

V. V. S. S. S. Chakravarthy
Wendy Flores-Fuentes
Vikrant Bhateja
B. N. Biswal *Editors*

Advances in Micro-Electronics, Embedded Systems and IoT

Proceedings of Sixth International
Conference on Microelectronics,
Electromagnetics and
Telecommunications (ICMEET 2021),
Volume 1

Lecture Notes in Electrical Engineering

Volume 838

Series Editors

Leopoldo Angrisani, Department of Electrical and Information Technologies Engineering, University of Napoli Federico II, Naples, Italy

Marco Arteaga, Departament de Control y Robótica, Universidad Nacional Autónoma de México, Coyoacán, Mexico

Bijaya Ketan Panigrahi, Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi, India
Samarjit Chakraborty, Fakultät für Elektrotechnik und Informationstechnik, TU München, Munich, Germany

Jiming Chen, Zhejiang University, Hangzhou, Zhejiang, China

Shanben Chen, Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

Tan Kay Chen, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore

Rüdiger Dillmann, Humanoids and Intelligent Systems Laboratory, Karlsruhe Institute for Technology, Karlsruhe, Germany

Haibin Duan, Beijing University of Aeronautics and Astronautics, Beijing, China

Gianluigi Ferrari, Università di Parma, Parma, Italy

Manuel Ferre, Centre for Automation and Robotics CAR (UPM-CSIC), Universidad Politécnica de Madrid, Madrid, Spain

Sandra Hirche, Department of Electrical Engineering and Information Science, Technische Universität München, Munich, Germany

Faryar Jabbari, Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA

Limin Jia, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Alaa Khamis, German University in Egypt El Tagamoa El Khames, New Cairo City, Egypt

Torsten Kroeger, Stanford University, Stanford, CA, USA

Yong Li, Hunan University, Changsha, Hunan, China

Qilian Liang, Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA
Ferran Martín, Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Tan Cher Ming, College of Engineering, Nanyang Technological University, Singapore, Singapore

Wolfgang Minker, Institute of Information Technology, University of Ulm, Ulm, Germany

Pradeep Misra, Department of Electrical Engineering, Wright State University, Dayton, OH, USA

Sebastian Möller, Quality and Usability Laboratory, TU Berlin, Berlin, Germany

Subhas Mukhopadhyay, School of Engineering & Advanced Technology, Massey University, Palmerston North, Manawatu-Wanganui, New Zealand

Cun-Zheng Ning, Electrical Engineering, Arizona State University, Tempe, AZ, USA

Toyoaki Nishida, Graduate School of Informatics, Kyoto University, Kyoto, Japan

Federica Pascucci, Dipartimento di Ingegneria, Università degli Studi “Roma Tre”, Rome, Italy

Yong Qin, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Gan Woon Seng, School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Joachim Speidel, Institute of Telecommunications, Universität Stuttgart, Stuttgart, Germany

Germano Veiga, Campus da FEUP, INESC Porto, Porto, Portugal

Haitao Wu, Academy of Opto-electronics, Chinese Academy of Sciences, Beijing, China

Walter Zamboni, DIEM - Università degli studi di Salerno, Fisciano, Salerno, Italy

Junjie James Zhang, Charlotte, NC, USA

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases.****

More information about this series at <https://link.springer.com/bookseries/7818>

V. V. S. S. S. Chakravarthy ·
Wendy Flores-Fuentes · Vikrant Bhateja ·
B. N. Biswal
Editors

Advances in Micro-Electronics, Embedded Systems and IoT

Proceedings of Sixth International Conference
on Microelectronics, Electromagnetics
and Telecommunications (ICMEET 2021),
Volume 1



Springer

Editors

V. V. S. S. Chakravarthy
Department of Electronics
and Communication Engineering
Raghu Institute of Technology
Visakhapatnam, Andhra Pradesh, India

Vikrant Bhateja
Department of Electronics
and Communication Engineering
Shri Ramswaroop Memorial College
of Engineering and Management
(SRMCEM)
Lucknow, Uttar Pradesh, India

Dr. A.P.J. Abdul Kalam Technical
University
Lucknow, Uttar Pradesh, India

Wendy Flores-Fuentes
Autonomous University of Baja California
Mexicali, Baja California, Mexico

B. N. Biswal
Bhubaneswar Engineering College
Bhubaneswar, Odisha, India

ISSN 1876-1100

Lecture Notes in Electrical Engineering

ISBN 978-981-16-8549-1

ISBN 978-981-16-8550-7 (eBook)

<https://doi.org/10.1007/978-981-16-8550-7>

ISSN 1876-1119 (electronic)

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

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Organizing Committee

Chief Patron

Er. Prabhat Ranjan Mallick, Chairman KGI and Founder of BEC

Co-chief Patron

Er. Alok Ranjan Mallick, Honorable Chairman, BEC

Convener

Prof. (Dr.) B. N. Biswal, Director (A&A), BEC, BBSR

General Chairs

Prof. Jaume Anguera, Universitat Ramon Llull, Spain

Dr. Suresh Chandra Satapathy, School of Computer Engineering, Kalinga Institute of Industrial Technology (Deemed to be University) Bhubaneswar, Odisha, India

Dr. Vikrant Bhateja, Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM), Lucknow, Uttar Pradesh, India

Program Chairs

Prof. P. Satish Rama Chowdary, Department of Electronics and Communication Engineering, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India
Dr. V. V. S. S. Chakravarthy, Department of Electronics and Communication Engineering Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India

Organizing Committee

Dr. S. Samal (Professor, Aeronautical Engineering, BEC)
Prof. B. P. Mishra (HOD, Mechanical Engineering, BEC)
Prof. P. M. Dash (HOD, EE, BEC BBSR,)
Prof. M. K. Swain (HOD, CSE, BEC, BBSR)
Prof. S. J. Kar (CSE Department), BEC
Prof. G. S. Behera (EE Department), BEC
Prof. D. Panda (H&Sc Department), BEC
Dr. N. Chatterjee (H&Sc Department), BEC
Dr. U. Panigrahi (Mechanical Department), BEC
Dr. S. Sahoo (EE Department), BEC

Technical Advisory Committee

Prof. Xin-She Yang, Middlesex University, London, UK
Prof. Jaume Anguera, Universitat Ramon Llull, Spain
Prof. Nihad I. Dib, Jordan University of Science and Technology, Irbid, Jordan
Prof. Ganapati Panda, IIT Bhubaneswar, Odisha, India
Dr. Steven L. Fernandes, Creighton University, USA
Col. Prof. Dr. G. S. N. Raju, Vice Chancellor, CUTMAP
Dr. Seyedali Mirjalili, Griffith University, Queensland, Australia
Dr. Aurora Andujar, Fractus, Spain
Prof. R. K. Mishra, Berhampur University, Odisha, India
Dr. T. Srinivas, IISC, Bengaluru
Prof. N. Bheema Rao, NIT Warangal
Prof. G. Sasibhushana Rao, Andhra University, Visakhapatnam
Dr. Babita Majhi, GG Central University, Bilaspur
Mr. V. Balaji, Head-Wireless Deployment, Reliance Corporate Park, Navi Mumbai
Dr. Lakshminarayana Sadasivuni, Andhra University (Retd.), Visakhapatnam
Prof. P. Rajesh Kumar, Andhra University, Visakhapatnam
Prof. A. Mallikarjuna Prasad, JNTUK
Prof. P. Mallikarjuna Rao, Andhra University, Visakhapatnam

Dr. Ch. Srinivasa Rao, JNTUK Vizianagaram

Dr. Tummala Surya Narayana Murthy, UCEV, JNTUK Vizianagaram

Dr. Tumati Venkateswara Rao, Sir C. R. Reddy College of Engineering, Eluru

Dr. Sudheer Kumar Terlapu, Shri Vishnu Engineering College for Women (A),
Bhimavaram

Dr. Sonali Dash, Raghu Institute of Technology

Dr. M. Vamsi Krishna, Professor, Dhanekula Institute of Engineering and Technology

Prof. Somya Goyal, Manipal University Jaipur, Jaipur, Rajasthan

Preface

This book is a compilation of research papers presented in the Sixth International Conference on Microelectronics, Electromagnetics, and Telecommunications (ICMEET 2021) which was held at Bhubaneswar Engineering College, Bhubaneswar, Odisha, India, during August 27–28, 2021. Prior to this, five versions of the conference were organized consecutively from 2015 to 2019. Every time, the proceedings are published in Lecture Notes in Electrical Engineering Series of Springer.

The 6th ICMEET is an international colloquium, which aims to bring together academic scientists, researchers, and scholars to discuss the recent developments and future trends in the fields of microelectronics, electromagnetics, and telecommunication. Microelectronics research investigates semiconductor materials and device physics for developing electronic devices and integrated circuits with data/energy efficient performance in terms of speed, power consumption, and functionality. The books published so far with the title ICMEET in LNEE discussed various topics like analog, digital and mixed signal circuits, bio-medical circuits and systems, RF circuit design, microwave and millimeter wave circuits, green circuits and systems, analog and digital signal processing, nanoelectronics and giga scale systems, VLSI circuits and systems, SoC and NoC, MEMS and NEMS, VLSI digital signal processing, wireless communications, cognitive radio, and data communication.

ICMEET received 321 submissions. Each paper was peer-reviewed by at least two well-qualified reviewers and the members of the programme committee. Finally, 101 papers were accepted for publication out of which 51 papers are included in this volume titled *Advances in Micro-Electronics, Embedded Systems and IoT*—which is a collection of manuscripts with theme focusing on integrated circuits, microelectronics, VLSI, embedded systems, optimization and metaheuristics, IoT, and automation.

