused on the potential care technologies

(continued)

Table 1 (continued)

S. No.	Table column head	Author	Area	Year	Description
8	Minghao Wang	Security areas of 6G network		2020	In this paper, they worked on the security of 6G network and its advancement feature. It focused on the critical areas of the 6G network, which done real-time intelligent edge, distributed artificial intelligence, intelligent radio system, and 3D intercoms. It focused on security so it could suggest molecular technology
9	Syed Junaid Navaz	Quantum machine learning		2019	Research is the data-driven adaptive method. It provides like quantum computing. In this method, the author set their targets in computing like enhanced mobile broadband and ultra-reliable low latency communication or massive machine-type communication

(continued)

Table 1 (continued)

S. No.	Table column head	Author	Area	Year	Description
10	Fengxiao Tang	6G vehicular network		2019	This paper provides a new concept that the 6G network provides vehicular connectivity using machine learning techniques. They created a new network and named it 6G vehicular network. 6G vehicular network, including the evaluation of intelligent radio. It focused on the network perspective with machine learning are: (1) Machine learning with vehicular networking (2) Machine learning for vehicular communication (3) Machine learning for vehicular security
11	Dias Yaacub	6G connectivity		2020	In this paper, they focused on 6G connectivity. It also focused on the 6G connectivity challenges. They served on the rural areas for 6G connectivity. The connectivity they defined on the 6G network is not connected, under-connected, connected, hyperconnected

(continued)

Table 1 (continued)

S. No.	Table column head	Author	Area	Year	Description
12	Charalampou Sergiou	Complex systems of 5G and 6G		2020	<p>In this paper, the authors have focused on the complex system and complex network. They reviewed the various complex network models. They encourage the collaboration of 5G and 6G networks. They provide the new network concepts—</p> <ul style="list-style-type: none"> (1) Dynamic topology (2) The frequencies (3) Traffic management <p>During the time of backhaul process. They observed the characteristics of small-world networks, random networks—the network used in the different disciplines</p>

(continued)

Table 1 (continued)

S. No.	Table column head	Author	Area	Year	Description
13	Shrongng Zhan	Price reduction		2020	<p>In this paper, they focused on when 6G is available for customers. It is simply called cost reduction technology. It made the 1000 time price reduction from a customs point of view. The 6G developed AI-assisted communications, which help people. It provides a customized network policy to the customer, so the price reduction takes place. The challenges faced by this technology are-</p> <ul style="list-style-type: none"> (1) Management of wireless big data (2) Portable and low latency algorithm (3) Hardware design
14	Harish Visvanathan	6G communication		2020	<p>In this paper, they focused on 6G connectivity. The 6G system technologies required new man-machine interfaces created multiple devices in unison. In this paper, they focused on the fundamental dimension and design of the 6G network. The 6G design consists of the spectrum, data, compute, spectrum efficiency, energy, space</p>

(continued)

Table 1 (continued)

S. No.	Table column head	Author	Area	Year	Description
15	Rajesh Gupta	Unnamed aerial vehicles connectivity with 6G network	2020	In this paper, they focused on the connectivity of drones with a 6G network. They concentrate on how UAV systems improved their performance with less cost. They highlight the UAV communication with 6G environment. They concentrate on communication system security	
16	Harsh Tataria	6G opportunities and design	2021	In this paper, they focused on three most important characteristics that linked to the next decade of lifestyle and societal changes, impacting the design and outlook with 6G networks, are: (1) High-fidelity holographic techniques adapted by society (2) Connectivity for all things like home automation vehicular networks. (3) Time engineered applications like smartwatches and smart devices connectivity	
17	Mohammed Najah Mahndi	6G Connectivity with IoT	2021	In this paper, they focused on the projected 6G system architecture. 6G is an autonomous network that connects with three trending techniques: IoT, machine learning, and energy. They also discussed potential technologies which are connected with the 6G architecture	

(continued)

Table 1 (continued)

S. No.	Table column head	Area	Year	Description
Author				
18	Edward. J. Oughton	6G Infrastructure with techno-economic features	2021	This paper focused on the 4G, and 5G strategies quantitatively perform in viably delivering universal broadband coverage. It also focused on the 6G standardization and enhancement. They also discuss the spectrum pricing policy

work done by researcher in 6G in which year. Few specifications of 6G technologies done by some authors are discussed below.

Giordan et al. [1] focused on the 5G and 6G technologies. It focuses on the wireless sensor networks, provides a full-stack on the system-level perspective on 6G, and provides a completely new status to communication. In this, Macro Giordan also discussed and provided particular envisioned on the novel approach of novel disruptive communication technology and gave about the innovative architecture of network also integer the intelligence of 6G network.

Huang et al. [2] have received several 6G network architecture techniques. It mainly focused on the green 6G network and how the 6G green architecture can symbolize the ubiquitous 3D coverage and introduction of the general AI on the advanced network. It is also focused on the 6G communication factor-like terrestrial communication space communication, and also, the running projects on the 6G network are the smart city, and distributed AI Networks.

Taria et al. [3] have provided the vision about the 6G network. It will deliver on the theoretical study on the 6G network and provides a technological extension of the 5G network. 6G network gives the plethora of autonomous services for tablets and mobile phones. The smart city connectivity also depends upon the 6G network. It also focused on the quantum communication of the network.

Huoet et al. [4] propose hardware that gives a multiplexing solution on the energy-efficient design of user equipment solutions. It also highlights the massive MIMO phased array with multiple access techniques used in the 5G wireless technology. 5G enables the many applications and applications scenarios along with cellular services. It also focused on the 5G most crucial feature, Vo Wi-Fi voice over Wi-Fi, and it will be the more viable and trending product nowadays.

Kalbande et al. [5] is mainly focused on the 6G communication network setup and utilization phase initialization phase and its full-duplex methodology of radio wave transmission for communication in India. Implementing a 6G network that can support the various developing techniques gets the high-speed data in a relatively low nanosecond time. It discussed the critical enabling feature of 6G technologies like cutting-edge technology and WISDOM stands for wireless innovation system for dynamic operating mega communication and second transmission remote observed of neon antennas.

Katz et al. [6] provide a new technology enhancement and improvement in the 6G technology. It discussed the six genesis flagship program of the 6G network.

Lu and Zheng [7] reviewed the 6G technology and focused on the connectivity feature of 6G, which includes intelligent connectivity, deep connectivity, and holographic connectivity. It shows the tactile Internet connectivity seamlessly. It provides information regarding the development of 6G in countries like the USA, China, South Korea, Japan, UK, and Finland. It gives the information of 6G dimensionality and working aspects on applications.

Wang et al. [8] have reviewed the 6G network as well as future investigations. It focuses on the quality of the 6G network, its security, and privacy. It shows the main key aspects of 6G, just like edge computing, intelligent ratio, distributed artificial intelligence, and blockchain technology. A unique feature of 6G technology

is molecular communications, the natural phenomenon of nanoscale structures with living entities. Quantum communication technology provides a great potential to 6G network. It can provide absolute security and the right communication technology for long-distance communication.

Nawaz et al. [9] focused on machine learning for 6G communications. The 6G network is performed beyond connectivity. 6G is a massive, rapidly connected network that can proceed to the users to use an efficient way. This paper concatenated machine learning and quantum computing and produced a new technique quantum machine learning (QML)-based network on the 6G communication. It gives an overview of the service of QML. It enhances mobile broadband. It provides ultra-low latency communication (ULLC). It can be done communications in massive machine type.

Tharde et al. [10] have focused on the communication and challenges and its applications of 6G. It shows that how 6G worked on the hyperaccurate position or location. It focused on the complex networks for the communication perspective, focusing on the 5G and 6G mobile communication networks. The collaboration between the 5G and 6G networks works on complex systems and behaviors. Complex systems are increasing rapidly due to the different domains of data. The challenges in 6G mobile networks are introduced just like dynamic topology THz frequencies. The usage of 6G networks by artificial intelligence (AI) and machine learning (ML) blockchain technology has achieved more significant improvement in data communication.

Tang et al. [11] were mainly focused on the 6G network connectivity with AI and machine learning techniques. In the future, 6G vehicular network enabling AI toward the end. The 6G vehicular network has revolutionized the domain of intelligent ratio [IR], network intelligence, and self-indexing with fast communications. The multi-radio is the main challenge of the 6G network with AI, vehicular communications, vehicular networking, and network security.

Yacoub et al. [12] have mainly focused on the deployment of communication networks in rural areas. It surveys the connectivity of the rural area with specific countries. It works on the rural connectivity problem of the networks. It reaches the point that a particular region has what kind of problem of connectivity. It can focus on all the connectivity levels on the 2G and 3G networks like not connected, under-connected, bonded, and hyperconnected.

Zhang et al. [13] have mainly focused on the reduction methods with 6G networks. 6G in current time can remove the different technical uncertainties and efforts toward the innovation of 6G. It can focus on the features of the 6G network and told about that why the 6G network had done the price reduction ten times than the 2G network. The 6G network has the following features. It has a peak throughput of 1Tb/s. It is having low latency. 6G has a reliability of 99.9%. The connection density increased 1000 times. The mobility increased further enhanced from 500 km/h to 1000 Km/h.

Vishvanathem et al. [14] focused on the requirements of the 6G network in the timeframe. It provides some area specifications, where a 6G network is needed. It focused on the machine interfaces created on the local devices acting in the union. It focused on ubiquitous computing, multi-sensory data.

Gupta et al. [15] focused on the cryptographic-based solution of the 6G network. They also suggested a new model of a 6G network based on the blockchain, reducing network latency and bandwidth issues. The main aim of this technique is to ensure data security and privacy, lower prices, and provide enhanced connectivity. This paper focused on the unnamed aerial vehicle, which is commonly said the drone, its connectivity with 6G network. UAV is mainly designed for rescue operations, urban planning, weather monitoring, and precision agriculture.

Tataria et al. [16] focused on the 6G network application, which provides open system interconnection of the stack with 6G applications. These applications required an order of magnitude and more spectrum. The 6G network has more frequency bands. 6G is the following generation network with core networks coding techniques, new modulation formation real-time techniques, and multiple access methods. They also focused on the enhancement of society with high-fidelity holographic systems.

Mahdi et al. [17] focused on the connectivity of a 6G network with three highly demanded technology: Internet of things, energy, and machine learning. They have done critical analysis on the 5G network as well as the 6G network. The 6G network is a self-sufficient network that covers all insights. It is a self-healing and protective network.

Oughton et al. [18] are mainly focused on cellular technologies like 5G, 6G. They performed a quantitative assessment on 5G policies and showed how 5G policy supports the 6G policy. The 6G network provides the quality of service. 6G network having the superior foundation for the upcoming techno-economic environment. 6G cost-effective approaches used to deliver the network connectivity to broad geographic areas. The connection is between the physical world to the actual world. It focused on the holographic telepresence for both work and social relationships.

6G is the latest communication technology; in the literature, the summarization of all existing work has not been explained adequately. In this paper, the authors have reviewed the latest recent three-year relative research work, which contributes significantly to 6G and its applications in different domains, like 6G used in IoT, UAV, smart cities, and many more. As per the authors' best knowledge, previously published research reviews have concentrations on 6G technology compared with 5G or previous communication technologies.

4 Contribution of 6G Toward the Smart Society and Ecosystem

Smart society can be formed with the help of 6G. Millions of sensors should be installed in vehicles, roads, houses, buildings, and all those sensors have high-speed connectivity and integrate with human activities or human wearable devices. So the tasks are performed by humans in a smarter way and improved living standards. This is the way when 6G contributed by maintaining the connectivity. When an intelligent society is formed, that is also having the responsibility of preserving the ecosystems

make it proper applications and improvement in biotechnology, bioinformatics. The ecosystem should be preserve when measuring the quality of soil its humidity index; with the help of smart devices, these smart devices could send the relevant information to the preservers for 6G connectivity. Wireless brain interaction with devices should be done quickly to reduce the transportation of humans, which should be better for the ecosystem. 6G makes the ecosystem connections whether you lie any part of the ecosystem like oceans, mountains and forests. It provides autonomous edge-based connectivity to all the systems.

5 6G Collaboration with Machine Learning Techniques

6G will create drastic changes in the healthcare sector. It breaks the time and space barriers in the healthcare sector. It used the quality of its features of service means QoS form its network intelligence. It may collaborate with machine learning algorithms to provide better results in the healthcare sector. It may reduce the time-space complexity. It is used with unsupervised machine learning algorithms. Many machine learning algorithms are used for feature extraction and feature detection. It may extract the features from images, and 6G intelligence capabilities in the network send those extracted features with high speed. It can make the feature extraction process very fast. Due to that, holographic connectivity treatment could be done inefficient way in a short span of time. It can concatenate with the machine learning algorithm KNN, which detects similar neighbors at a certain metric. It performs this process in a very highly efficient way. It could perform pattern recognition. It uses in the classification of glucose patterns when 6G collaborates with deep learning techniques like ANN that makes connections with human brain signals and provides results, the MRI and CT scan systems, and become modernized and efficient regarding time reduction. When 6G technology collaborated with reinforcement learning, it provides better results according to the performance of the system. 6G concatenates with machine learning; it provides decisions automatically for any implementation regarding machine learning or any other technology.