Several special sessions were offered by eminent professors on many cutting-edge technologies. The conference also featured two keynotes by Prof. (Dr.) Ran Cheng, Southern University of Science and Technology, China, with title “When Evolutionary Computation Meets Big Data: Challenges and Opportunities” and by Dr. Kamlesh Verma, Scientist, Instruments Research and Development Establishment

(IRDE), DRDO, Ministry of Defense, Govt. of India, with title “Visual Tracking in Un-stabilized Videos.” The inaugural of the conference was graced by eminent people like Prof. (Dr.) Ran Cheng, Southern University of Science and Technology, and Prof. (Dr.) Naeem Hanoon, University of Teknologi, Malaysia, as Guests of Honor along with Prof. (Dr.) A. K. Nayak, Advisor—IIBM Group of Institutions, Former President and Fellow CSI as Chief Guest.

We would like to thank Chief Patron Er. Prabhat Ranjan Mallick, Chairman KGI and Founder of BEC and Co-chief Patron and Er. Alok Ranjan Mallick Honorable Chairman, BEC, for their commitment and dedication to host the conference in their organization. Also, we wish to extend our gratitude to Convener, Prof. (Dr.) B. N. Biswal, Director (A&A), BEC, BBSR, for his coordination through the success of the conference. We congratulate all the heads of the departments and the staff members of Bhubaneswar Engineering College, Bhubaneswar, for their efforts in coordinating the proceedings of the program.

Special thanks to all session chairs, track managers, and reviewers for their excellent support. Last but certainly not least, our sincere thanks go to all the authors who submitted papers and all the presenters for their contributions and fruitful discussions that made this conference a great success.

Visakhapatnam, India

Mexicali, Mexico

Lucknow, India

Bhubaneswar, India

V. V. S. S. S. Chakravarthy

Wendy Flores-Fuentes

Vikrant Bhateja

B. N. Biswal

Contents

Synthesis of KNN Algorithm in FPGA Technology	1
Ravi Payal and Amit Prakash Singh	
Study of Architecture and Performance Analysis of Approximation Multipliers	11
E. Jagadeeswara Rao, K. Murali Krishna, Md. Khasim, and T. Anjaiah	
Design and Analysis of Energy-Efficient Logic Gates Using INDEP Short Gate FinFETs at 10 nm Technology Node	19
Umayia Mushtaq, Md. Waseem Akram, and Dinesh Prasad	
SIMPLE-DRR: A New Energy-Efficient Multi-hop Routing Protocol in WBANs for Health Monitoring	29
Subba Reddy Chavva and Ravi Sankar Sangam	
Design of Class AB and Class C Amplifiers	41
Tejendra Kumar Singh, Deepti Kakkar, and Sukhwinder Singh	
LightGBM Model for Credit Card Fraud Discovery	51
Appala Srinuvasu Muttipati, Sangeeta Viswanadham, Radha Dharavathu, and Jayalakshmi Nema	
PV System-Based Switched Capacitor DC–DC Converter for High Voltage Gain Using Fuzzy-PID Controller	59
Balachandra Pattanaik and Yohannes Bekuma Bakare	
Start-up Circuits for Zero Current Reduction in Bandgap Reference Circuits	67
Anushree, Anushka Dixit, Nandini Singh, and Jasdeep Kaur	
Recycle of Plastic Material for Fabrication of Paver Blocks	79
K. Ananthi, S. Priyadharsini, A. Ijas Ahamed, B. Viney Rangasamy, and R. Prabhu Surya	

Handwritten Cursive English Character Recognition Using DAG-CNN	89
P. V. Bhagyasree, Ajay James, N. D. Bisna, and K. S. Vipin Kumar	
Automation of Security Scanning Tools Using K8's Environment for Commercial Telecom Product	103
Pooja and K. S. Shushrutha	
Rad-Hard Model SOI FinTFET for Spacecraft Application	113
Ajay Kumar Dharmireddy, Sreenivasa Rao Ijjada, K. V. Gayathri, K. Srilatha, K. Sahithi, M. Sushma, and K. Madhavi	
Direct Synthesis Method-Based PID Controller for Critically Damped Time Delay Systems	121
E. Govinda Kumar, R. Vinoth Kumar, and D. Deivasigamani	
Exploiting Full-Duplex Relaying in Vehicular Cooperative NOMA for Residual Self-interference in Amplify and Forward and Decode and Forward with Incremental Relaying	131
Sravani Potula, Sreenivasa Rao Ijjada, and Karunakar Reddy Santhamgari	
Lower Order Harmonic Torque Reduction in v/f Controlled Induction Motor Drive	143
Shivam Yadav, Anshul Kumar Mishra, Jayant Mani Tripathi, and Aseem Chandel	
Attention-Based Comparison of Automatic Image Caption Generation Encoders	157
Cheboyina Sindhu NagaDurga and T. Anuradha	
Implementation of Image Denoising with Reversible Multipliers Using Xilinx System Generator	169
Bandlapalli Sudharani and Gunapati Sreenivasulu	
Development of an Automatic Fruit Classification Using Hybrid Deep Learning Model for Super Markets	181
B. Sridhar, K. Sitharam Sai Kiran, N. Likhitha, K. P. V. A. Vardhan, and K. V. Nikhil	
Design a Low-Cost Air Pollution Monitoring IoT System	191
B. Sridhar, R. Mounika, P. Nagendra Babu, Y. Pavan Kumar, and S. Y. Ravi Raja	
Discrete Wavelet Transform-Based Fusion of Mammogram Images for Contrast Improvement	203
Anushka Dikshit, Vikrant Bhateja, and Ashruti Rai	

Assessment of Land Degradation Dynamics Using Spectral Angle Mapper Method and Demographic Analytics	211
Amar Kumar Kathwas and Rakesh Saur	
Modelling Flash Flood Vulnerability and Sensitivity Dynamics of Jiadhal River Basin of Eastern Himalayan Range Using Space Technology and AHP	225
Rakesh Saur and Virendra Singh Rathore	
Automating Vehicle Insurance Process Using Smart Contract and Ethereum	237
Saroj Kumar Nanda, Sandeep Kumar Panda, Madhabananda Das, and Suresh Chandra Satapathy	
Energy-Aware Load Balancing in Dynamic Cloud Environment Using Nature-Inspired Technique	249
Bivasa Ranjan Parida and Amiya Kumar Rath	
Analyzing Intake of Water by a Human Using Meteorological Conditions	259
M. Rekha Madhuri, T. Sriya Datta, M. Nithisha, and Y. Sandeep	
Driver Drowsiness Detection and Object Detection for Vehicles Using Yolov3	273
D. Sampath, V. V. Krishna Chaitanya, B. Dhanvanth Kumar, and G. Geetha	
Traffic Management System Using AI and IoT	285
Priya Geda and Sumitra Motade	
Autonomous UAV with Human Detection	297
Kshitij Karnawat, Nihar Asare, Sumit Anilkumar Singh, and Anju Kulkarni	
Team Sports Result Prediction Using Machine Learning and IoT	305
Lydia D. Isaac and I. Janani	
A Hybrid SVM–ABC Model for Monthly Stream Flow Forecasting	315
Ujjawal K. Singh, Baidyanath Kumar, Nabin K. Gantayet, Abinash Sahoo, Sandeep Samantaray, and Nihar Ranjan Mohanta	
Optimizing Power Saving in Cloud Computing Environments Through Server Consolidation	325
Parthasarathi Pattnayak	
Siamese Network with Transfer Learning for Similar Query Retrieval in Online Health Community Forums	337
B. Athira and Sumam Mary Idicula	

Emergency Bot in Healthcare Using Industry 4.0	347
Sahil Sobhani, Sejal Balasaheb Shirale, Sajal Saxena, Vartika Paharia, and Somya Goyal	
Intelligent Traffic Management System Using Industry 4.0	357
Aman Panwar, Saagar Bafna, Arjun Raghav, and Somya Goyal	
ELPR: Energy Efficient Load Balanced PRoPHET Routing for Delay Tolerant Networks	365
Amit Malik, Amita Rani, and Milan Kumar	
LSTM-Based IDS System for Security of IoT	377
Ajay Kumar and Amita Rani	
Securing Smart Homes Using Face Recognition	391
Mehul Sinha, Raj Chaurasiya, Arav Pandey, Yuvraj Singh, and Somya Goyal	
Smart Quiz for Brain Stormers	399
Aditya Kumar, Rishabh Gupta, Naivedhya Sharma, and Somya Goyal	
Crash Detection-Based Fleet Tracking System Using VANETs	407
Parveen, RishiPal Singh, and Sushil Kumar	
FPGA Implementation of 16-Bit Wallace Multiplier Using HCA	419
G. Challa Ram, M. Venkata Subbarao, D. Girish Kumar, and Sudheer Kumar Terlapu	
A Comprehensive Study on Fruit Odour Detection and Classification Techniques Using eNose	429
Kalidindi Lakshmi Divya and V. Vijaya Baskar	
Multivariate Time Series Analysis for Predicting Number of Disease Cases	437
Ramatulasi Tammineni and G. N. V. G. Sirisha	
Difference Differential Op-Amp Based CMOS Adiabatic All Pass Filter	447
P. Mullangi and John Blessy Mandru	
Parameter Boosted Approach to Collaborative Filtering Based Recommender System	457
Swasti Singhal, Nilotpal Pathak, and Bhawna Singh	
Improving Survival Rate by Estimating the Progression of Pulmonary Fibrosis	465
Jyothi Gutala, Navya Sri Kalepalli, Madhuri Rudrapati, and G. Kalyani	
Analytical Variable Execution of GDI Vedic Multiplier Using FinFET Full Adder	475
S. K. Dilshad and G. Sai Krishna Santosh	