In Fig. 2, the proposed methodology combines machine learning algorithm convolutional neural network (CNN) with 6G. This technique takes the input data from real-time scenarios, this observed data is passed through different layers, and the feature map is extracted as an output of a particular layer. This feature map/representation is used for classification/prediction. The data is collected in real-time and used for classification. The result will be sent to various applications by superfast connecting technology 6G to outperform any digital, physical, or biological world.

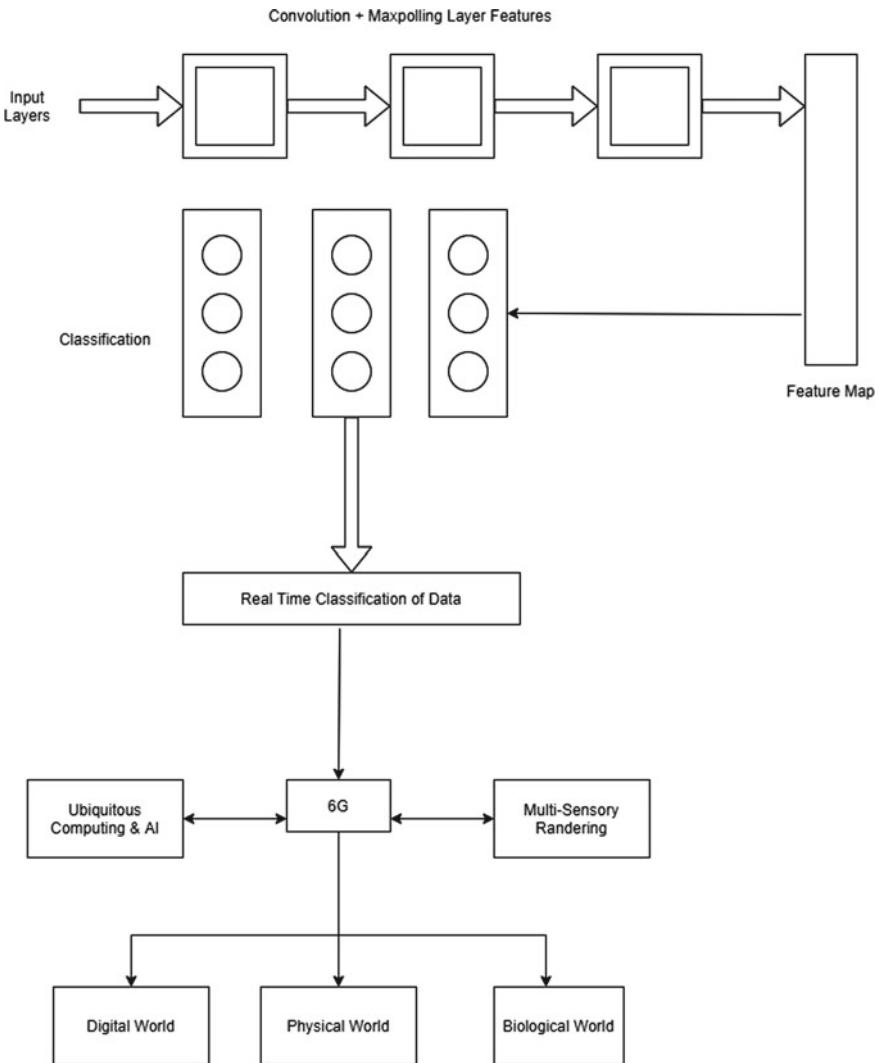


Fig. 2 Proposed methodology of 6G working with machine learning

6 Conclusion

The demand for high-speed data is increasing day by day due to the rapid proliferation of all kinds of smart devices, leading us to the era of the 6G network. 6G network is a sustainable and promising network. It includes QoS features. It may dramatically improve the data transmission policy and become a potential provider of the seamless network society. This paper may serve as an enlightening guide for future research work in 6G communications.

In the future, 6G has a wide realistic perspective considering the communicational part of it and the other fields like defense, aerospace engineering, ethical hacking, electronics, electrochemical science, and artificial intelligence.

References

1. Giordani M, Polese M, Mezzavilla M, Rangan S, Zorzi M (March 2019) Towards 6G networks: use cases and technologies. ArXiv 55–61
2. Huang T, Yang W, Wu J, Ma J, Zhang X, Zhang D (2019) A survey on green 6G network: architecture and technologies. IEEE Access 7:175758–175768. <https://doi.org/10.1109/ACCESS.2019.2957648>
3. Tariq F, Khandaker MRA, Wong KK, Imran M, Bennis M, Debbah M (Feb 2019) A speculative study on 6G. ArXiv
4. Huo Y, Dong X, Xu W, Yuen M (2019) Enabling multi-functional 5G and beyond user equipment: a survey and tutorial. IEEE Access 7:116975–117008. <https://doi.org/10.1109/ACCESS.2019.293629>
5. Kalbande D, Khan Z, Haji S, Haji R (2019) 6G-next gen mobile wireless communication approach. Proceedings of the 3rd international conference on electronics and communication and aerospace technology, ICECA 2019, 1–6. <https://doi.org/10.1109/ICECA.2019.8821934>
6. Katz M, Matinmikko-blue M, Latva-aho M (2019) 2018 IEEE Latin-American conference on communications. 2018 IEEE 10th Latin-American conference on communications (LATINCOM). pp 1–1. <https://doi.org/10.1109/latincom.2018.8613237>
7. Lu Y, Zheng X (2020) 6G: A survey on technologies, scenarios, challenges, and the related issues. J Ind Inf Integr 19(July):100158. <https://doi.org/10.1016/j.jii.2020.100158>
8. Wang M, Zhu T, Zhang T, Zhang J, Yu S, Zhou W (2020) Security and privacy in 6G networks: new areas and new challenges. Digital Commun Networks 6(3):281–291. <https://doi.org/10.1016/j.dcan.2020.07.003>
9. Nawaz SJ, Sharma SK, Wyne S, Patwary MN, Asaduzzaman M (2019) Quantum machine learning for 6G communication networks: state-of-the-art and vision for the future. IEEE Access 7(c):46317–46350. <https://doi.org/10.1109/ACCESS.2019.2909490>
10. Sergiou C, Lestas M, Antoniou P, Liaskos C, Pitsillides A (2020) Complex systems: a communication networks perspective towards 6G. IEEE Access 8:89007–89030. <https://doi.org/10.1109/ACCESS.2020.2993527>
11. Tang F, Kawamoto Y, Kato N, Liu J (2020) Future intelligent and secure vehicular network toward 6G: machine-learning approaches. Proc IEEE 108(2):292–307. <https://doi.org/10.1109/JPROC.2019.2954595>
12. Yaacoub E, Alouini MS (2019) A key 6G challenge and opportunity—connecting the remaining 4 billions: a survey on rural connectivity. ArXiv 108(4). <https://doi.org/10.36227/techrxiv.1025336>
13. Zhang S, Xiang C, Xu S (2020) 6G: connecting everything by 1000 times price reduction. ArXiv 1(March):107–115. <https://doi.org/10.1109/ojvt.2020.2980003>
14. Viswanathan H, Mogensen PE (2020) Communications in the 6G Era. IEEE Access 8:57063–57074. <https://doi.org/10.1109/ACCESS.2020.29817>
15. Rajesh G, et al (2021) Blockchain-assisted secure UAV communication in 6G environment: architecture, opportunities, and challenges. IET communications
16. Harsh T, et al (2021) 6G wireless systems: vision, requirements, challenges, insights, and opportunities. Proceedings of the IEEE
17. Mohammed Najah M, et al. (2021) From 5G to 6G technology: meets energy, internet-of-things and machine learning: a survey. Appl Sci 11:17:8117

18. Oughton EJ, Jha A (2021) Supportive 5G infrastructure policies are essential for universal 6G: assessment using an open-source techno-economic simulation model utilizing remote sensing. IEEE Access 9:101924–101945

MRI Breast Tumor Extraction Using Possibilistic C Means and Classification Using Convolutional Neural Network



R. Sumathi and V. Vasudevan

Abstract Breast cancer is the most leading cancer disease which demolishes many women lives for the past few decades. It can be prevented and reduce the death rate by proper treatment and by diagnosis at early stage. Mammogram and MRI scanning are preferable for analyzing the internal functionality of breast in detail, and in our study, we used MRI images with various sequences like T1-W, T2-W and T1-contrast enhanced for validation. Our approach uses gaussian filters for preprocessing, probabilistic C means for tumor segmentation and convolutional neural network for tumor classification. Our approach yields 98% segmentation accuracy and 95% classification accuracy, and its performance is compared with the existing methods like FCM, FCM-VES, hybrid K-means for segmentation and SVM, KNN and HOS-SVM for classification. We apply various qualitative measures such as entropy, eccentricity, dice coefficient, MSE, PSNR, F-score, sensitivity, specificity and accuracy for validation. For our work, we utilized online dataset like BI RADS and clinical dataset and proved that our approach needs average of 6 s for processing the breast cancer images.

Keywords MRI imaging · Fuzzy contrast-limited adaptive histogram equalization · Possibilistic C means · Convolutional neural network · Performance metrics

1 Introduction

MRI imaging is the most powerful scanning which briefs the anatomy structure of breast tissue, tumor in detail which aids the physician to make decision for early diagnosis. Mammogram screening is used to capture the breast lesions, but for dense

R. Sumathi (✉)

Department of Computer Science and Engineering, School of Computing, Kalasalingam Academy of Research and Education, Krishnankovil, Tamil Nadu, India

V. Vasudevan

School of Computing, Kalasalingam Academy of Research and Education, Krishnankovil, Tamil Nadu, India

breast, it is not preferable, so MRI screening is used to study the growth of tumor and lesion in breast. Clustering is used to link each pixel in image with a class label and grouped together with same class labels. Possibilistic C means clustering is used to reduce insensitive noise and able to identify the outlier data efficiently. Due to its high sensitiveness in initialization process, it generates coincident clusters easily. Classification is used to understand the basic domain and to improve prediction of huge dataset in medical applications. To ensure the efficient classification for large datasets, many classification techniques were used [1]. Various classification techniques are used in medical imaging; among them, CNN plays a vital role in classifying the tumor part as benign and malignant. It automatically detects the various features in the given image without human intervention. Its convolution and pooling operations play a major role in classifying the objects with efficient manner. Our proposed approach focused on PCM for brain tumor part extraction and CNN for classification with various images sequences like T1-contrast enhanced, T1-W and T2-W images. To handle the missing pixels, PCM is suitable with needful parameters for detecting the tumor part and its features are fed as input for CNN for classifying the tumor and ensured its performance is far better than SVM, KNN, HOSSVM techniques.

2 Related Works

Integration of BDR-CNN-GCN technique is used to detect the malignant in mammogram breast cancers images with 8 layer for CNN and yields 96.1% accuracy with MIAS dataset and its performance is compared with various techniques like SVM, IBBO, WEE and RSPNN [2]. In this approach, BRD and CNN are used for image level feature extraction, and GCN aids in training the model. Limitation is mammogram imaging is suitable for less dense breast cancer detection. Reference [3] focused on pathological images of breast cancer with three steps; in the first step, K-means with auto-encoder is used for clustering, and enhanced loss function is applied for classification and is performed in the second step; performance comparison is done with various existing approaches and ensured its classification accuracy is 95% which is superior state-of-art techniques and is done in the last step. They proved that the integration of deep learning, K -means and auto-encoder with enhanced loss function is suitable for classifying breast histopathological image. Data mining concept like decision tree classification is integrated with enabled class learning approach [4] for detecting and categorizing the breast cancer as benign and malignant. With the combination of random forest and tree voting analysis model, breast tumor is classified with online data set (WBCD) and proves its classification accuracy is superior than CART model [5]. To overcome the missing data and non-normal distribution, Naïve Bayesian classifier is used for reducing the dimension and utilized Hoeffding tree for classification and ensured its classification accuracy is 96%, whereas KNN, CDA and J48 obtained less than proposed approach, WBC online dataset is used for

validating the approach [6]. Integration of genetic programming with DGP framework is utilized for classification of breast cancer tumor as benign and malignant with UCI ML dataset and ensured its accuracy is superior than BPNN, Koza & Rice, Ave and GONN classifiers. Combination of CNN and data augmentation [7] based on pixel intensity framework is designed to classify the MRI breast cancer images and produces 98.3% classification accuracy. With 200 breast images collected from clinical and BI RADS for validation and proved that with the limited dataset the CNN model is capable for classifying the various size of masses. Various classification techniques [8] are analysed in terms of classification accuracy with various online datasets and ensured that CNN is superior that deep learning technique. In terms of performance, they categorized as quantitative, qualitative measures which ensure the best model for MRI breast classification. Reference [9] suggests the combination of morphological and dynamic features approach for categorizing the MRI DCE breast images as normal, benign and malignant. They utilized 26 breast cancer images and extracted 54 morphological features and 98 dynamic features for training and yield 91.7% classification accuracy and proved its accuracy is superior to manual segmentation. To classify breast tissue, deep convolution neural network [11] is used for classification and obtained 96% accuracy using BI-RADS dataset. Various artificial intelligence (AI)-based classification [12] of MRI breast images are done and yield less false negative value than other machine learning techniques [13]. The major contribution of this approach is to assimilate PCM with CNN for detection and classification with various MRI breast image sequences with fewer spans of time.