Contents	xv
Classified Authentication System with IoT and Dashboard	487
Chilakala Sudhamani, K. Bharath Kumar, G. V. Hari Prasad, and N. Renuka	
Waste Management Data Analytics and Solution for Domestic Waste Management to Improve Soil Quality	497
Suwarna Gothane, K. Srujan Raju, and B. Kavitha Rani	
Low-Cost Irrigation and Laser Fencing Surveillance System for the Paddy Fields	507
Abdul Subhani Shaik, Manir Ahmed, and Merugu Suresh	
Hybridization of Modified Cuckoo-Moth Flame Optimization for Effective Route Recovery of Networks	517
Satish Kumar Patnala, Venubabu Rachapudi, S Anjali Devi, Sasibhushana Rao Pappu, and T. Subha Mastan Rao	
Analysis of Password Protected Documents Using Statistical Approaches on High Performance Computing	533
Ajeet Singh, Vikas Tiwari, Allu Swamy Naidu, Appala Naidu Tentu, K. Surjan Raju, and Ashutosh Saxena	
Author Index	547

About the Editors

V. V. S. S. S. Chakravarthy is a professor in the Department of Electronics and Communication Engineering at Raghu Institute of Technology, Visakhapatnam. He is a senior member of Communication, Signal Processing and Antenna and Propagation Societies of IEEE. He is serving as a vice chair of IEEE COMSOC/SPS Joint Chapters of IEEE Vizag Bay Section. His research interests include computational intelligence, smart antenna, data modelling, machine learning and evolutionary computing tools. He has 18 years of teaching. He served as a co-editor to proceedings of third and fifth International Conference on Microelectronics, Electromagnetics and Telecommunications published in Lecture Notes in Electrical Engineering. He is also a life member of professional bodies like Instrumentation Society of India, International Computer Science and Engineering Society (ICSES) and Soft Computing Research Society. He published more than 40 journal and conference papers along with one Book Chapter which are indexed in SCOPUS and SCI.

Wendy Flores-Fuentes received the master's degree in engineering from Technological Institute of Mexicali in 2006, and the Ph.D. degree in science, applied physics, with emphasis on optoelectronic scanning systems for SHM, from Autonomous University of Baja California in June 2014. She has more than 115 publications which includes journal articles in Elsevier, IEEE Emerald and Springer, book chapters and books in Intech, IGI global and Springer, proceedings articles in IEEE. She has been a panel reviewer of Taylor and Francis, IEEE, Elsevier, and EEMJ (Gh. Asachi Technical University of Iasi). Currently, she is a full-time professor-researcher at Universidad Autónoma de Baja California, at the Faculty of Engineering.

Dr. Vikrant Bhateja is associate professor in Department of Electronics & Communication Engineering (ECE), Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM), Lucknow (Affiliated to AKTU) and also the Dean (Academics) in the same college. He is doctorate in ECE (Bio-Medical Imaging) with a total academic teaching experience of 19 years with around 180 publications in reputed international conferences, journals and online book chapter contributions; out of which 31 papers are published in SCIE indexed high impact factored journals. Among the international conference publications, four papers have received “Best

Paper Award". Among the SCIE publications, one paper published in Review of Scientific Instruments (RSI) Journal (under American International Publishers) has been selected as "Editor Choice Paper of the Issue" in 2016. He has been instrumental in chairing/co-chairing around 25 international conferences in India and abroad as Publication/TPC chair and edited 45 book volumes from Springer-Nature as a corresponding/co-editor/author on date. He has delivered nearly 20 keynotes, invited talks in international conferences, ATAL, TEQIP and other AICTE sponsored FDPs and STTPs. He is Editor-in-Chief of IGI Global—International Journal of Natural Computing and Research (IJNCR) an ACM & DBLP indexed journal since 2017. He has guest edited Special Issues in reputed SCIE indexed journals under Springer Nature and Elsevier. He is Senior Member of IEEE and Life Member of CSI.

B. N. Biswal is currently working as a director (Academics) in Bhubaneswar Engineering College, Bhubaneswar, India. He has about 24 years of experience in teaching and research. He has completed his B.Tech. in Electronics Engineering from Baba Saheb Ambedkar Marathwada university, M.Tech in Communication System Engineering from VSSUT Burla and Ph.D. in Computer Science From Utkal University, Vanivihar Bhubaneswar. He is a member of IEEE and life member of ISTE, IAENG and CSI. He is the editorial member of Springer AISC-199, 247, 327, 328. He has been delivering lectures to the students of UG and PG in the areas such as analog/digital electronics and circuits, microprocessor and microcontrollers, computer organization and advanced architecture, digital communication technique and computer networking, digital signal processing and VLSI design, embedded system and computational intelligence. He is a good organizer holding national and international conferences and workshops. He has published many papers in national/international conferences, journals and books.

Synthesis of KNN Algorithm in FPGA Technology



Ravi Payal and Amit Prakash Singh

Abstract Machine learning is coming everywhere, and currently in every field, it is contributing in terms of different applications. Machine learning is a type of artificial intelligence (AI), which enables a program to learn, without being explicitly programmed and without human intervention. It is broadly divided into two types, namely supervised and unsupervised learning. Different types of Machine learning algorithms are there but slowly and slowly wrt hardware design also it is coming. KNN algorithm is one of the most important algorithms, and it is typically used for pattern recognition, data mining, image recognition, text categorization, and predictive analysis. Here, we tried to implement KNN algorithm by using HDL. It is divided into two parts. First one is to calculating distance between testing point and training datasheet points, and the second part consists of sorting the distances calculated and defining the respective class/group of the testing point. The code is also mapped to FPGA library for the synthesis process.

Keywords KNN · Verilog HDL · Machine learning · Simulation · Synthesis

1 Introduction

Machine learning is an application of artificial intelligence (AI) which provides ability to the systems to automatically learn and improve the performance of systems based on experience. Currently, VLSI EDA industry is trying to evolve hardware related to machine learning algorithms. The progress is slow as most of the problems in EDA industry are statistical in nature. Over the last few years, it has been seen that machine learning applications have become the parts of human life. It is implemented for various tasks which are related to our life and also expected that in a coming year more tasks will be designed. There are lot of algorithms related to machine learning,

R. Payal (✉)
CDAC, Noida, India

A. P. Singh
GGSIPU, New Delhi, India

and one of them is k-nearest neighbor (KNN) algorithm. KNN algorithm is one of the simplest and important algorithms, and it is used for classification and regression problem. In this paper, we present a parallel implementation of KNN algorithms using Verilog HDL with respect to FPGA technology.

The rest of the paper is organized as follows; Sect. 2 shows other work related to KNN implementation. Section 3 describes the KNN algorithm and other VLSI terms used for the implementations. Section 4 presents the KNN block diagram and its explanation, and Sect. 5 describes the results.

2 Previous Work

Lot of work in past was done for KNN algorithm. The work is related to KNN algorithm optimization with respect to any desired application, for memory efficient techniques, neural network, and many more. On [1] paper author talks about a new novel techniques based on WLAN location estimation which is used to calculate the location of persons. Authors of this paper proposed a new novel algorithm called dynamic KNN (D-KNN) which uses AOA, i.e., angle of arrival and KNN as a hybrid method. Authors suggested that D-KNN selects the best nearest neighbors by adaptive antenna system (AAS) for determining the user locative area. Simulation results show that KNN performance is much improved. Paper [2] talks about reducing the effect of constraint related to memory access in K-nearest neighbors (KNN). Authors implemented the KNN on FPGA by using high-level synthesis by using two reduction methods named low precision data representation and principal component analysis-based filtering. Authors through experiment suggested that optimized MPCAF-kNN is better with respect to time and energy efficiency with respect to existing KNN. Coming on next paper, Paper [3] speaks about how traditional HDL-based designs are difficult for accelerating machine learning algorithm in FPGA. Author uses Intel FPGA SDK CAD tool for the experimental purpose. Authors use IFSO for exploring acceleration of KNN algorithm using SRAD simulation using FPGA. The optimized algorithm is mapped on two FPGA named Intel Stratix A7 and Intel Arria 10 GX. In paper related to adaptive implementation of KNN on FPGA [4], authors proposed two adaptive FPGA architectures of KNN classifier and compared their implementation. For the one architecture, the proposed hardware implementation outperformed GPP by 76X and 68X for second. Authors also proposed dynamic partial reconfiguration architecture of KNN classifier. In the paper related to KNN algorithm on FPGA based on heterogeneous computing system based on open CL [5]. Authors propose a solution to speed up KNN algorithm on FPGA. Authors use OpenCL for this purpose. Author designed a bubble sort algorithm for optimizing the KNN algorithm. Authors used both GPU- and FPGA-based architecture for the experiment purpose. The experiment results show that in terms of speed GPU performance is better and for power FPGA is better. In paper VLSI Implementation of Restricted Coulomb Energy Neural Network with Improved Learning Scheme [6], authors propose new modified neural network learning algorithm. Author uses the VLSI design techniques for experiment

implementation. Paper talks about reducing the neurons at the same radius. The algorithm divides each neuron region and measures its performance. Verilog HDL is used for this purpose and synthesized using Synopsys Design Compiler. In paper A Study on Various Data Mining Algorithms Pertaining to VLSI Cell Partitioning [7], authors presented the study of various machine learning algorithms for solving VLSI partitioning problem. The paper compares the various algorithm performances. The experiment result shows that KNN algorithm gives better classification results than SVM. Paper also tells that object function is reduced in the fuzzy C-means algorithm with less number of iterations as compared to KNN. In the paper, two main data mining approaches, namely classification and clustering, are analyzed. VLSI cell partitioning techniques are used in the experiments. The study of various papers suggested that there is a huge possibility of KNN algorithms in various fields, lot of applications can be designed by using KNN and various research applications can be designed. Slowly and slowly KNN algorithm is also coming in the research area related to VLSI in both front and backend sides. Lot of applications are implemented on FPGA.

3 KNN Algorithm

KNN algorithm is very popular among the researchers, and currently, lot of applications are developing under this algorithm. KNN comes under supervised machine learning algorithm which can be used to solve regression and classification problems. This algorithm checks the similarity between new case and available cases and based on comparison put the new case into the category which matches nearest to available categories. KNN stores all the data and categorizes new data on matching points. So, whenever new data comes, then it can be categorized into a well match category by using KNN. This algorithm stores the dataset and only at the time of classification performs action on the dataset. Since algorithm not learn the from the training set immediately, KNN also known as lazy learner algorithm and non-parametric algorithm. Let us understand by the example, suppose we have images of animal that look similar to horse, goat, and cow, but we want to know that either it is a horse, goat, or cow. So for this, we implement KNN algorithm. Now KNN model will find the similar features of new data set to horse, goat and cow images and on similarity of features data will be classified as either horse or goat or cow (Fig. 1).



Fig. 1 Input and output matching using KNN

Now, point is that how KNN algorithm works. The steps-wise explanation of algorithm is as follows.

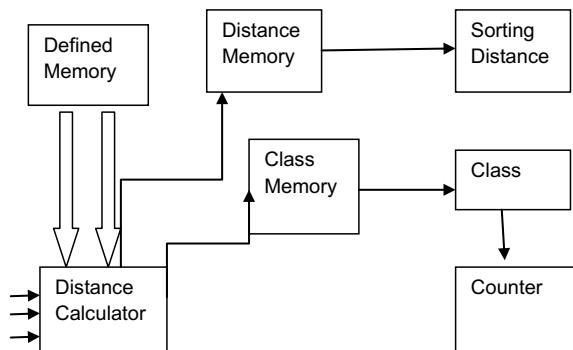
- (a) First we select the K number of the neighbors.
 - (b) Next is to calculate the Euclidean distance of K number of neighbors.
 - (c) Now as per the calculated Euclidean distance, we take the K-nearest neighbors.
 - (d) Now, determine the number of the data points in every category.
 - (e) Category for which number of neighbor is maximum to them assign the new data point and by this required model is complete.

Simplicity, versatility, fast calculation time, high accuracy, and no assumption about data are some of the advantages of this algorithm which makes KNN very popular. On the other-side, dependency in the K value and high computation cost makes KNN algorithm little.

4 Block Diagram

The block diagram is shown in Fig. 2. Block consists of seven major blocks shown in Fig. 2. The seven major blocks for implementation are defined, namely memory block, distance calculator, distance memory, sorting distance, class memory, class, and counter. To find the K-nearest neighbors, divide the K-nearest neighbors into two major parts. One is to calculate the distance between provided input vector and every reference vector, and second is to find the K smallest distance. Subtractors, multipliers, and accumulators structure is required for distance calculation. For calculation of distance, dataset is stored in memory. The data is stored in terms of array in memory. There are 128 locations for x points and y points, and total 8 locations are for defining class. Here, distance is calculated from between all the defined points and testing data point. For calculating distance concept of Euclidean distance, measurement is used. Euclidean distance basically consists of multipliers, subtractor, and accumulators. This got the reference points from memory register and calculate the distance with testing point. The distance calculated is stored in defined memory

Fig. 2 Block diagram for KNN implementation



register and calculate distance with testing point. The distance calculated is stored in another memory register where it is stored in the form of 16 bits distance and group values in 8 bit register. When distance is calculated, then distance is sorted to get the K-nearest distance for testing points. Vote counter checks the frequency of the class label. Class which got highest votes is assigned to the unknown testing data point. For finding the K-nearest neighbors, the calculation is divided into two parts. First is to calculate the distance between input vector and every reference vector, and second is to find the k smallest distance.

5 Implementation and Result

For implementation of KNN algorithm, HDL code was written in Verilog. For different blocks, code was written in HDL. The tool used for writing HDL code was Questa Sim, and after that, the code was synthesized with respect to Spartan 3E with device no 3S100EVQ100 with target frequency 66.66 MHz. Synthesis is a process of converting HDL code into a gate-level netlist where net represents the connection between each gates. For targeting the code Xilinx FPGA-based technology Spartan 3E was used. The tool used for synthesis purpose is precision synthesis (Fig. 3).

The precision synthesis tool is of Mentor Graphics Company, and it is very advanced tool used widely for synthesis process. The HDL code of distance calculation module was written and successfully synthesized on the precision synthesis tool. The RTL block of distance module is shown in Fig. 4.

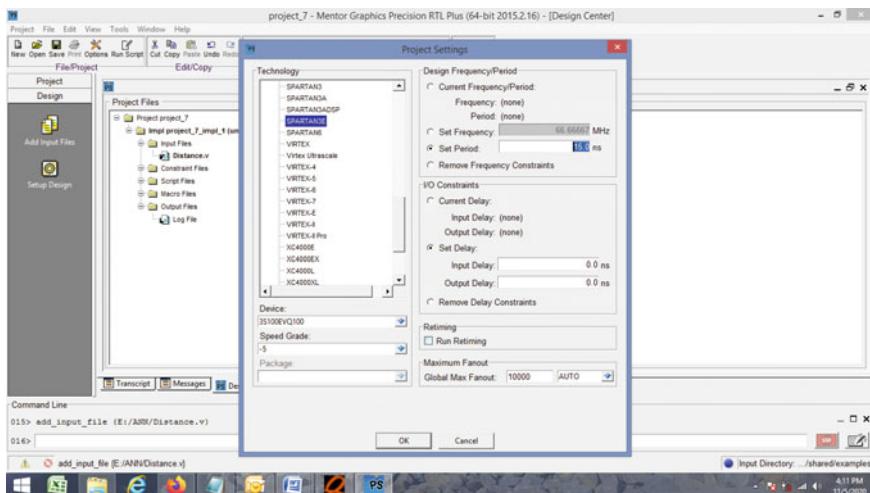


Fig. 3 Precision synthesis environment

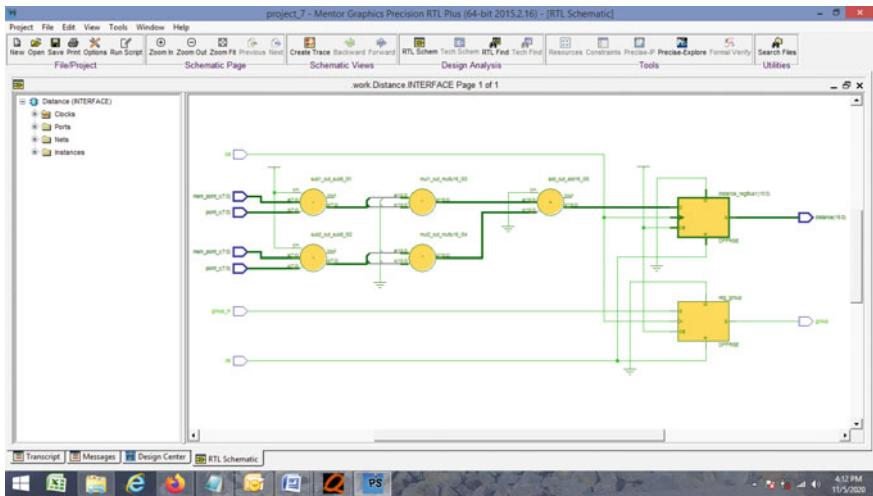


Fig. 4 RTL block diagram of distance module

The technology schematic diagram is also shown in Fig. 5. The distance calculation module is successfully synthesized with respect to desired design constraint.

From Table 1, it is clear that total number of interconnection between gates are 231 for distance calculation block, number of block multiple consumed by the distance block is 144, block multiplier is 2, and for storage, total 17 flip-flops stages were consumed. The LUT utilized by the distance block is 32. The slack generated for

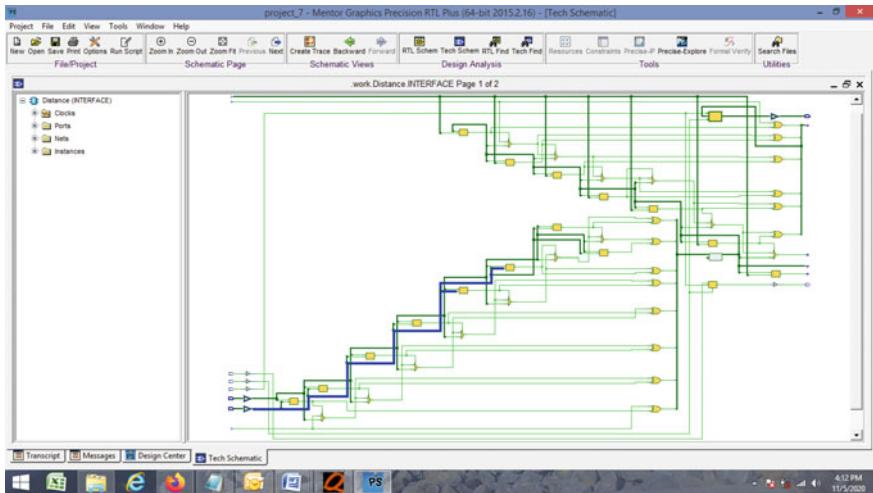


Fig. 5 Technology schematic of distance module

Table 1 Area report of distance block

S. No.	Name of content	No. of components
1	Number of nets	231
2	Number of block multiplier Dffs	144
3	Number of block multipliers	2
4	Number of Dffs or latches	17
5	Number of LUTs	32
6	Number of MUX CARRYs	29

distance calculation block is 6.61 ns. The slack is known as difference between required time and desired time. Another major block designed was sorting block. This block was coded by using Verilog HDL. The block was synthesized with respect to target technology 3S100EVQ100. The RTL code was synthesized, and the RTL schematic is shown in Fig. 6.