3 Methods and Materials

3.1 *MRI Breast Dataset*

For our validation, we used MIAS, BI-RADS MRI breast images with various sequences like T1-W, T2-W, T1-contrast enhanced. We used 212 (160 Benign and 52 malignant) MRI breast images for ensuring the efficiency of proposed approach. For Implementation MATLAB 2019(b) is used with Intel I5 Processor with 4 Ghz processor and 8 GB RAM. The working principle for our proposed approach is defined in Fig. 1.

3.2 *Possibilistic C Means*

It partitions the clusters based on the possibilistic membership and yields powerful clustering result, and due to its noise ratio, many medical applications like brain tumor, breast cancer and skin cancer-related diseases preferred [14]. To overcome the constrained member ship in FCM, Possibilistic C means relaxes the constraint on

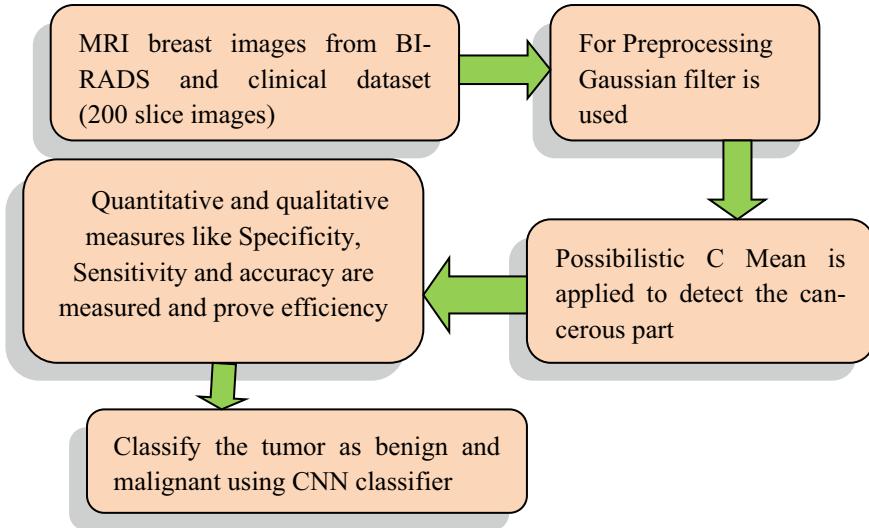


Fig. 1 Flow diagram for PCM–CNN approach

the membership of each data point with degree of possibility [15]. The degrees are defined to characterize the categories based on the features obtained from the clusters and help to categorize the members of clusters. It reposes the row sum constraints defined in FCM technique. Its main aim is to define each membership value lies between 0 and 1. Objective function if PCM defined as [10, 16].

$$J_m(U, V) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m \|x_k - v_i\|^2 + \sum_{i=1}^c n_i \sum_{j=1}^n (1 - u_{ij})^m \quad (1)$$

Steps for PCM processing are as follows:

- Step 1: Initialize the membership matrix U, and choose the values of m , c and ε .
 Step 2: Cluster center is updated by using the formula

$$v_i = \frac{\sum_{j=1}^n (u_{ij})^m x_j}{\sum_{j=1}^n (u_{ij})^m} \quad (2)$$

Step 3: Evaluate η_i using the following equation

$$\eta_i = \frac{\sum_{j=1}^n (u_{ij})^m (d_{ij})^2}{\sum_{j=1}^n (u_{ij})^m} \quad (3)$$

Step 4: Update the membership function of U using the following equation

$$u_{ij} = \frac{1}{1 + (d_{ij}/\eta_i)^{1/(m-1)}}. \quad (4)$$

Step 5: If $\|u_{ij} - u'_{ij}\|^2 \leq \varepsilon$, then stop the process else go to step 2.

Where n represents the number of patterns, c represents the number of clusters, d_{ik} represents the distance measures, m defines the fuzzy degree, and $U \{u_{ij}\}$ represents the fuzzy partition for the given image X [17]. Use the above steps to process the input image using PCM, and from the result, various features like entropy, correlation, eccentricity, variance, mean, homogeneity, contrast, kurtosis, and energy are fed as input for CNN model for classification.

3.3 Convolutional Neural Network

CNN is a member of deep neural network which analyzes the visual images in detail. For classifying medical dataset, CNN is preferred and achieves good results compared with KNN, SVM, etc. We used seven layers of CNN with nine features set with three convolution layers, three maxpooling layers and one fully connected. Output of first layer is fed as input for maxpooling layer, and its output is connected with fully connected layer to obtain the desired result. The necessary features for extraction from MRI breast images are contrast, entropy, eccentricity, mean, standard deviation, variance, homogeneity, kurtosis and smoothness.

3.4 Qualitative Measures

Dice similarity coefficient ensures the segmentation quality by measuring the degree of overlap between PCM segmented output and expert segmented output.

$$DSC(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

where X represents the ground truth images, and Y represents the proposed segmented output image. We can measure the level of quality based on the range (0, 1) and obtained 0.94 as DSC value ensured its segmentation quality.

3.5 Classification Accuracy

Sensitivity, specificity and accuracy ensure the quality of accurate detection and classification of breast tumor as benign, malignant and normal.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (7)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

whereas true positive (TP) represents the identification of correct tumor part, TN represents the identification of non-tumor part, FP represents the incorrect tumor part, and FN represents the incorrect detection of tumor part.

4 Results and Discussion

Our proposed method detects the tumor part using PCM with two classes of cluster, number of iteration is 90 for processing, and the obtained output is compared with K -means, FCM. From Fig. 2, it shows that PCM detects the abnormal part of breast images in accurate.

From Fig. 2, segmented output ensures that our proposed approach is suitable for extracting various sequences of MRI breast images. To detect the correct tumor part in MRI breast using PCM is compared with existing methods like FCM which yields 92%, FCM-VES yields 97% segmentation accuracy and our proposed method PCM

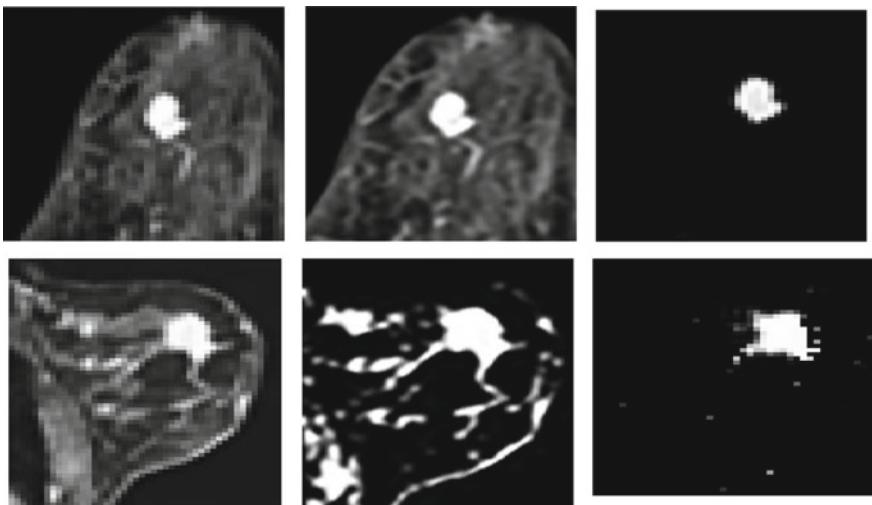


Fig. 2 **a** input image **b** Preprocessed image **c** PCM

yields 97.9% segmentation accuracy. Segmentation accuracy comparison [18, 19] is shown in Fig. 3.

Classification accuracy is ensured by its comparison classification accuracy with existing techniques like KNN, ANN, SVM, etc.

From Fig. 4, it was proved that classification accuracy of CNN is superior to existing techniques [20, 21]. From the obtained result, it was ensured that CNN classification with PCM segmentation is suitable for MRI breast tumor detection and classification. Table 1 shows the MSE and PSNR values obtained from various classification techniques like KNN, SVM and ANN; from this value, it observed that segmentation quality is ensued with PSNR average value is better than existing techniques. Processing time is more important for medical image processing, our proposed segmentation processing time is less than 4 minutes, whereas FCM processed in 11 s, *K*-means utilizes 7 s, and our PCM uses 6 s of duration for detection of tumor portion MRI breast images is shown in Fig. 5.

Fig. 3 Segmentation accuracy comparison

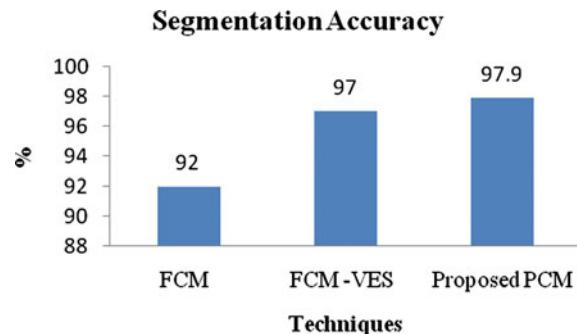


Fig. 4 Classification accuracy comparison

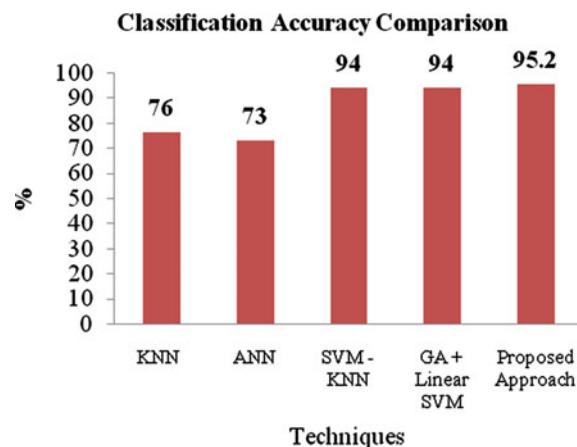
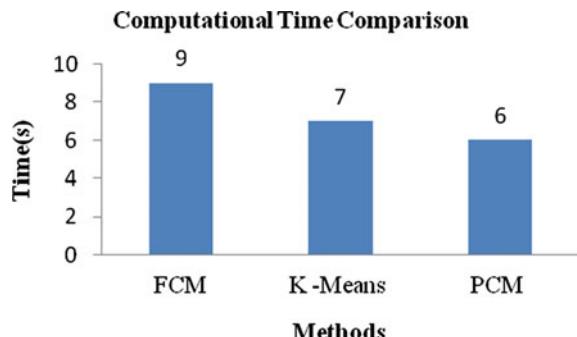


Table 1 MSE PSNR comparison

Technique	MSE	PSNR
FCM	1.22	45 decibel
FCM-VES	1.03	52 decibel
KPCM	0.94	56 decibel
Proposed	0.023	58 decibel

Fig. 5 Computational time comparison

5 Conclusion

Our proposed framework possibilistic C means with CNN approach is used to detect and classify the tumor in MRI breast sequences with less duration of time. Accuracy and efficiency are ensured with various qualitative measures and computational time. Segmentation quality is ensured with obtained MSE and PSNR values of various segmentation techniques. Forming the integration of PCM with CNN approach aids the radiologist for fast processing and reduces manual error. Our future work may focus on deep learning techniques for both classifications of breast images.

References

1. Filipczuk P, Kowal M, Obuchowicz A (2011) Automatic breast cancer diagnosis based on K-means clustering and adaptive thresholding hybrid segmentation. In: Choraś RS (eds) Image processing and communications challenges 3. Advances in intelligent and soft computing, vol 102
2. Zhang Y-D, Satapathy SC, Guttler DS, Górriz JM, Wang S-H (2021) Improved breast cancer classification through combining graph convolutional network and convolutional neural network. Inf Process Manage 58(2)
3. Acharya S, Alsadoon A, Prasad PWC (2020) Deep convolutional network for breast cancer classification: enhanced loss function. J Supercomputer 76:8548–8565
4. Ghiasi MM, Zendehboudi S (2021) Application of decision tree-based ensemble learning in the classification of breast cancer. Comput Biol Med 128