The technology schematic diagram of sorting distance is shown in Fig. 7.

The sorting block is little complex to design, and this can be easily known from number of lookup table (LUT) consumed. 2154 LUT blocks were consumed by the sorting module. Number of storage spaces consumed is 393. Sorting block is little more complex and consumes more area compared to distance calculation block (Table 2).

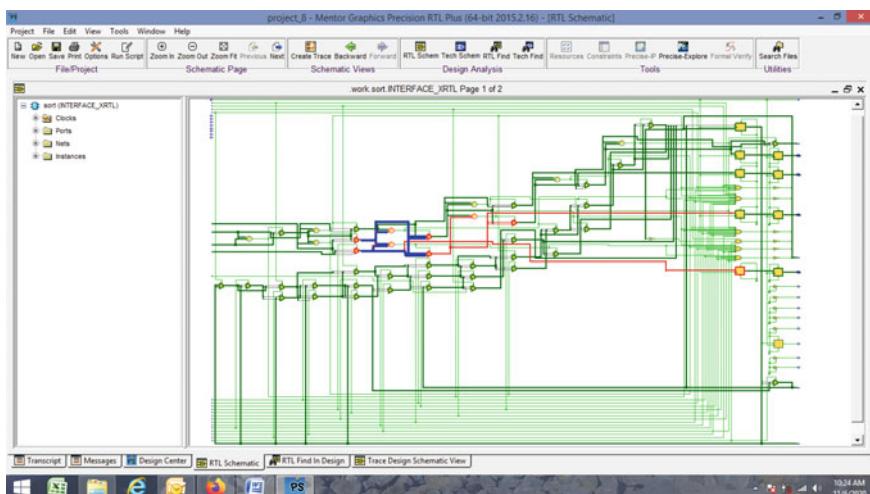


Fig. 6 RTL block diagram of sorting distance

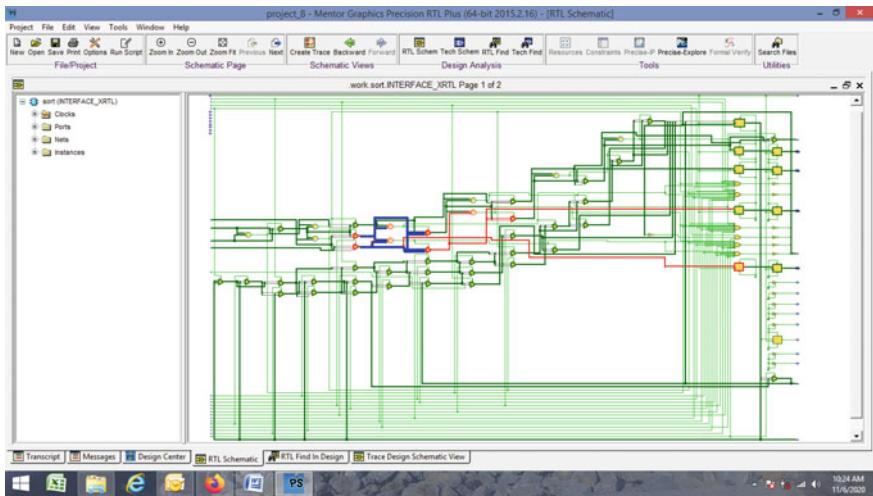


Fig. 7 Technology schematic diagram of sorting distance

Table 2 Area report of sorting

S. No.	Name of content	No. of components
1	Number of nets	3391
2	Number of Dffs or latches	393
3	Number of LUTs	2154
4	Number of MUX CARRYs	448

6 Conclusion and Future Work

The code for KNN algorithm was written in Verilog HDL, and for finding the K-nearest neighbors, we divided it into two parts. One is to calculate the distance between input vector and reference vector, and second is to find the K small distance. Currently, we coded the block diagrams in HDL and mapped to Xilinx FPGA technology Spartan 3E. Compared to Python and OpenCL, one can say that describing the behavior of KNN in HDL is more difficult and complex. Lot of challenges were come while writing the code. For synthesis process, special care was taken for writing the code, and certain keywords of Verilog HDL were avoided during code description. There is a huge scope of research in KNN algorithm implementation for the researchers. Lot of applications related to the area of machine learning and artificial intelligence are designed. Currently, the work on KNN algorithm is going on, and in the next stage, this algorithm will be implemented on FPGA Hardware. Currently, synthesis results are saying that approximately 2290 LUTs will be taken by FPGA technology. The number of LUTs can be minimized more by using effective coding style and proper techniques. This is the future research area.

References

1. Roshanaei M, Maleki M (2009) Dynamic-KNN: a novel locating method in WLAN based on angle of arrival. In: 2009 IEEE symposium on industrial electronics & applications, vol 2. IEEE, pp 722–726
2. Song X, Xie T, Fischer S (2019) A memory-access-efficient adaptive implementation of kNN on FPGA through HLS. In: 2019 IEEE 37th international conference on computer design (ICCD). IEEE, pp 177–180
3. Acceleration of k-nearest neighbor and SRAD algorithms using Intel FPGA SDK for OpenCL
4. Hussain HM, Benkrid K, Seker H (2012) An adaptive implementation of a dynamically reconfigurable K-nearest neighbour classifier on FPGA. In: 2012 NASA/ESA conference on adaptive hardware and systems (AHS). IEEE, pp 205–212
5. Pu Y, Peng J, Huang L, Chen J (2015) An efficient KNN algorithm implemented on FPGA based heterogeneous computing system using OpenCL. In: 2015 IEEE 23rd annual international symposium on field-programmable custom computing machines. IEEE, pp 167–170
6. Cho J, Jung Y, Lee S, Jung Y (2019) VLSI implementation of restricted coulomb energy neural network with improved learning scheme. Electronics 8(5):563
7. Sujitha R, Manikandan R (2014) A study on various data mining algorithms pertaining to VLSI cell partitioning

Study of Architecture and Performance Analysis of Approximation Multipliers



E. Jagadeeswara Rao, K. Murali Krishna, Md. Khasim, and T. Anjaiah

Abstract In modern society, rapidly change the technology of different applications such as Image processing (IP) and Digital Signal Processing (DSP) applications, etc., low power and high-speed Multiplier (MUL) play a major role in IP applications like biomedical and wireless communication. This paper mainly gives a systematic review of approximate multipliers (AMs) more used for shaping and smoothing applications. Also, explain systematic results of different AMs mention other research and focus on rounding based AM. Finally, this paper compares the results and gives the merits and demerits of different AMs.

Keywords Approximation multiplier · ROBA · Logarithm multiplier

1 Introduction

Achieving higher energy efficiency (EFF) has become the main objective for electronic systems and computing systems due to its limited capacity of battery, for achieving the higher EFF in digital systems can be done by improving different stages in software and hardware (SOHA) architecture (ARCH). The performance (PER) of MUL can be judged by its accuracy (ACC), speed, power consumption (PC) and time delay (TD), we have identified the different multipliers (MULs) and compared its results with our ROBA MUL, and the results are noted. So, we have compared various MULs and calculated the ACC, PC, area and TD. The digital circuits should have high ACC, low PC with minimum TD and be occupied in less

E. Jagadeeswara Rao (✉) · K. Murali Krishna

Department of Electronics and Communication Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, AP, India

Department of Electronics and Communication Engineering, Vignan's Institute of Information Technology (VIIT), Duvvada, Visakhapatnam District, AP, India

Md. Khasim · T. Anjaiah

Department of Electronics and Communication Engineering, Aditya College of Engineering & Technology, Surampalem, AP, India

area. However, if any three conditions are satisfied among those, it is efficient (EFFT) depending on our applications (APPS). If we want to design a more accurate circuit, then ACC is our primary criterion to implement the circuit. If we want to design a circuit with low PC and high ACC, we should comprise speed, like over the digital circuits (DC) designed by VLSI.

The DSP is mainly responsible for increasing the system's response, performing the heavy use of arithmetic operations like addition, subtraction, multiplication (MULP) and division. Using DSP [1], this design provides results with high ACC and eventually in power reduction. Moreover, the DSP also plays a crucial role in IP circuits. It is displaying an image with high quality using some software, i.e., computer algorithms (ALG). The main focus of this paper is on AMs circuits that are critical components for DC designing purposes, so we have observed and justified the MULs that can have the most negligible loss. In Sect. 2, the different researchers designed the AMs with techniques for improving the PER, and Sects. 3 and 4 give the table and explain methodology and results. Finally, conclude this paper in Sect. 5.

2 Systematic Literature Survey

In 1993, Schulte and Swartzlander [1] proposed truncated multiplication (TM) with correction (CORR) constant and concluded that the truncated MULs reduce the errors (ERR) by rounding the ERR when CORR needed. In 1996, Kidambi et al. [2] introduced area EFFT MULs for DSP and concluded that a sign-magnitude MUL has employed this ARCH to reduce 50% area than parallel MULs. In 1999, Hong and Kim [3] introduced about low-power (LOP) parallel MUL design for DSP APP coefficient optimization (OPT) using the novel multiplier (NM) and concluded that this methodology has developed to design power MULs using OPT and also modifications in ARCH.

In 2009, Kelly and Phillips [4] introduced approximate (APPR) signed integers MULs for arithmetic data value manipulation and concluded that the feasibility of AMs is faster than exact multipliers (EMs). In 2010, Kyaw and Goh [5] introduced LOP high-speed MUL for ERR tolerant APPS and concluded that the proposed MUL is divided into two categories—the MULP part is done by convention method, which has high ACC, and non-MULP part permits less amount of ERR.