5. Alhayali RAI, Ahmed MA, Mohialden YM, Ali AH (2020) Efficient method for breast cancer classification based on ensemble hoeffding tree and naïve Bayes. *Indonesian J Electr Eng Comput Sci* 18(2):1074–1080
6. Devarriya D, Gulati C, Mansaramani V, Sakalle A, Bhardwaj A (2020) Unbalanced breast cancer data classification using novel fitness functions in genetic programming. *Expert Syst Appl* 140
7. Yurttakal AH, Erbay H, İkizceli T (2020) Detection of breast cancer via deep convolution neural networks using MRI images. *Multimedia Tools Appl* 79:15555–15573
8. Murtaza G, Shuib L, Abdul Wahab A (2020) Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artif Intell Rev* 53:1655–1720
9. Fusco R, DiMarzo M, Sansone C (2017) Breast DCE-MRI: lesion classification using dynamic and morphological features by means of a multiple classifier system. *Eur Radiol Express* 1(10)
10. Khairi R, Rustam Z, Utama S (2019) Possibilistics C-means (PCM) algorithm for the hepatocellular carcinoma (HCC) classificatio. IOP conference series: materials science and engineering, vol 1546, pp 052038
11. Borkowski K, Rossi C, Cirlitsis A, Marcon M, Hejduk P, Stieb S, Boss A, Berger N (2020) Fully automatic classification of breast MRI background parenchymal enhancement using a transfer learning approach. *Medicine* 99(29):e21243
12. Dalmış MU, Gubern-Mérida A, Vreemann S, Bult P, Karssemeijer N, Mann R, Teuwen J (2019) Artificial intelligence-based classification of breast lesions imaged with a multiparametric breast MRI protocol with ultrafast DCE-MRI, T2, and DWI. *Invest Radiol* 54(6):325–332
13. Sutton EJ, Onish N, Fehr DA, Dashevsky BZ, Sadinski M, Pinker K, Martinez DF, Brogi E et al (2020) A machine learning model that classifies breast cancer pathologic complete response on MRI post-neoadjuvant chemotherapy. *Breast Cancer Res* 22(57):1–12
14. Xie X, Shi F, Nitu J, Tang X (2018) Breast ultrasound image classification and segmentation using convolutional neural networks. In: Hong R, Cheng WH, Yamasaki T, Wang M, Ngo CW (eds) Advances in multimedia information processing—PCM 2018. PCM 2018. Lecture notes in computer science, vol 11166
15. Kalist V, Ganesan P, Sathish BS, Jenitha JMM, Shaik KB (2015) Possibilistic-fuzzy C-means clustering approach for the segmentation of satellite images in HSL color space. *Procedia Comput Sci* 57:49–56
16. Krishnapuram R, Keller JM (1996) The possibilistic c-means algorithm: insights and recommendations. *IEEE Trans Fuzzy Syst* 4(3):385–393
17. Özdemir Ö, Kaya A (2019) Comparison of FCM, PCM, FPCM and PFPCM algorithms in clustering methods. *Afyon Kocatepe Univ J Sci Eng* 19:92–102
18. Pawlovsky AP, Nagahashi M (2014) A method to select a good setting for the kNN algorithm when using it for breast cancer prognosis. *IEEE-EMBS International conference on biomedical and health informatics (BHI)*. pp 189–192
19. Sonar P, Bhosle U, Choudhury C (2017) Mammography classification using modified hybrid SVM-KNN. 2017 international conference on signal processing and communication (ICSPC). pp 305–311
20. Veeraraghavan H, Dashevsky BZ, Onishi N (2018) Appearance constrained semi-automatic segmentation from DCE-MRI is reproducible and feasible for breast cancer. *Radiomics: A Feasibility Study* *Sci Rep* 8
21. Zhang C, Li L (2014) 3D segmentation of masses in DCE-MRI images using FCM and adaptive MRF. *Proceedings, medical imaging 2014: image processing*, vol 9034. p 90344I

An EEG Atomized Artefact Removal Algorithm: A Review



Rudra Bhanu Satpathy and G. P. Ramesh

Abstract Signals from electroencephalograms (EEGs) are part of a growing body of biomedical data, which is used as EEG signals can be interfered with by artefacts created during the range and dispensation of the information, which can obscure the main features and information quality of signal. In order to confirm that the EEG signal does not lose any of its major attributes when diagnosing human neurological diseases such as epilepsy, tumours, and traumatic issues, these artefacts should be a review of EEG applications in health care is also provided, as well as a summary of challenges, research gaps, and future directions.

Keywords EEG · EMG · EOG · ICA · EEMD · DWT · Artefact removal

1 Introduction

Though neurons and other cells are interconnected in the nervous system, it is an extremely complex organ system that controls the body's functionality. A human body's external interface is comprised of the intellect, vertebral cord, and nerves. The intellect and vertebral cord are two examples of the dominant anxious system, which is responsible for controlling all of the body's activities. He/she also serves to drive information transmission and reception by way of the nervous system. Now, precise recordings of the brain's activity can offer a treasure of evidence regarding the body's functioning. In order to analyse the electrochemical nature of brain signals, electroencephalography (EEG) can be used [1]. Electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) are the only practical methods available for measuring intellect action exterior of an experimental set (fNIRS).

Biomedical signal evaluation accuracy is compromised. As these EEG data are extremely enervated and diverse through some non-cerebral instincts known as artefacts or noise, they are difficult to analyse and interpret. It is possible to classify these artefacts as either physiological or non-physiological. Variations in the desired

R. B. Satpathy (✉) · G. P. Ramesh

Department of ECE, St. Peters Institute of Higher Education and Research, Chennai, India

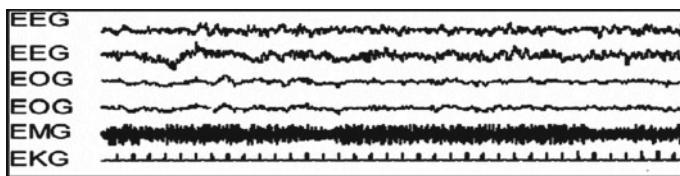


Fig. 1 Records of the EOG, EEG, EMG, and ECG (EKG) are overlaid

signal in the body cause physiological artefacts to appear. Figure 1 shows that eye crusade-connected artefacts, cardiac data, breathing data, and muscle rigidity data dignified by EMG are the most common sources of physiological artefacts, while electrode pop and alternating current effects are responsible for non-physiological artefacts.

Some of the most important artefacts that affect the excellence and evidence of signals are shown in Fig. 1, which results in false signals. For improved investigation and analysis of humanoid nervous illnesses, it is therefore necessary to identify and remove artefacts from the desired signal. Various research papers are analysed in this manuscript. Conclusion and summary of the experiments and openings in EEG signal artefact elimination, as well as prospects for improvement, are provided by this information.

Figure 2 denotes the pure EEG signal plot and its different artifactual plot. It is common for EEG epochs with signal intensities greater than the threshold to be rejected [2–4]. Due to this approach's inflexibility, meaningless information is lost. Aside from that, these artefacts will be masked by the EEG signal itself. Because of this, the important information will be lost if a threshold-based rejection is used.

Figure 3 designates the generalised block diagram for EEG artifactual elimination. To solve this problem, an automated component-based artefact separation approach is needed [5, 6]. Linear signal decomposition must be transformed into different source components in order to be useful.

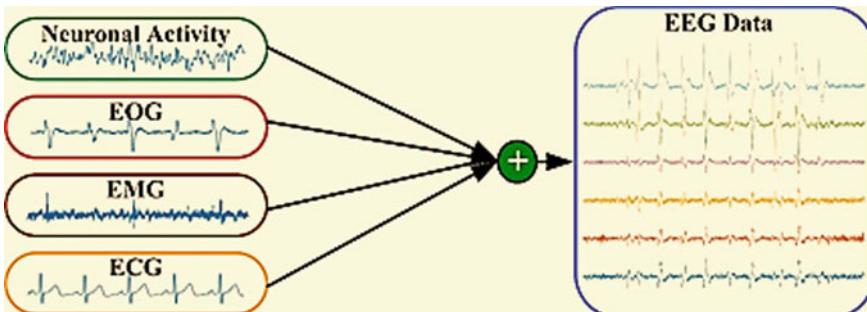


Fig. 2 EEG signal with various artefacts

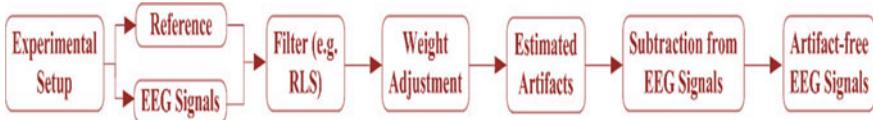


Fig. 3 EEG artefact removal block diagram

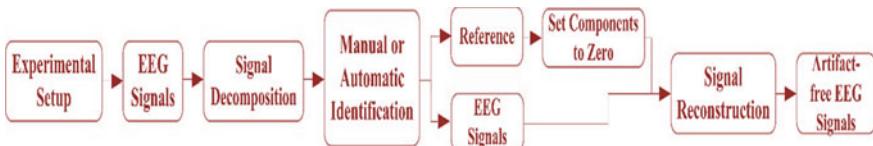


Fig. 4 Component-based EEG artefact removal block diagram

Figure 4 represents the EEG artefact elimination based on component-based analysis. Depending on the source type, the decomposed components will provide information. This is accomplished by collecting and analysing artefacts information from separate sources, then reconstructing the final signal in order to eliminate the effects of these sources [7–10].

The following procedures are the utmost generally applied to remove artefacts commencing EEG signal:

1. Blind source separations (ICA, CCA, and GECCA);
2. EEMD;
3. Wavelet transform (DWT and SWT).

According to a comprehensive study, BSS-ICA algorithms are frequently used to suppress artefacts in single- and two-stage processing. Although it relies on higher-order statistics, this algorithm can be time-consuming. CCA procedures are also chosen above ICA procedures because of their uncomplicatedness (founded on second-order indicators). EEMD methodologies are smeared to one-channel indicator alterations. In addition to wavelet transform algorithms, some algorithms centred on neural systems and optimization processes are also used to suppress artefacts [11–15].

EMG, EOG, and ECG are the three most common EEG signal artefacts to occur. They are categorised by artefact types and methods of elimination. The major assessment focuses on the investigation done to remove EOG artefacts, followed by a discussion of EMG artefact removal and automatic detection and removal of artefacts [16–19].

EOG is the most dominant artefact among all the other artefacts. Due to eye activities and eye blinks, EOG artefacts are affecting the EEG data at frontal electrodes. Signals from these sources will binge through the scalp and infect the clean EEG data as a consequence. These EOG artefacts are spectrally overlapping with EEG signals, making it difficult to remove them by using conventional methods [20–23].

This algorithm was developed by [24] and compared to recursive least squares (RLS) for the purpose of eliminating EOG artefacts from the EEG data.

The short time Fourier transform (STFT) was applied to reduce the computational time for EOG artefact removal while requiring less memory. To eliminate rapid eye movements, [25–29] proposes a wavelet transform-centred adaptive sifting methodology. They also found that the EEMD-CCA method was more efficient, taking less time to compute, and having improved signal and artefact ratio (SAR) and association constant. Soomro et al. [30, 31] established that the EEMD-CCA method was supplementary effective, taking fewer period to compute, besides having improved SAR and correlation coefficient.

Artefact detection and removal have improved in [32, 33]. Although the detection method is not very accurate, it is centred on the normalized correlation coefficient (NCC), and the EOG artefact is removed by applying the EEMD methodology. To remove EOG artefacts, [34–38] suggested using sample entropy enhanced wavelet ICA, which outperformed Zeroing-ICA and Wavelet ICA by a wide margin. A wavelet transform is used to identify and eliminate the independent blink component [39].

FOOBI outperformed ICA in removing ocular artefacts in [95]. The ORICA algorithm developed by Kuan-Ju Huang [40] eliminates eye blink artefacts based on sample entropy.

According to [41], automatic detection and suppression of eye artefacts can be achieved with the DWT algorithm when compared to SWT. According to the results, DWT is 25 times faster than SWT at removing EOG artefacts. Information pertaining to the nervous system, on the other hand, is less well preserved. Wavelet neural network algorithm was used to remove EOG artefacts in real time (WNN). Following training, artefacts are removed in accordance with the EOG behaviours that were learned. When used in real time, this method is extra computationally proficient than ICA [42–44]. Using an amalgamation of ICA and WNN, [45] proposes a more effective way to eradicate EOG commencing EEG signals. Recognition procedures of this type are multifaceted and require a lot of computation time to execute. References [46–48] propose a wavelet-based approach to eliminate EOG with CCA as well, and it performs better.

Healthcare systems with a single channel are preferred over those with multiple channels permitted to diminish the complication of the medical systems. To remove ocular artefacts, [49] proposed the use of EEMD-PCA on one channel of EEG signal. In order to eradicate eye blink artefacts from one-channel EEG, [50–53] developed the complete EEMD (CEEMD) and ICA algorithms. Performance is also compared and shown to be superior to that of the WICA, EMD-ICA, and EEMD-ICA. EOG artefact removal method is also based on this. They determined that SWT with algebraic verge outperformed DWT in terms of preserving neural information in the EEG, while the DWT with arithmetic verge had a quicker implementation time than the other method.

On the basis of wavelet-enhanced canonical correlation analysis, Zhao suggests a technique for automatically eradicating ocular artefacts commencing EEG records (WCCA). Over popular methods such as CCA, ICA, and WICA, WCCA offers

two distinct advantages. A subjective visual inspection of artefact components is not required, because the principal canonical constituents found by CCA for every dataset, which are correspondingly the furthermost collective components among leftward and rightwards hemispheres, are certainly associated to artefacts. According to second-generation EEG signals, WCCA removed the greatest number of ocular artefacts with the least amount of cerebral information lost [54–58].

According to [59], some research has also motivated on adaptive artefact elimination for both the EOG and the EMG. First, the artefacts are classified, and then, second-order blind identification (SOBI)-SWT and CCA-SWT are applied to remove EOG besides EMG artefacts, respectively. Compared to existing methods of artefact removal, this adaptive algorithm produced better results [60].

Some researchers have recently focused on the removal of EMG artefacts. As a result of swallowing, walking, and talking, the potential for artefacts was generated. Compared to the brain signal, EMG artefacts have a wider spectrum. On top of that, the duration and frequency of this EMG can be straightforwardly detached commencing the structure. According to [61], EMD-ICA amalgamation procedure is operational for small SNR, while 2 T-EMD or contrast maximization 2 (CoM2) is more effective for high SNR. If mathematical complication is taken into deliberation, DWT or CCA is preferred.

As demonstrated by B. Mijovic [8], there are a number of blind source separation (BSS) methods as well as independent component analysis (ICA) available for decomposing the signal into its components in multichannel measurements, while for single-channel analysis, SCICA and WICA can be used to separate the signal into its components. It was concluded that SCICA had the worst performance in this paper. Even though the WICA technique performs less well than the EEMD-ICA technique, it is still superior to it.

Authors [66] have also proposed a multivariate-EMD approach for removing EMG artefacts and compared it to ICA based on SNR and MSE. Muscle artefacts, on the other hand, obscure EEG signals and complicate their analysis, as demonstrated by [62]. A method for cancelling muscular artefacts has been proposed by the authors for the one-channel EEG circumstance, which is commonly used. Ensemble empirical mode decomposition (EEMD) algorithm and joint blind source separation (JBSS) techniques are combined in this approach.

DWT and ICA are combined in WICA, which makes use of multi-perseverance and multi-dimensional investigates in tandem. EMG signals were first processed using WICA [59]. This is because WICA projects data into an area where severance is developed and artefact structures are completely utilised. There are additional super-Gaussian features in the probability density function (pdf) of wavelet coefficients than there are in the fresh data. Expending a mention channel and WICA, Li et al. applied automatic EOG artefact decline and demonstrated that WICA significantly recovers the SNR of EEG signals as well as anti-noise capability [63].

According to the authors, the enactment of EEMD-CCA was marginally greater to that of EEMD-IVA. By using EEMD-MCCA, this approach is further extended [15]. According to [50, 51], epilepsy patients' EMG artefacts were suppressed using

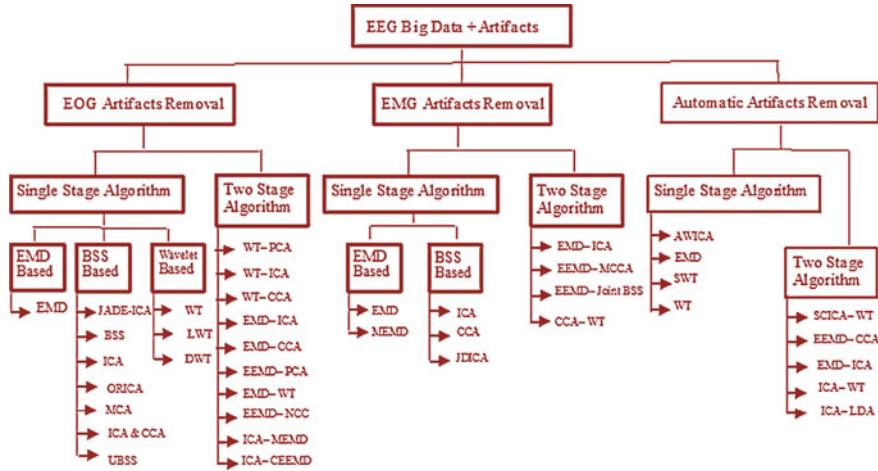


Fig. 5 Taxonomy of EEG artefact removal algorithms

CCA and CCA-WT methodologies. In a table, all EEG artefacts removal research papers are compared based on a few key characteristics.

Applications of EEG in the health care system have grown along with the use of ambulatory devices. There are some unintended signals (artefacts) that must be removed from real-time applications in order to recover the investigation and identification of humanoid nervous illnesses for healthcare resolutions. Among the furthermore common EEG signal artefact elements, EMG, EOG, ECG, and motion artefacts are the most troublesome. Figure 5 depicts the classification of artefacts elimination procedures according to artefact type.

Above diagram 5 shows a summary of the different algorithms that have been established to eradicate artefacts from the EEG data. Algorithms are categorised based on the types of artefacts they produce. According to the directly above shown figure, cascaded procedures are additional operative than single-stage algorithms at removing artefacts. Further, the type of signal used in the analysis is also important. While ICA algorithm or wavelet transform (WT) procedures are applied to subdue the artefacts in multichannel signals, single-channel signals are first converted to multichannel using EMD-based approaches, followed by BSS or WT-based algorithms.

Due to the frontalis and temporal muscles, EMG artefacts are the most common cause. To remove muscle artefacts, [10] used the EEMD-ICA method; however, the EEMD-CCA algorithm was developed to improve this process by [16]. The performance of EEMD-CCA is compared to that of EEMD-ICA, and it is shown to be superior. EEMD-MCCA is also used in [15] to improve the removal of EMG artefacts by aggregating PSNR and decreasing RMSE parameters compared to other motion artefact elimination methods. To remove EMG artefacts, [50] implemented the CCA-WT algorithm.

Attributable to eyelid crusade and eye blinking, electrooculograms (EOGs) are the most corrupting artefacts in EEG signals. To suppress EOG artefacts, [69] applied the Haar type wavelet-dependent ICA method, using entropy as the arithmetical degree. Additionally, in [42], an automatic EOG artefact detection method using WICA was employed, and a modified multi-scale entropy statistical measure was considered. When plotting the ROC curve, it becomes clear that the sensitivity and specificity have increased significantly. To remove EOG artefacts, CCA method [47] reduces complication and calculation period of the artefact elimination procedure.

To adaptively remove eye blink, EOG and EMG have used the involuntary recognition and rectification of artefact procedure with the independent component ensemble. Aimed at the automatic recognition besides correction of artefacts in EEG signals, [19] has used the ICA-LDA algorithm.

Table 1 summarises the results of the study and investigation, which show that certain artefact elimination procedures are operative depending on the category of involved artefacts.

As shown in Table 1, blind source separation (BSS) procedures are the utmost operative at removing EOG artefacts, while CCA-WT is the utmost operative at removing EMG artefacts.

There are four main factors that determine how effective an artefact removal method is: PSNR, RMSE, and correlation coefficient. Subsequently, many proficient

Table 1 Different EEG artefacts and its solicitations

Type of artefact	Artefact removal algorithms
Electrooculogram (EOG)	In many cases, a single-stage BSS is used (PCA, ICA, and CCA) There are few ways to use WT-BSS for multichannel inputs Single-channel input two-stage approaches are compatible with EEMD-BSS EOG artefact suppression is most effective when done in two stages cascaded
Electromyogram (EMG)	Single-stage approaches use EEMD and BSS algorithms Single-channel, two-stage EEMD-BSS is used In a multichannel two-stage approach, the CCA-WT is applied. EMG artefact suppression is most effective when BSS is used in conjunction with SWT
Automatic artefact detection and removal	Single-stage approaches are used, for example, EMD, ICA, and SWT There is an application of EEMD-BSS for a single-channel cascaded approach There is a BSS-WT applied to a multichannel two-stage approach There are also neuro-fuzzy and optimization algorithms

methods for instance CCA, ICA, DWT, SWT, EEMD, and others can be used to remove dissimilar artefacts (EOG, EMG and so on) commencing an contributed EEG data can be done more quickly, reliably and accurately using these. Both single-channel and multiple-channel EEG signals can be used with these methods to remove artefacts. As well as varying recording duration and sampling rate, the input EEG signal can also have a wide range of With certain conditions, for instance SNR standards, integer of channels, and types of illnesses, e.g. some artefact removal methods.

To begin with, it is important to note that the great PSNR assessment corresponds to better EEG data excellence, while the lowest RMS error value indicates improved artefact removal. It is clear from the improved correlation coefficient [18] that the artefacts in the noisy EEG indicator can be improved and recognised besides separated, leading to better source separation. Techniques for removing artefacts can be made more complex or better depending on the algorithm's speed and accuracy. Artefact elimination approaches and their calculation period affect the complexity of the methodology. As an artifact-removal algorithm, ICA takes much longer to compute than CCA, EEMD, or DWT also requires the least amount of computational time [22]. How quickly artefact removal methods can be executed will be influenced by the computational time, and consequently, the algorithm's speed will be slowed by the longer computation.

Artefact removal techniques such as those described above are extremely important and helpful in diagnosing neurological diseases such as epilepsy, tumours, and sleep apnea, among other conditions. EEG artefact removal is still being researched, which will lead to an additional precise identification and treatment of nervous sicknesses in the future.

2 Conclusion

These papers have been thoroughly studied, and it has been determined that artefact elimination approaches are essential primary steps for EEG data. Medical professionals will be able to use these cleaned EEG signals to make more accurate diagnoses and analyses. There is a lot of research done on removing EOG, EMG, besides muscle artefacts. In the inclusive evaluation work, the methods used to remove various artefacts in the EEG signals are grouped into different categories.

There are many algorithms like EEMD, DWT, SWT, ICA, and CCA, and occasionally, amalgamations of these approaches are the most commonly used artefact removal algorithms. Some performance evaluation parameters, such as peak signal-to-noise ratio (PSNR) parameter, root mean square error (RMSE) parameter, and association factor, have been applied to compare these methods and demonstrate their effectiveness through simulations. Final conclusion: Blind source separation (BSS) methods are the utmost extensively applied procedures to eradicate EOG artefacts commencing EEG data, according to research and analysis.

Using the BSS algorithms, you can easily remove artefacts by identifying the artefact source. It is also extra operative in eradicating EMG artefacts from EEG data.

They will even out the EMG artefacts and extensive range unpredictability present in the EEG data, while conserving neural evidence. As a result of the assessment's investigation of the aforementioned investigation papers, it has been determined that a cascade of diverse artefact elimination procedures can be supplementary operative in removing. Because of this, n's analysis and diagnosis will benefit from the signal's processing.

References

1. Mahadevan AA, Mugler D (2008) Ballistocardiogram artifact removal in EEG-fMRI signals using discrete Hermite transforms. *Signal Process* 2(2):839–853
2. Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artefacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34(4):1443–1449
3. Casson AJ, Villegas ER (2008) On data reduction in EEG monitoring: comparison between ambulatory and non-ambulatory recordings. International IEEE EMBS conference vancouver, pp 5885–5888
4. Mert A, Akan A (2013) Hilbert-Huang transform based hierarchical clustering for EEG denoising. *Signal processing conference (EUSIPCO)*, vol 1–5
5. Qazzaz A, Ali NK, Ahmed S, Islam MS, Ariff MI (2015) Selection of mother wavelets thresholding method in denoising multichannel EEG signals during working memory task. *Sensors* 15:29015–29035
6. Hassan AR, Bhuiyan MIH (2016) Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. *Biocybernetics Biomed Eng* 36:248–255
7. Turnip A (2014) “JADE-ICA algorithm for EOG artifact removal in EEG recording. technology, informatics, management, engineering, and environment. TIME-E, pp 270–274
8. Mijovic BV, Huffel SV (2010) Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *Biomed Eng* 57(9):2188–2196
9. Majumdar CA, Morshed BI (2015) Real-time hybrid ocular artifact detection and removal for single channel EEG. *Electro/information technology (EIT)*. pp 330–334
10. Burger C, Heever D (2015) Removal of EOG artefacts by combining wavelet neural network and independent component analysis. *Biomed Signal Process Control* 15:67–79
11. Mosquera CG, Vázquez AN (2010) Automatic removal of ocular artefacts using adaptive filtering and independent component analysis for electroencephalogram data. *Signal Process IET* 6(2):99–106
12. Kuo CY, Wei SK, Tsai PW (2013) Ensemble empirical mode decomposition with supervised cluster analysis. *Adv Adapt Data Anal* 5(1):1–13
13. Teng CZ, Wang G (2014) The removal of EMG artifact from EEG signals by the multivariate empirical mode decomposition. *Signal Processing, Communications and Computing*, pp 873–876
14. Zhao C, Qiu T (2015) An automatic ocular artifacts removal method based on wavelet-enhanced canonical correlation analysis. *Engineering in medicine and biology society (EMBC)*, pp. 4191–4194
15. Donoho DL, Johnstone IM, Kerkyacharian G, Picard D (1995) Wavelet shrinkage: asymptopia? *J R Stat Soc B* 57:301–369
16. Moretti DV, Babiloni F, Carducci F, Cincotti F, Remondini E, Rossini PM, Salinari S, Babiloni C (2003) Computerized processing of EEG–EOG–EMG artefacts for multicentric studies in EEG oscillations and event-related potentials. *Int J Psychophysiol* 47:199–216
17. Hosna G, Erfanian A (2010) A fully automatic ocular artefact suppression from EEG data using higher order statistics: improved performance by wavelet analysis. *Med Eng Phys* 32(7):720–729