In 2010, Mahdiani and Ahmadi [6] introduced bio-inspired imprecise computational blocks for EFF VLSI implementation of soft-computing APPS. They concluded that this APPR hardware was implemented in cost and PER.

In 2011, Kulkarni and Gupta [7] introduced trading ACC for power with under designed multiplier (UDM) ARCH. They concluded that this ARCH gives fewer amounts of ERR with fewer power losses. It achieves a better signal-to-noise ratio (SNR) and does not suffer from the overhead problem. In 2011, Venkatesan et al. [8] introduced MACACO: Modeling and Analyze of Circuits for APPR computing and concluded that (MACACO) is a new ARCH which deals with both SOHA, and it affects the approximation at the output of ALG.

In 2012, Kang and Kang [9] introduced about ACC-configurable adder (ACA) for approximation arithmetic design (AAD). This paper concluded that an adder (ADat) has high ACC during runtime but has low EFF. In 2012, Alioto [10] introduced low-power VLSI circuit design and explained that the proposed multiplier is much efficient regarding power. Liang et al. [11] introduced proposed AM for the Reliable of APPR and Probability Adders in 2013. Using approximate mirror adders (AMA) concluded that this proposal has better reliability and power EFF to perform computing APPS. In 2013, Lin and Lin [12] introduced high ACC AMs with ERR correction and concluded that it has lower TD and better ACC efficient MUL but has lower EFF.

In 2013, Gupta [13] introduced low-power DSP using APPR adders using ripple carry adders (RCA). They concluded that proposed adders have more energy efficient by having frequent less ERR.

In 2014, Bhardwaj et al. [14] introduced power and area-EFF APPR Wallace tree multiplier (AWTM). This paper concluded that produced the images quality of the almost same with only around 30% EFF in area and PC. In 2015, Momeni et al. [15] introduced the designing and analyzing of APPR compressors (COMPs) for MULP and concluded that it deals with COMPs using MUL. This COMP shows a reduction in the amount of transistor count, which results in low PC.

Osta et al. [16],] introduced AMs based on Inexact ADD for Energy EFF Data Process used META MULs and concluded that improved EFF with small ERR and minor ACC in 2015. In 2015 Mukherjee [17] introduced energy-efficient MULs for high-speed DSP, and in this paper, he concluded that by using approximate half adder (AHA) and approximate full adder (AFA) the area reduces up to 35% and propagation delay also reduces.

In 2016, Narayananamoorthy et al. [18] introduced energy-efficient APPR MULP for DSP and classification APPS. The author concluded that the proposed AM is improved ACC and energy EFF. In 2017, Akbari et al. [19] introduced dual-quality 4:2 COMP for utilizing in dynamic ACC-configurable (DAC) MULs. They concluded that using a 32 bits Dadda multiplier (DM), COMP was employed with ACC and is energy efficient.

In 2017, Begum [20] introduced about design of FIR filter using ROBA MUL and concluded that the proposed MUL shows better PER in terms of area, power and TD. In 2017, Venu Kumar [21] proposed implementing ROBA MUL using APPR MUL for high-speed and energy EFF DSP APPS and concluded that the proposed MUL, which had high ACC, improves speed and PC with a small amount of ERR.

In 2018, Yasodha [22] proposed the design of Roba MUL using Booth signed multiplier (BSM) and Brent Kung adder (BKA) and concluded that the APPR MUL, which is based on rounding, is designed using BSM And BKA, which is more efficient than the previous APPR and accurate MULs. In 2018, Susi [23] proposed DAC 64 bit DM using dual-quality 4:2 COMPs and concluded that the proposed APPR COMPs are improved compared with those of the DM employing the exact COMP.

In 2018, Marimuthu et al. [24] proposed fragmentation programmable array (FPGA) implementation of rounded-based AM for improving the PER of DSP. The high ACC of AM is dependent on the rounded value of 2n-bit input. The high ACC

requested MUL is based on the input inverse of the 2n. In 2018 Tikkireddy and Manikanta [25] proposed that Reliable and accurate low, dense MULs and concluded that the proposed one has High speed and energy-efficient AM has achieved.

In 2018, Arya and Rahul [26] proposed a new proposal in VLSI-based MULs for ERR fewer DSP APPS and concluded that MULPs have high execution time in DSP systems. The proposed trend has low ERR. In 2018, Nagajyothi and Ancia [27] introduced a high-speed and energy-efficient Roba MUL using Ladner Fischeradder (LF) adder. They concluded that rounding is designed using LF, which is more efficient than the previous approximate and accurate multipliers. An AM (ROBA) using LF adder implementation was proposed in the paper. In 2018, Venkateshwarlu and Papa Rao [28] introduced the low-power design of cryptography using ROBA MUL. They concluded that a new ARCH had been established to achieve fast execution using the cryptography model. From 2016 to 2018, Durgesh Nandan et al. suggested few modifications in the existing logarithm multiplier, which open new possibilities of utilizing logarithm multiplier as a critical component of approximation applications.

In 2018, Vaeztourshizi et al. [29] introduced an energy EFF, yet highly-ACC and APPR non-iterative divider (AID). They concluded that a proposed divider has a low error and also improves MULP as its primary operation. In 2018, Gayathri et al. [30] introduced dual-quality 4:2 COMP-based configurable MULs using Vedic multiplier (VM). They concluded that four COMPs were proposed using reversible logic that can switch between exact and APPR modes. In APPR modes and exact model, the delay times and PC are the same but different in ACC. A thorough review of newly developed approximation multipliers was conducted by Jagadeeswara Rao et al. [31] (nearly 45 approximate multipliers). During the research, it was discovered that there is currently no approximation multiplier that offers superior design metrics and error metrics.

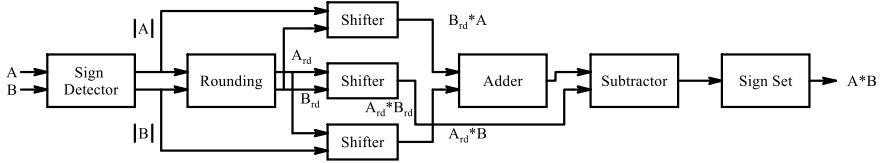
3 Methodology

The primary purpose of the AM is to perform MULP on the numbers (NUMs) that are power n to base 2. Let us denote round NUMs of input of A and B by A_v and B_v , and then, the MULP of the two numbers is

$$A \times B = (A_{rd} \times B) + (A \times B_{rd}) + (A \times B) - (A_{rd} \times B_{rd}) \quad (1)$$

The terms $A_{rd} \times B$, $A \times B_{rd}$ and $A_{rd} \times B_{rd}$ can be obtained just by using logical shift operation. The hardware implementation is also quite complex, so simplify the MULP process, and the expression is further simplified in (2). The AMs architecture has shown in Fig. 1.

$$A \times B \simeq (A_{rd} \times B) + (A \times B_{rd}) - (A_{rd} \times B_{rd}) \quad (2)$$

**Fig. 1** AMs architecture

4 Results

The PER analysis of different AMs are mentioned here and the results that we have compared in terms of PC and TD of different types of AMs shown in Table 1. We draw a comparative graph of power consumption which is shown in Fig. 2. Delay graph is shown in Fig. 3, and the errors comparison graph is shown in Fig. 4. As a result, we found that ROBA performed based on all parameters out of all available multipliers. The power consumptions are in the order of (mW). In the case of an 8-bit multiplier, the largest power consumption is observed in DM [23] of 240 mW, while the smallest power consumption is observed in ACA [9] of 5.96 mW which is seen in Fig. 1. From Fig. 3, it is noted that the delays of the designs are of the order of (ns). Considering an 8-bit multiplier, largest delay is observed in VM [32] of 66 ns, while the smallest delay is observed in ROBA [21] of 8.56 ns. Error is measured

Table 1 Performance analysis comparison of various AMs

Design	Power (mw)	Delay (ns)	Errors (%)
ACA [9]	5.76	25.7	0.3
AWTM [14]	55.76	9.6	0.68
FPGA [24]	13.7	14.72	0.36
META [16]	57	18.4	1.23
DM [23]	240	23.87	0.2
BKA [22]	102	30.77	0.63
VM [32]	109	66	5.3
ROBA [21]	5.37	8.56	0.2
LF ADDER [27]	129	30.6	2.86
TM [1]	51.5	16.24	5.36
NM [3]	63.7	15.08	8.52
MACACO [8]	52.56	14.72	3.5
RCA [13]	68.04	10.7	2.36
AHA, AFA [17]	83	25.646	0.3
UDM [7]	37.9	33	0.56
AMA [11]	31.5	31.84	1.50
AID [29]	21.2	22.52	2.2