18. Rilling G, Flandrin P, Calves PG, On empirical mode decomposition and its algorithms. IEEE-EURASIP workshop on nonlinear signal and image processing NSIP-03
19. Abdullah H, Cvetkovic D (2013) Double density wavelet for EEG signal denoising. International conference on machine learning and computer science (IMLCS'2013), Kuala Lumpur (Malaysia)
20. Ghandeharion H, Erfanian A (2006) A fully automatic method for ocular artifact suppression from EEG data using wavelet transform and independent component analysis. Engineering in medicine and biology society. pp 5265–5268
21. Daly IS, Putz GM (2015) FORCe: fully online and automated artifact removal for brain-computer interfacing. *Neural Syst Rehabil Eng* 23(5):725–736
22. Matiko JB, Tudor J (2013) Real time eye blink noise removal from EEG signals using morphological component analysis. Engineering in medicine and biology society (EMBC). pp 13–16
23. Hu JW, She J (2015) Removal of EOG and EMG artifacts from EEG using combination of functional link neural network and adaptive neural fuzzy inference system. *Neurocomputing* 151:278–287
24. Zheng J, Cheng J, Yang Y (2014) Partly EEMD: an improved noise assisted method for eliminating mode mixing. *Signal Process* 96:362–374
25. Huang KJ, Liao JC, Shih WY, Feng CW, Chang JC, Chou CC, Fang WC (2013) A real-time processing flow for ICA based EEG acquisition system with eye blink artifact elimination. IEEE workshop on signal processing systems. pp 237–240
26. Huang KL, Fang W (2013) A realtime processing flow for ICA based EEG acquisition system with eyeblink artifact elimination. *Signal processing systems (SiPS)*. pp 237–240
27. Sweeney KT, Onaral B (2012) A methodology for validating artifact removal techniques for physiological signals. *Inf Technol Biomed* 16(5):918–926
28. Sweeney KT, Ward TE (2013) The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique. *Biomed Eng* 60(1):97–105
29. Anastasiadou MC, Mitsis GD (2015) Automatic detection and removal of muscle artifacts from scalp EEG recordings in patients with epilepsy. Engineering in medicine and biology society. pp 1946–1950
30. Soomro MH, Jatoi MA (2013) Automatic eye-blink artifact removal method based on EMD-CCA. *Complex medical engineering (CME)*. pp 186–190
31. Soomro MH, Yusoff M (2014) Comparison of blind source separation methods for removal of eye blink artifacts from EEG. *Intelligent and advanced systems (ICIAS)*. pp 1–6
32. Anastasiadou MC, Mitsis GD (2014) Detection and removal of muscle artifacts from scalp EEG recordings in patients with epilepsy. *Bioinformatics and bioengineering*. pp 291–296
33. Betta MG, Menicucci D (2013) Detection and removal of ocular artifacts from EEG signals for an automated REM sleep analysis. Engineering in medicine and biology society (EMBC). pp 5079–5082
34. Jansen M, White TP, Mullinger KJ, Iddle EB, Gowland PA, Francis ST, Bowtell R, Liddle PF (2012) Motion-related artefacts in EEG predict neuronally plausible patterns of activation in fMRI data. *Neuroimage* 59(1):261–270
35. Islam MK, Rastegarnia A, Yang Z (2015) A wavelet-based artifact reduction from scalp EEG for epileptic seizer detection. *IEEE J Biomed Health Inform*
36. Mowla MR, Paramesran R (2015) Artifacts-matched blind source separation and wavelet transform for multichannel EEG denoising. *Biomed Signal Process Control* 22:111–118
37. Akhtar MT, James CJ (2012) Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data. *Signal Process* 92:401–416
38. Babu PA, Prasad K (2011) Removal of ocular artifacts from EEG signals using adaptive threshold PCA and wavelet transforms. *Communication systems and network technologies*. pp 572–576
39. Jadhav PA, Naik G (2014) Automated detection and correction of eye blink and muscular artefacts in EEG signal for analysis of autism spectrum disorder. Engineering in medicine and biology society. pp 1881–1884

40. Zhao QH, Peng H (2014) Automatic identification and removal of ocular artifact in EEG-improved adaptive predictor filtering for portable application. *NanoBioscience* 13(2):109–118
41. Croft RJ, Barry RJ (2000) EOG correction: which regression should we use? *Psychophysiology* 37:123–125
42. Mahajan R, Morshed BI (2014) Unsupervised eye blink artifact denoising of EEG data with modified multi-scale sample entropy, Kurtosis, and wavelet-ICA. *Biomed Health Inf* 19(1):158–165
43. Vazquez RR, Maquin D (2007) EEG ocular artefacts and noise removal. *Engineering in medicine and biology society*. pp 5445–5448
44. R. S. Patel and N. Mariyappa (2015) Ocular artifact suppression from EEG using ensemble empirical mode decomposition with principal component analysis. *Comput Electr Eng*
45. Haoni S, Chauxie W, Pandey M. Application of Hilbert-Huang transform in generating spectrum-compatible earthquake time histories. *ISRN Signal Process* 2011:1–17. Article Id 563678
46. Khatun SM, Morshed BI (2015) Comparative analysis of wavelet based approaches for reliable removal of ocular artifacts from single channel EEG. *Electro/information technology (EIT)*, pp. 335–340
47. Postalcioglu S, Erkan K, Bolat ED (2005) Comparison of Kalman filter and wavelet filter for denoising. *Int Conf Neural Networks Brain* 2:951–954
48. Priyadharsini SS, Rajan SE (2014) An efficient method for the removal of ECG artefact from measured EEG signal using PSO algorithm. *Int J Adv Soft Comput Appl* 6(1):1–19
49. Sanei S, Chambers JA (2007) Introduction of EEG- EEG signal processing. Wiley. ISBN-10: 0470025816
50. Lagerlund TD, Sharbrough FW, Busacker NE (1997) Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *J Clin Neurophysiol* 14(1):73–82
51. Chiong TL (2006) Sleep: a comprehensive handbook. Wiley
52. Radüntz TS, Meffert B (2015) EEG artifact elimination by extraction of ICA-component features using image processing algorithms. *J Neurosci Methods* 243:84–93
53. Emir U, Akgul CB, Akin A, Harmanci K (2003) Wavelet denoising versus ICA denoising for functional optical imaging. International IEEE EMBS conference on neural engineering. IEEE
54. Chang WC, Im C (2015) Detection of eye blink artifacts from single prefrontal channel electroencephalogram. *Comput Methods Programs Biomed* 124:19–30
55. Hsu WL, Chen I (2012) Wavelet-based envelope features with automatic EOG artifact removal: application to single-trial EEG data. *Expert Syst Appl* 39:2743–2749
56. Blume WT, Kaibara M, Young GB (2002) Atlas of adult electroencephalography. Lippincott Williams and Wilkins, Philadelphia
57. Chen XH, Peng H (2014) Removal of muscle artifacts from single-channel EEG based on ensemble empirical mode decomposition and multi-set canonical correlation analysis. *J Appl Math* 1–11
58. Chen XL, Ward RK (2014) A preliminary study of muscular artifact cancellation in single-channel EEG. *Sensors* 14:18370–18389
59. Lu Y, Oruklu E, Sanei J (2013) Chirplet signal and empirical mode decompositions of ultrasonic signals for echo detection and estimation. *J Signal Inf Process* 4:149–157
60. Liu Z, Zhang Z (2015) The improved algorithm of the EMD decomposition based on cubic spline interpolation. *Signal Process Res* 4:63–68
61. Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal* 1(1):1–41
62. Ge SH, Hong X (2014) A fully automatic ocular artifact removal from EEG based on fourth-order tensor method. *Biomed Eng Letter* 455–463

63. Sardouie SH, Merlet I (2015) An efficient Jacobi-like deflationary ICA algorithm: application to EEG denoising. *Signal Process Lett* 22(8):1198–1202
64. Bizopoulos PA, Fotiadis DI (2013) An automatic electroencephalography blinking artefact detection and removal method based on template matching and ensemble empirical mode decomposition. *Engineering in medicine and biology society*. pp 5853–5856
65. Kanoga S, Mitsukura Y (2015) Eye blink artifact rejection in single-channel electroencephalographic signals by complete ensemble empirical mode decomposition and independent component analysis. *Engineering in medicine and biology society*. pp 121–124

A Journey of Artificial Intelligence and Its Evolution to Edge Intelligence



P. Britto Corthis and G. P. Ramesh

Abstract The progressions in the innovations and the expansion in the computerized scaling down are making gadgets become more intelligent. The development of Internet of Things (IoT) and the cloud platform have made things stunningly better. Moving all information to the cloud for handling has steadily neglected to meet the continuous prerequisite of IoT benefits because of high network traffic and delay. The restrictions in cloud platform in resource allocation, service providing, and delay in transmission prompt us to move from cloud to edge. Edge computing (EC) moves the computation part from the cloud to the edge nodes (ENs). This will considerably improve the quality of service (QoS) of the IoT programs that demand less latency. With the forward leaps in deep learning, recent research has seen a blasting of artificial intelligence (AI) applications in various fields like suggestion systems, decision systems, and surveillance systems. The rise of the artificial intelligence in the edge computing has ended up being main focus as it boosts up the speed and number of the IOT applications. Edge computing that pushes processing and services from the network center to the network edge, widely accepted as a productive solution. The outcome is edge AI, or edge knowledge (EI) is starting to get a huge measure of interest. The research on edge intelligence is only in the starting stage. We do a comprehensive survey of the recent research efforts on evolution of EI, the security challenges encountered by distributed ENs as they are more defenseless against attacks for shortened computing equipment and storage. The incredible learning capacity of AI empowers the framework to recognize malignant assaults more precisely and effectively.

Keywords Artificial intelligence • Edge nodes • Edge computing • IoT

P. Britto Corthis (✉)

Research Scholar, Department of ECE, St. Peter's Institute of Higher Education and Research, Chennai, India

G. P. Ramesh

Professor, Department of ECE, St. Peter's Institute of Higher Education and Research, Chennai, India

1 Introduction

With the ubiquitous use of sensors in the real world, an expanding number of physical entities are being connected to the Internet of Things (IoT) via sensors in order to share information. IoT technology is widely used in a variety of applications, including smart cities, home automation, wearable healthcare products, and environmental sensing [1–3]. In classical IoT services, the data from networked sensors and other widgets must be sent to cloud servers for consolidation and computing. They are groomed and returned to the device after the consolidations are completed. Though the cloud reduces the technical stress of sensors and devices, massive channeling overhead of the data cannot be avoided. There is a tremendous increase in the amount of electronic devices used in the world, and it is projected to expand to another level by 22 [4]. Growth of information is rapid. At the same time, we also have upgrades in network devices, bandwidth, and memory devices to manage the projected data. However, bandwidth expansion lags significantly behind data expansion. To overcome the aforementioned barrier, a new computing prototype known as edge computing (EC) has been introduced, which receives global attention. The computational duties are pushed to the network's edge by EC [5, 6]. When it comes to user privacy, edge computing outperforms cloud computing due to network delays during data transmission. Edge computing also minimizes network bandwidth stress while cutting data center energy consumption. Natural data triggered by smart devices does not need to be synced to the cloud in EC. All computations are done at edges and transmitted, thus reducing the latency time and delays.

Conversely, EC sets out many security challenges and multiply the target locations [7] of the system. The different factors are:

- (a) Layout: The edge nodes are much diffused at heterogenous regions on the edge of the entire network [8]. It is difficult for any system to unite all smart devices under a principal unit. The opposer can attack the ENs that have safety defects. Through that affected node, opposer can enter into the system very easily.
- (b) Resource limitation: As the functionality of edge nodes are inadequate in most of the computations, strong security algorithms cannot be applied on those nodes. Some common attacks on ENs lead to a great damage in the case of EC.
- (c) Heterogeneous environment: Extensive sweep of networks and technologies are involved in EC. Due to this diversified environment, it is very burdensome to make a single or consistent safety mechanism or policy or algorithm.

In order to pacify the protection pitfalls triggered by the features of edge computing, lot of security methods, mechanisms, protocols, and algorithms are materialized [9, 10]. The security mechanisms are mostly lodged on the algorithms that are used currently in the fields of intrusion detection, authentication control, privacy preservation, or access control. Inspiringly, the advent of artificial intelligence (AI) supports EC with mighty solutions to the above-mentioned privacy issues. In the case of DoS attack and DDoS attack, machine learning (ML) can be used to monitor and analyze the requests and attacks from all edge nodes. ML algorithms clip the

malignant request patterns by assessing former datasets [11, 12]. All smart devices connected in IoT contain secret, emotional, and sympathetic information [13]. As we discussed earlier, application of regular cryptographic methods for preserving the secrecy of data, encryption of data demand raised mathematical overhead. Deployment of such solutions on edge nodes will not be a beneficial one. Instead of common methods, distributed machine learning (DML) mechanisms can be used with proper dataset training on ENs to lessen the stress during data transmission [14].

As the number of devices connected to the smart system is raising, we will face accessing issues. We need to have a proper access control mechanism to differentiate different requests travel across the sensors and devices for the access of devices or data. The authentication process is a complex as the devices are heterogenous. Only with proper authentication device can access other devices or data. [15, 16]. Using classification algorithm in ML, ENs are classified into categories based on the access permissions [17]. High priority devices can be allowed to access to devices for which they posted the request message. As various researches indicate the advancements in EC with AI, AI has systematically been enforced in many smart applications to provide edge security [18, 19]. However, there are still many dares in the training process in ML. Training efficiency will get increased if we provide clean data for training. But the ground of edge computing, attacks and malicious intrusion are very common in request response sessions [20]. So it is mandatory for the system to be vigilant to get benefited [21].

This review focuses on current technologies, mechanisms, and achievements in edge computing. We have also presented the related topics to edge computing and evolution of edge computing.

2 Basic Concepts

2.1 *IoT and Cloud*

IoT is to develop a worldwide organization of things where everything is associated with the Internet, thus understanding the interconnection of all articles/sensor utilizing Internet innovations. With IoT innovation, gadgets can communicate data to one another and a few gadgets; sensors cooperate to do a job without the mediation of people. IoT can be applied in different enterprises by implanting sensors into items like clinical hardware, home gear, transports, carrying out the incorporation of human culture, and actual world. The three regular IoT administrations are remote observing and control, keen home, and catastrophic event expectation. IoT permits clients to control the gadgets associated with the Internet and screen a situation distantly, which carries liberal accommodation to our life and well-being. With the use of IoT, keen home items contain an immense potential to turn out to be more smart and flexible, ready to serve clients better. Assume that when you go into your room from an external perspective, the climate control system is turned on and

acclimated to an agreeable temperature consequently for you. Numerous goods, like as cleaning robots, will relieve you of duty, and even lights may be turned on and off without the assistance of anyone else and with no manual exertion. As a result, it is easy to understand how smart homes are one of the most visible examples of how IoT administrations make our lives easier and more pleasant. IoT assumes a significant part in the expectation of catastrophes like seismic tremors, floods, dry season, and torrents. Sensors sent outside are designated to accumulate information from the surrounding climate and the prepared information might uncover critical data about the coming regular cataclysm, in this manner setting aside sufficient time for us to eliminate individuals from the war zone and keep away from property misfortune, however, much as could reasonably be expected. Up until now, on the subject of advantages IoT achieves, we have just alluded to a glimpse of something larger. Certainly, IoT has filled in as an incredible motor driving insurgencies in numerous conventional disconnected enterprises. However, IoT is as yet in its underlying state, it has a wide application range which is simply restricted by human's creative mind, and it will undoubtedly impact pretty much every part of our life sooner rather than later. Numerous applications are improved by incorporating the IoT and distributed computing.

2.2 Artificial Intelligence (AI)

In the era of IoT, artificial intelligence has touched spotlight and acquired colossal consideration. It was first authored in 1956. Basically, AI is a way to deal with construct astute machines fit for doing errands as people do. This is clearly an exceptionally wide definition, and it can allude from Apple Siri to Google AlphaGo and too strong advances yet to be designed. In mimicking human insight, AI frameworks ordinarily show a portion of the accompanying practices related with human knowledge: arranging, picking up, thinking, critical thinking, information portrayal, discernment, movement, and control and, less significantly, social insight, and innovativeness. During the previous long term's turn of events, AI has encountered rise, fall, and again rise and fall. The most recent ascent of AI after 2010s was somewhat because of the leap forwards made by profound learning, a strategy that has accomplished human-level precision in some fascinating regions.

2.3 IoT and Artificial Intelligence

Developments in all fields has grabbed the eye of IoT-empowered smart fields by co-ordinating AI system with advanced network innovations while sending media content. An integration of various working systems and wearable gadgets can serve the every single corner of the world. However, few difficulties are faced by the end users. Both IoT and artificial intelligence are mainly used in medical field to reduce the burden of

medicos. Usage of smart devices in medical field, the progress is reformed. Currently, the AI-induced edge computing is very fundamental in all manufacturing practices and resolves significant issues [201]. In manufacturing industries, IoT gadgets are constantly sending information to the distant worker to check and calculate the results [22]. The sensors and smart devices connected over IoT are distributed. However, the small gadgets involved are normally very prone to battery drain problem as the battery lifetime is less [23, 24]. There is a compulsory need for traditional batteries in smart sensor devices to be consistently re-energized or supplanted. There comes the recovery sensors which reduces the difficulty of changing batteries often. They will find the minor flaws in energy systems and recover the system [25]. The edge intelligence present in the network gathers information in the squares present, sends the needed data to cloud, analyzes the data, and does the computation for monitoring system.

2.4 Edge Computing

Edge computing (EC) is a revolutionary innovation that processes and stores data at the network's edge, close to smart devices, sensors, and client computers [26]. Mostly, the input devices are mobile phones and other personal devices. Edge computing is distributed in nature, but brings the computations nearer to the devices to reduce the transfer of data over the network which is prone to any cryptographic attacks. It is unnecessary to transfer the sensitive information to the cloud center and then back to smart devices. We face transmission bandwidth issues, delay in transfer issues, and security issues in the distributed environment. Also it is not possible for the cloud to manage such a huge data generated by large-scale mobile devices. Cloud speed and processing efficiency will be reduced [27, 28]. We do have some advanced cloud mechanisms like AR and VR [26]. They have higher prerequisites for low delay and quick action time. The inconsistency between efficient processing of data and providing secure communication brings the edge computing to the IoT industry.

2.4.1 Benefits of EC

- (a) Latency: Rather than sending all information to cloud, calculations and processing are done at the edge which is very nearer to smart devices, consequently speeding up the actions and lessening the latency [29].
- (b) Privacy and Security: The data are made local in ENs in EC and provides protection against threats and attacks by providing cryptographic mechanisms at edge level. With the advent of various security algorithms, data protection is assured to an extent and data loss has been expertly scaled down [29].
- (c) Energy utilized: in EC, some portion of processing assignments is offloaded to a few ENs, which calms the weight of data transfer capacity [26] as well as diminishes the energy utilization in the cloud.

2.5 IoT Framework with Edge Computing

The present IoT administrations are predominantly cloud-based and incorporated with the goal that all information handling and investigations must be finished in cloud itself. With the success of IoT, more IoT gadgets are requesting low dormancy and high response time. But cloud is incapable to offer help continuously for this drastic improvement of IoT. Just making use of best benefits of EC, we can yield creamy IoT services. A layered structure can be designed for mapping IoT system with EC. Each layer has its own functionalities. We can see the layers below.

Device layer has all electronic gadgets like mobiles, PCs, vehicles, and personal devices, and every one of the gadgets are furnished with various types of detectors or sensors like RFID, smart sensors, and QR code. All these gadgets are capable of transmitting data packets to one another in real time. As they are collected and transmitted real time and retrieved from heterogenous devices, the data will vary.

Network layer is a pathway between cloud, edge, and smart gadgets. It is linking the detected gadgets in IoT system. The information received from detected gadgets are sent through various associated transmission technologies. The transmission technology may use different protocols for effective and quality communication. The necessity for cloud-edge cooperation has grown in recent years. Smart cities, smart farms, smart homes, public safety, and automated cars are just a few examples of IoT services that use edge computing. Edge layer tackles partially the issues of deficient transfer speed and high conveyance delay. To handle an incredible amount of information from IoT gadgets productively and precisely, part of edge servers are involved in preparation, processing, managing, and storing of information. The outcomes are sent to mapped gadgets or transferred to the cloud layer for additional investigation or storing through the network layer. Cloud layer is the backbone of entire administrative system of IoT. It is typically made out of enormous cloud data stores with exceptional processing power. In this framework, cloud layer is used for additional and advanced handling of data from the edge layer. When the edge layer's resources fail, the cloud layer can provide processing help in the form of virtual machines. When an edge layer becomes malevolent or ineffective, a cloud layer with appropriate security mechanisms can take over the function of the edge layer. As a result, the combination of IoT with EC has emerged as a viable option which has a prominent role in this field. Few areas where both were implemented successfully are smart home, smart city, smart farms, smart industries, automated vehicles, and unmanned vehicles.

3 Journey of Edge Intelligence

3.1 Edge Intelligence

The AI applications are flourishing with the growth of deep learning and the numerous enhancements in equipment designs. Plenty of bytes of data need processing at network edge which leads to the integration of edge computing with artificial intelligence. Either we can apply AI on edge or we can use AI for edge. When we use AI for edge, most of the important problems will be solved with optimal solutions. In the case of other, we have to build AI models training on the edge.

3.1.1 Benefits of Edge Intelligence

The combination of AI and edge processing is normal since there is a reasonable convergence between them. In particular, edge computing targets in connecting huge number of communitarian edge gadgets and edge servers to handle the produced information in its range. At this point, AI can be used as an intelligent agent like a human intelligence by learning the data. So the benefits can be doubled. Traditional edge computing gadgets have no intelligent abilities. These gadgets are answerable for handling adjacent information based on the algorithms given in the edges servers, for example, feature extraction [30] and communicating information to cloud servers. When we introduce machine leaning in edges, edge devices become without even the cloud [31]. This decreases the delay and response time [32]. There are various reasons why do we have to move intelligence from cloud to edge.

Security becomes a challenge if we use cloud servers for storing and processing as it is very huge. The distance between the device and the data is a barrier when we try to transmit the result to the device [33]. As the cloud handles very large amount of data, delay is inevitable. But in the case of smart home, smart station, smart transport, and other smart systems, we need instant actions without delay. Therefore, we prefer edge computing. The devices which are connected to the IoT are continuously sending information to the cloud for storing and processing (e.g., camera, sensors). As the number of devices connected to the system and the data size increases drastically, the bandwidth consumption required for this communication is also increased, which in turn reduces the performance of the system [30]. During transmission, the data are vulnerable to threat and attacks which destroy the originality of data [30].

Application of machine learning algorithms to these edge systems has a few difficulties:

We need higher-end devices to run machine learning algorithms effectively. But the edge devices are heterogenous and differ in their performances. This increases the complexity in computations and performance which thrives us to develop some lower-end machine learning algorithms [34].

AI algorithms need more memory to store AI model and other required information. Thus, there is a need to plan AI procedures that can run on less resources

gadgets [34]. Deep learning algorithms are layered, with each layer which is used for fetching different features from data. This works well with edge than machine learning algorithms. Because it processes the required information, other irrelevant information can be automatically sent to cloud. The implementation of deep learning in edge computing has many challenges.

- Deep learning methods need cloud [35].
- Edge devices are distributed all over the network and the mechanism to find the appropriate device is required [36].
- Programs which partitions the data automatically among the different edge devices for proper processing need to be written [36].

3.1.2 Driving AI to the Edge

- (1) Data from network edge make AI as a fully functional one.
- (2) Edge computing helps AI to prosper with various applications and loaded data.
- (3) AI advances in edge computing.
- (4) Edge computing is popular and trendy if it accepts AI algorithms.

AI has vital role in rapid analysis of voluminous information. There is a great demand in integration of both edge processing and AI. At present, there is no formal and universally presentable edge intelligence. Many researchers gave analysis report on this issue. Few researchers accepted that the extent of edge should not be limited with AI models and must be extended with cloud. Zhang et al. [25] characterized edge intelligence as the ability factor to make edge to execute AI algorithms.

4 Conclusion

We carefully examined and analyzed the attributes and benefits of intelligence edge computing (IEC), as well as the difficulties and prospective use cases. Driving by prospering of both AI and IoT, there is a rigid need to push the AI from the cloud to the organization edge. To satisfy this pattern, edge computing has been broadly perceived as a promising answer for intense AI applications in less resourced areas. The consolidation of edge computing and AI brings edge intelligence (EI). With its capacity of handling information near end users, which is a significant interest for IoT applications, EC becomes a very appealing option. Though the edge intelligence is in its very initial stage, it attracts more researchers and industries to get into the deep study of that. This paper tries to give the list of desirable investigation openings through a compact and successful effective classification. Clearly, we discussed about Internet of Things, cloud platform, artificial intelligence, and edge computing. Also, we clearly gave the roadmap from cloud to edge intelligence. We believe that the connections given in this paper between IoT and cloud, AI, EC will promote the research to the next level. We attempted to give some illuminating musings on the

arising field of edge intelligence. We trust that this article can invigorate productive debates on potential future research directions for edge intelligence.