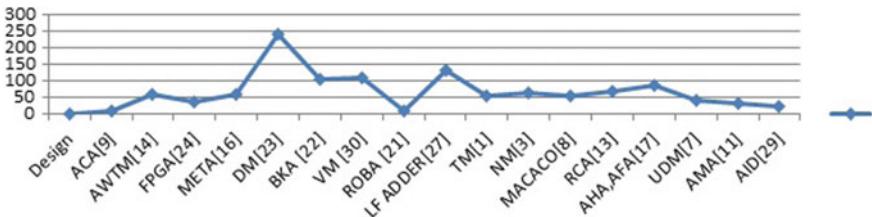


Fig. 2 Graph for various multipliers power consumption

Fig. 3 Graph for various multipliers delay comparisons

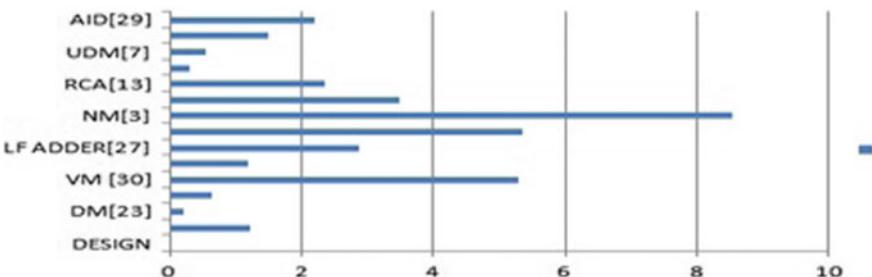
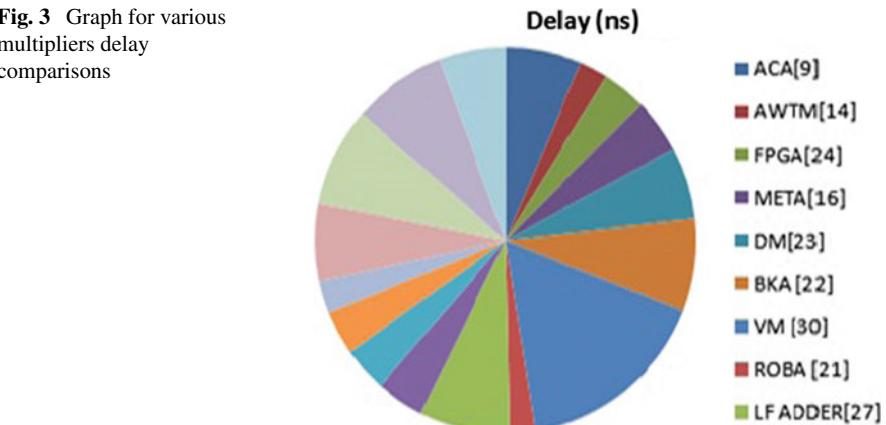


Fig. 4 Graph for various multipliers errors comparisons

according to percentage. For an 8-bit multiplier, the largest error observes NM [3] of 8.52%. Similarly, for an 8-bit multiplier, the smallest error noted in AFA [17] and ACA [9] of 0.3% is shown in Fig. 4.

5 Conclusion

Many authors conclude that the different AMs are designed and introduced an efficient multiplier until the literature review, but they were partially concluded. In this paper, we discussed AMs, based on the development in AMs and which one is best in what suitable condition. This study includes the different techniques used by researchers to improve the speed of AMs. Still, few researchers did not focus on the error present in approximate computing when number of bits increases the AMs. Finally, we are concluded that the ROBA is the best choice for designers based on power is the primary concern. But the FPGA is the most EFF design in terms of area and not in PC. This conclusion is hoped to contribute significant changes in DSP and IP areas.

References

1. Schulte MJ, Swartzlander E (1999) Truncated multiplication (TM) with correction constant. *IEEE Trans Electron Comput* 41(3):388–396
2. Kidambi SS (1996) Area efficient multipliers for digital signal processing. *IEEE Trans Circuits Syst* 11: Analog Digit signal Process 43(2):90–95
3. Hong S, Kim S (1999) Low power parallel multiplier (novel multiplier) design for DSP application through coefficient optimization. In: Proceeding on 12th annual IEEE international ASIC/SOC conference, Washington, DC, USA, pp 286–290
4. Kelly DR, Phillips BJ (2009) Approximate signed binary integer multipliers for arithmetic data value speculation. *IEEE Trans Integr Tech Circuits* 992:255–265
5. Kyaw KY, Goh WL (2010) Low-power high-speed multiplier for error tolerant application. In: Proceeding on IEEE international conference of electronic devices and solid-state circuits (EDSSC), Hong Kong, China, pp 1–4
6. Mahdiani HR, Ahmadi A (2010) Bio-inspired imprecise computational blocks for efficient VLSI implementation of soft-computing applications. *IEEE Trans Circuits Syst: Regul Pap* 57(4):850–862
7. Kulkarni P, Gupta P (2011) Trading accuracy for power with an under-designed multiplier architecture. *IEEE Trans Electron Circuits* 56(68):1–7
8. Venkatesan R, Agarwal A, Roy K (2011) MACACO: modeling and analysis of circuits for approximate computing. In: Proceeding on IEEE/ACM international conference on computer-aided design (ICCAD), San Jose, CA, USA, pp 667–673
9. Kahng AB, Kang S (2012) Accuracy-configurable adder (ACA) for approximate arithmetic designs. In: Proceeding on DAC design automation conference, San Francisco, CA, USA, pp 820–825
10. Alioto M (2012) Ultra-low power VLSI circuit design demystified and explained. *IEEE Trans Circuits Syst: Regul Pap* 59(1):3–29
11. Liang J, Han J, Lombardi F (2013) New metrics for the reliability of approximate and probabilistic adders. *IEEE Trans Comput* 62(9):1760–1771
12. Lin C-h, Lin I-C (2013) High accuracy approximate multiplier with error correction. In: Proceeding on IEEE 31st international conference on computer design (ICCD), Asheville, NC, USA, pp 33–38
13. Gupta V (2013) Low power digital signal processing using approximate adders (RCA). *IEEE Trans Comput-Aided Des Integr Circuits Syst* 32(1):124–137

14. Bhardwaj K, Mane PS, Henkel J (2014) Power- and area-efficient approximate wallace tree multiplier (AWTM) for error-resilient system. In: Proceeding on 15th international symposium on quality electronic design, Santa Clara, CA, USA, pp 263–269
15. Momeni A, Han J, Montuschi P, Lombardi F (2015) Design and analysis of approximate compressors for multiplication. *IEEE Trans Comput* 64(4):984–994
16. Osta M, Ibrahim A, Valle M (2015) Approximate multipliers based on inexact adders for energy-efficient data processing. In: Proceeding on new generation of CAS (NGCAS), Genova, Italy, pp 125–128
17. Mukherjee S (2015) Energy efficient multipliers for high-speed DSP applications. *Int J Comput Sci Mob Comput* 4(6):66–75
18. Narayananamoorthy S, Moghaddam HA, Liu Z, Park T, Kim NS (2015) Energy-efficient approximate multiplication for digital signal processing and classification of applications. *IEEE Trans Very Large Scale Integr* 23(6):1180–1184
19. Akbari O, Kamal M, Afzali-Kusha A, Pedram M (2017) Dual-quality 4:2 compressors for utilizing in dynamic accuracy configurable multipliers. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 25(4):1352–1361
20. Begum S, Vijaya Kumar M (2017) Design of FIR filter using rounding based approximate (ROBA) multiplier. *Int J Sci Eng Technol Res* 22(6):4483–4489
21. Venu Kumar K (2017) Implementation of ROBA multiplier using approximate multiplier for high-speed & energy-efficient DSP applications. *Int J Innov Technol* 5(9):1824–1828
22. Yasodha P (2018) Design of Roba multiplier using booth signed multiplier and Brent Kung adder. *Int J Eng Sci Invent (IJESI)* 7(4):8–14
23. Susi S (2018) Dynamic accuracy configurable 64 bit multipliers using dual-quality 4:2 compressors. *J Emerg Technol Innov Res (JETIR)* 7(5):1032–1041
24. Marimuthu D, Bharath S (2018) FPGA implementation of rounded based approximate multiplier for efficient performance of digital signal processing. In: Proceeding on international conference on progressive research in applied sciences, engineering and technology, India, pp 19–26
25. Tikkireddy S, Manikanta SVG (2018) Reliable and accurate low dense multipliers. *Int J Eng Adv Res Sci Techno* 3(2):689–696
26. Arya N, Rahul M (2018) A new trend in VLSI based multipliers for error resilient DSP applications. *Int Res J Eng Technol (IRJET)* 5(4):3866–3869
27. Nagajyothi NR, Ancia A (2018) A high-speed and energy-efficient Roba multiplier using LF adder. *Pramana Res J* 8(12):52–58
28. Venkateshwarlu M, Papa Rao Ch (2018) Low power design of cryptography by using ROBA multiplier. *Int J Manag Technol Eng* 8(11):2502–2508
29. Vaeztourshizi M, Kamal M, Afzali-Kusha A, Pedram M (2018) An energy-efficient, yet highly-accurate, approximate non-iterative divider. In: Proceedings of the international symposium on low power electronics and design (ISLPED '18). Association for Computing Machinery, New York, NY, USA, pp 1–6
30. Medha Gayathri I, Sindhoori P, Sravan Kumar Reddy P (2018) Design and synthesis of multiplier using dual quality 4:2 compressors with higher speed and lower power consumption. *Int J Res Advent Technol* 16(6):1082–1089
31. Jagadeeswara Rao E, Samundiswary P (2021) A review of approximate multipliers and its applications. In: Advances in automation, signal processing, instrumentation, and control. Lecture notes in electrical engineering. Springer, Singapore, pp 1381–1392
32. Nandan D, Kanungo J, Mahajan A (2017) An efficient VLSI architecture for Iterative Logarithmic Multiplier. In: Proceeding on IEEE 4th international conference on signal processing and integrated networks (SPIN), Noida, pp 419–423

Design and Analysis of Energy-Efficient Logic Gates Using INDEP Short Gate FinFETs at 10 nm Technology Node



Umayia Mushtaq, Md. Waseem Akram, and Dinesh Prasad

Abstract In this paper, basic logic gates are designed at 10 nm technology node using Fin Field Effect Transistor (FinFET), and comparative analysis is performed with the proposed one based on input dependent (INDEP) technique. The total power dissipation in case of FinFET INDEP NAND gate and FinFET INDEP NOR gate is reduced by 63.28 and 66.08%, while power delay product is reduced by 63.26% and 50.06%, respectively, in comparison with the FinFET NAND and NOR gate without INDEP technique. Comparative analysis is also performed between INDEP FinFET inverters with the one without INDEP technique. The simulation results show that the design of logic gates using INDEP FinFET is more efficient in comparison with the one without technique. The reliability of the logic gates is also checked using Monte Carlo approach which clearly depicts the improved performance parameters of FinFET logic gates designed using INDEP technique.