References

1. Ming FX, Habeeb RAA, Md Nasaruddin FHB, Gani AB (2019) Real-time carbon dioxide monitoring based on iot & cloud technologies. In: Proceedings of the 2019 8th international conference on software and computer applications. Cairo, pp 517–521
2. Moin S, Karim A, Safdar Z, Safdar K, Ahmed E, Imran M (2019) Securing IoTs in distributed blockchain: analysis, requirements and open issues. *Futur Gener Comput Syst* 100:325–343
3. Chu F, Yuan S, Peng Z (2006) Using machine learning techniques to identify botnet traffic. In: Encyclopedia of structural health monitoring. Wiley, Hoboken, NJ, pp 967–974
4. Xiao Y, Jia Y, Liu C, Cheng X, Yu J, Lv W (2019) Edge computing security: state of the art and challenges. *Proc IEEE* 107(8):1608–1631
5. Shi W, Dustdar S (2016) *e promise of edge computing. *Computer* 49(5):78–81
6. Xia X, Chen F, He Q, Grundy J, Abdelrazek M, Jin H (2020) Cost-effective app data distribution in edge computing. *IEEE Trans Parallel Distrib Syst* 32(1):31–44
7. Manadhata PK, Wing JM (2011) A formal model for a system's attack surface. In: Moving target defense. Springer, Berlin, pp 1–28
8. Lai P, He Q, Abdelrazek M et al (2018) Optimal edge user allocation in edge computing with variable sized vector bin packing. In: Proceedings of the international conference on service-oriented computing. Springer, Hangzhou, pp 230–245
9. Lin J, Yu W, Zhang N, Yang X, Zhang H, Zhao W (2017) A survey on internet of things: architecture, enabling technologies, security and privacy, and applications. *IEEE Internet Things J* 4(5):1125–1142
10. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge computing: vision and challenges. *IEEE IoT J* 3(5):637–646
11. Amanullah MA, Habeeb RAA, Nasaruddin FH et al (2020) Deep learning and big data technologies for IoT security. *Comput Commun* 151:495–517
12. Chi X, Yan C, Wang H, Rafique W, Qi L (2020) Amplified locality-sensitive hashing-based recommender systems with privacy protection. *Concurr Comput Pract Exp*
13. Zhong W, Yin X, Zhang X et al (2020) Multi-dimensional quality-driven service recommendation with privacy-preservation in mobile edge environment. *Comput Commun* 157:116–123
14. Song G, Chai W (2018) Collaborative learning for deep neural networks. In: Proceedings of the advances in neural information processing systems. Montreal, pp 1832–1841
15. Khattak HA, Shah MA, Khan S, Ali I, Imran M (2019) Perception layer security in internet of things. *Futur Gener Comput Syst* 100:144–164
16. Hossain MM, Fotouhi M, Hasan R (2015) Towards an analysis of security issues, challenges, and open problems in the internet of things. In: Proceedings of the IEEE world congress on services. IEEE, New York, NY, pp 21–28
17. Li L, Goh T-T, Jin D (2020) How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. *Neural Comput Appl* 32(9):4387–4415
18. Xu X, Zhang X, Liu X, Jiang J, Qi L, Bhuiyan MZA (2020) Adaptive computation offloading with edge for 5G-envisioned internet of connected vehicles. *IEEE Trans Intel Transp Syst* 1–10
19. Xu X, Liu X, Xu Z, Dai F, Zhang X, Qi L (2019) Trustoriente IoT service placement for smart cities in edge computing. *IEEE Internet Things J* 7(5):4084–4091
20. Tahsien SM, Karimipour H, Spachos P (2020) Machine learning based solutions for security of internet of things (IoT): a survey. *J Netw Comput Appl* 161, Article ID 102630
21. Liang F, Hatcher WG, Liao W, Gao W, Yu W (2019) Machine learning for security and the internet of things: the good, the bad, and the ugly. *IEEE Access* 7:158126–158147

22. Lavassani M, Forsström S, Jennehag U, Zhang T (2018) Combining fog computing with sensor mote machine learning for industrial IoT. *Sensors* 18(5):1532
23. Sodhro AH, Pirbhulal S (2018) Power control algorithms for media transmission in remote healthcare systems. *IEEE Access* 6(1):42384–42393
24. Oyekanlu E, Scoles K (2018) Towards low-cost, “real-time, distributed signal and data processing for artificial intelligence applications at edges of large industrial and internet networks”. In: 2018 IEEE first international conference on artificial intelligence and knowledge engineering, pp 66–167
25. Zhang W, Zhang Z, Zeadally S, Chao H, Leung V (2019) MASM: a multiple-algorithm service model for energy-delay optimization in edge artificial intelligence. *IEEE Trans Ind Inform* (1):1–10
26. Shi W, Zhang X, Wang Y, Zhang Q (2019) Edge computing: state-of-the-art and future directions. *J Comput Res Devel* 56(1):69–89
27. Chen Y, Zhang N, Zhang Y, Chen X, Wu W, Shen XS (2019) TOFFEE: task offloading and frequency scaling for energy efficiency of mobile devices in mobile edge computing. *IEEE Trans Cloud Comput* 1
28. Wang L, Zhang X, Wang R, Yan C, Kou H, Qi L (2020) Diversified service recommendation with high accuracy and efficiency. *Knowl Syst* 204, Article ID 106196
29. Ziming Z, Fang L, Zhiping C, Nong X (2018) Edge computing: platforms, applications and challenges. *J Comput Res Devel* 55(2):327
30. Catarinucci L, De Donno D, Mainetti L, Palano L, Patrono L, Stefanizzi ML, Tarricone L (2015) An IoT-aware architecture for smart healthcare systems. *IEEE Internet Things J* 2:515–526
31. Alrawais A, Alhothaily A, Hu C, Cheng X (2017) Fog computing for the internet of things: security and privacy issues. *IEEE Internet Comput* 21:34–42
32. El-Sayed H, Sankar S, Prasad M, Puthal D, Gupta A, Mohanty M, Lin CT (2017) 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
33. Cao Y, Hou P, Brown D, Wang J, Chen S (2015) Distributed analytics and edge intelligence: pervasive health monitoring at the era of fog computing. In: Proceedings of the 2015 workshop on mobile big data, Hangzhou, 22–25 June 2015, pp 43–48
34. Hassan N, Gillani S, Ahmed E, Yaqoob I, Imran M (2018) The role of edge computing in internet of things. *IEEE Commun Mag* 56:110–115
35. Plastiras G, Terzi M, Kyrou C, Theocharides T (2018) Edge intelligence: challenges and opportunities of near-sensor machine learning applications. In: Proceedings of the 2018 IEEE 29th international conference on application-specific systems, architectures and processors (ASAP), Milano, 10–12 July 2018, pp 1–7
36. Li H, Ota K, Dong M (2018) Learning IoT in edge: deep learning for the internet of things with edge computing. *IEEE Netw* 32:96–101

Correction to: Micro-Electronics and Telecommunication Engineering



Devendra Kumar Sharma, Sheng-Lung Peng, Rohit Sharma,
and Dmitry A. Zaitsev

Correction to:

**D. K. Sharma et al. (eds.), *Micro-Electronics
and Telecommunication Engineering, Lecture Notes
in Networks and Systems 373,***
<https://doi.org/10.1007/978-981-16-8721-1>

In the original version of the book, the following belated corrections were incorporated:

In chapter 11, the affiliation “School of Engineering and Technology (SET), Sharda University, Greater Noida, India” of the author “Jaya Srivastava” has been changed to: Noida Institute of Engineering Technology (NIET), Greater Noida, India.

In chapter 29, the affiliations “Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India” and “SRM Institute of Science and Technology, Chennai, India” of author “G.Elavel Visuvanathan” and “T. Jaya” respectively have been changed to G. Elavel Visuvanathan, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India and T. Jaya, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India.

The chapter and book have been updated with the changes.

The updated versions of these chapters can be found at
https://doi.org/10.1007/978-981-16-8721-1_11
https://doi.org/10.1007/978-981-16-8721-1_29

Author Index

A

- Abbas, Amel H., 373
Abdullah, Dhuha Basheer, 405
Abirami, M. S., 269
Agarwal, Veral, 467
Ahmed, Sajjad, 661
Akanksha, 173
Akhil Jaichandra Reddy, B., 1
Akhtar, Faaiz, 257
Alaabedi, Yasir Abdulzahra Flaiyh, 383
Albaker, Baraa M., 415
Ali, Sazid, 149
Altufaili, Mohammed Mahdi Salih, 383
Al-Turfi, Mohammed N., 415
Anand, Rishabh, 47
Anand, Santosh, 593
Ansari, M. A., 17, 25, 37, 47
Anusha, M., 581
Anusha, R., 141
Arthi, R., 241
Arya, Gautam, 685
Ashish, Shivam, 729

B

- Babu, Jagadish, 221
Badoni, Pankaj, 561
Bagyammal, T., 647
Bala, Indu, 91, 97
Banerjee, Sourav, 183
Bansal, Nipun, 703
Basheer, Ayman, 439
Bhardwaj, Abhishek, 349
Bhardwaj, Priyanka, 477, 485
Bhatt, Chayan, 609

Bhowmik, Trisha, 349

- Bhushan, Bharat, 107, 119, 141, 309
Bindra, Mahin, 703
Bisht, Ankit, 517
Biswas, Utpal, 183
Britto Corthis, P., 817

C

- Chakraborty, Partha, 497
Chaudhary, Chirag, 747
Chauhan, Surendra Singh, 685
Chhabra, Manish, 541
Choudhary, Surya Deo, 533
Choudhury, Tanupriya, 517, 525
Chowdhury, Aninda, 149

D

- Dalip, 507
Das, Anwesha, 149
Das, Debashis, 183
Deepa J., 141
Deepika, 507
Desai, Nirav, 669
Devi, Bingi Manorama, 331
Dominic, Deepica MS.S., 729

E

- Elavel Visuvanathan, G., 297

F

- Ferman, Vian Adnan, 393

G

- Gayathri, M., 193
 Gomathy, C., 193
 Gumar, Hiba A., 415
 Gupta Das, Ritwika, 339
 Gupta, Keshav, 573
 Gupta, Rohan, 729
 Gupta, Sabhav, 573

H

- Hasan, Balqees Talal, 405
 Hassan, Hassan Jaleel, 439

I

- Islam, Saiful, 661

J

- Jayashree, J., 141
 Jaya, T., 297
 Joseph, Niju P., 331, 339
 Joshi, Sujata, 221

K

- Kapileswar, Nellor, 1
 Kariappa, Hruthik, 593
 Karn, Ashish, 561
 Karnwal, Nitin, 37
 Karthika, M., 581
 Katiyar, Neha, 775
 Kaviya, M., 289
 Kerur, S. S., 159
 Khant, Shailesh, 627
 Kiruthika, V., 281, 289
 Kishor, Kaushal, 541
 Komal, M., 81
 Kousalya, G., 61
 Krishna, Santhosh B. V., 251
 Kumar, Abhijit, 107
 Kumar, Avinash, 119, 349
 Kumar, Harish, 221
 Kumari, Reshma, 533
 Kumar, Naresh, 207
 Kumar, Prasanna R., 231
 Kumar, Rajneesh, 507
 Kurni, Muralidhar, 331
 Kushwaha, Manoj, 269

L

- Lakshmy, K. V., 717

M

- Madhurima, V., 759
 Mahipal, 257
 Malik, Ayasha, 107
 Malla, Satya yana V., 525
 Mandal, Danvir, 91, 97
 Manimegalai, C. T., 241
 Mani, Prashant, 457
 Manoj, K., 593
 Mohan, Bharathi G., 231
 Mousa, Hussein Ali, 383
 Msekhh, Alyaa A., 363
 Msekhh, Zahraa A., 363
 Mukherjee, Ritam, 149
 Muralidharan, Roshan, 593
 Muttasher, Gaida, 439

N

- Nagar, Piyush, 685
 Naik, Akruti, 669
 Nand, Parma, 119
 Naruka, Shivam, 573
 Nataraj, Neetha, 251
 Nihaz, Mahammad, 221

O

- Obaid, Ahmed J., 425
 Omer, Zaid T., 373

P

- Padmapriya, K., 759
 Padmavathi, B., 1
 Pandey, Adarsh, 573
 Pandey, Krishna Mohan, 207
 Pandey, Sachin Kumar, 131
 Pandey, Sneha, 533
 Panguluri, Sai Deepika, 717
 Paramasivam, C., 131
 Parameswaran, Latha, 647
 Parmar, Mohitsinh, 627
 Patel, Atul, 627
 Patel, Rachit, 467
 Phani Kumar, Polasi, 1
 Prajapati, Yogendra Narayan, 207

R

- Rajasekaran, S., 61
 Raju, Akhil, 221
 Ramesh, G. P., 805, 817
 Ramyapriyanandhini, G., 647

Reji, V., 241
Roy, Manidipa, 477, 485

S

Sabah, Noor, 447
Salimath, Nivedita, 251
Samanta, Debabrata, 331, 339
Sambhav, Saurabh, 553
Saritha, K., 331
Saroja, V. S., 81
Satpathy, Rudra Bhanu, 805
Shamim, Firdous, 149
Sharma, Bhavya, 703
Sharma, Charu, 323
Sharma, Hitesh Kumar, 517, 525
Sharma, Rahul, 541
Sharma, Rohit, 149, 349
Shekhar, Himanshu, 729
Shreyank, K., 81
Shreyas, C., 593
Sidhu, Jagroop Singh, 71
Sil, Riya, 149
Singhal, Ankur, 97
Singhal, Sunita, 609
Singh, Amit, 747
Singh, Kushall Pal, 775
Singh, Mukul, 37
Singh, Nidhi, 37
Singh, Nivedita, 17
Singh, Sanjay Kumar, 477, 485
Singh, Shilpi, 553
Singh, Vivek Pratap, 17
Sonker, Devesh, 173
Sowmya, L., 251
Sowmya, N., 81
Srinivasan, Chungath, 717
Srivastava, Jaya, 107
Srivastava, Jyoti, 775
Srivastava, Nishant, 457
Suhas, S., 81

Sultana, Sajeda, 497
Sumathi, R., 795
Swarnima, 533

T

Tanya, Reenie, 257
Tawfeeq, Mohammed Ali, 393
Thaker, Hetal, 669
Thakur, Suyash, 561
Tiwari, Manu, 685
Tyagi, Nipun, 467

U

Uttam Reddy, N., 1

V

Vaiapury, Karthikeyan, 647
Vaid, Rohit, 323
Vallikannu, A. L., 281
Vallikannu, R., 289
Vasudevan, V., 795
Velangi, Radha, 159
Verma, Rishabh, 25
Vijayashree J., 141
Vimalarani, G., 281
Vishal, 747

W

Walia, Gurjot Kaur, 71

Y

Yadav, Ranjeeta, 173
Yadav, Shivani, 251
Yousuff, Mohamed, 141
Yukta, K., 81