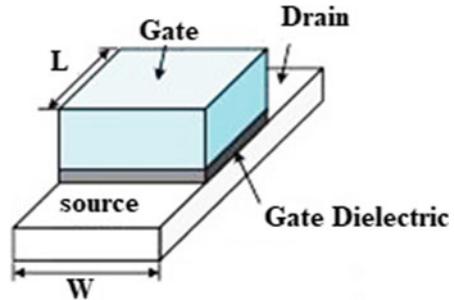
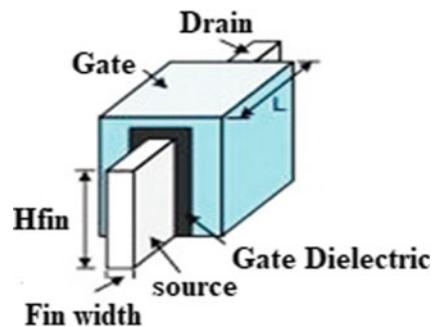
Keywords CMOS · FinFET · Power dissipation · INDEP · PDP

1 Introduction

Due to continuous scaling in deep sub-micrometer region, a conventional device like complementary metal-oxide semiconductor (CMOS) has reached its limits at 22 nm technology node. This occurs due to increase in short channel effects (SCEs), which mostly increases the leakage current. These effects can be minimized by using multi-gate transistors due to the reason that multi-gate transistors have better electrostatic control over the channel [1]. FinFET has come out as the best candidate among multi-gate devices from the fabrication point of view [2]. In addition to this, it provides better control of short channel effects and lower leakage current and better performance [3, 4]. The structures of planner device and FinFET device are shown in Figs. 1 and 2, respectively. Planner devices have the horizontal channel, while in case of FinFET, the channel is a thin vertical fin wrapped by the gate electrode [5].

U. Mushtaq (✉) · Md. Waseem Akram · D. Prasad

Department of Electronics and Communication Engineering, Faculty of Engineering & Technology, Jamia Millia Islamia, New Delhi 110025, India

Fig. 1 Planner device [5]**Fig. 2** FinFET device [5]

In FinFET devices, the current direction is parallel to die plane, but the channel formation is perpendicular to the plane of wafer [6]. In case of CMOS logic devices, various circuit-level leakage reduction techniques, such as LEakage Control Transistor (LECTOR) [7], Gated Leakage transistOR (GALEOR) [8], Stacking, multiple threshold CMOS (MTCMOS) [9], ONOFIC [10] and input dependent (INDEP) [11] have been devised from time to time to reduce the leakage power dissipation, propagation delay and power delay products, and among them, INDEP technique has proven as the best leakage reduction technique [12]. Due to the ease in fabrication and better performance parameters, FinFET is used to design 5G memory like content addressable memory (CAM) [13] and static random-access memory (SRAM) memory as well [14]. So, in this paper, FinFET logic gates are designed at 10 nm technology node, and comparative analysis is performed with the INDEP FinFET gates. The rest of the paper is organized as follows.

Section 2 provides the design methodology used to implement FinFET logic gates. Section 3 gives the comparative analysis of various performance parameters of FinFET logic gates and INDEP FinFET logic gates at 10 nm technology node. In addition to this, it consists of simulation results and discussion section. Besides, the reliability analysis using Monte Carlo simulations performed at $\pm 10\%$ variation under 3σ Gaussian distribution for process–voltage–temperature (PVT) parameters is also discussed in the same section. Finally, the conclusion and future work are provided in Sect. 4.

2 Design and Implementation of FinFET Logic Gates

The innovations in materials and device structures of MOSFET from the past few decades have resulted in better performance, besides reducing power dissipation, it has improved other parameters like propagation delay and power delay products. FinFET is a non-planar double gate device and hence removes scaling obstacles to large extent due to the reason that front and back gate can control the channel independently as well. This results in better control of threshold voltage and other process variations and hence provides a means to manage leakage power dissipation [15]. In addition to power dissipation, estimation of propagation delay in the chip is an important factor. For the bigger chips, it is necessary to know how much amount of delay is there between primary input and primary output across the critical path to get the maximum frequency of operation of the chip. Other parameter called power delay product (PDP) is the product of power consumption times propagation delay. It gives the idea about energy efficiency of logic gates or logic family in digital electronics. It gives the estimate of energy consumed per switching event. Therefore, in our circuit design using FinFET at 10 nm technology node, PDP should be low compared to conventional CMOS, resulting in minimum energy consumption. In this section, basic FinFET gates (NAND, NOR and inverter) are designed using HSPICE simulation tool at 10 nm technology node. Various performance parameters like power dissipation, propagation delay and power delay products are analyzed for different logic gates. In this section, comparison is provided between basic logic gates and the ones with input dependent (INDEP) technique. This INDEP technique was actually designed for conventional CMOS logic circuits [11] to reduce power dissipation in it. In our analysis, it is implemented to design FinFET logic gates as well as shown in Fig. 3.

This technique provides a better trade-off between power dissipation and propagation delay in digital logic circuits. This technique does not increase propagation delay to a large extent as in case of stacked transistors due to the proper input terminal selection of INDEP transistors. In our design, two extra FinFETs are inserted between pull-up and pull-down networks. The extra inserted FinFETs can be called as INDEP FinFETs (one n-type FinFET and one p-type FinFET) due to the reason that low leakage logic circuits are formed due to the controlling of input terminals of FinFET INDEP transistors. The inputs V_0 and V_1 as shown in Table 1 are chosen in such a way that it increases the number of OFF-state FinFETs from V_{DD} (supply voltage) to ground. This increase in number of OFF-state transistors which increases the resistance from V_{DD} (0.75 V) to ground hence results in decrease in leakage power dissipation. The propagation delay of this technique increases by small amount. But the power delay product reduces which makes our design of FinFET gates energy efficient.

Fig. 3 Schematic of INDEP technique using short gate FinFETs

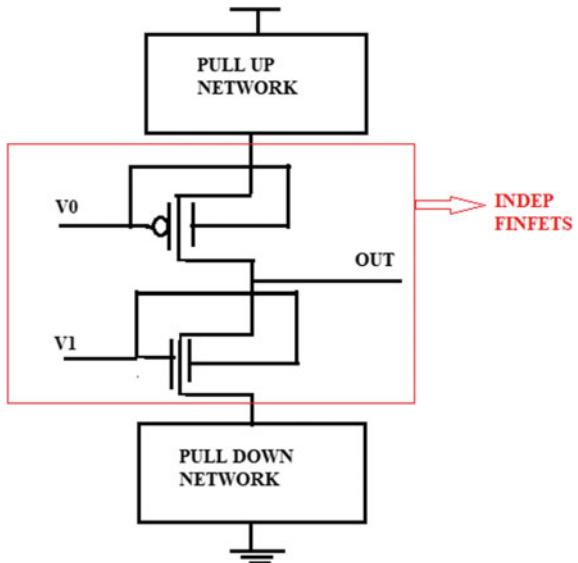


Table 1 Inputs to INDEP FinFETs

FinFET gate	Inputs	V_0	V_1
Inverter	0	0	0
	1	1	1
NAND	00	0	0
	01	0	0
	10	0	0
	11	1	1
NOR	00	0	0
	01	1	1
	10	1	1
	11	1	1

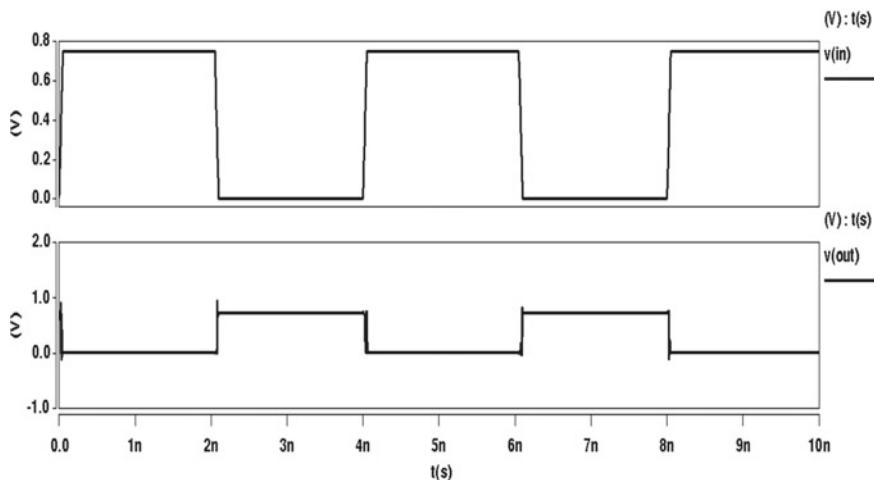
3 Simulation Results and Discussion

Various performance parameters of basic FinFET inverter, FinFET NAND and FinFET NOR are compared with INDEP FinFET inverter, INDEP FinFET NAND and INDEP FinFET NOR, respectively. The performance parameters are given in Table 2, and the waveforms obtained for different FinFET gates are shown in Figs. 4, 5, 6, 7, 8 and 9.

The analysis clearly shows that design of logic gates using FinFET results in decrease in leakage power dissipation and power delay product. This gets reduced further when INDEP technique is used in FinFET logic circuits. This clearly shows

Table 2 Comparison of performance parameters of basic logic gates using FinFETs with and without INDEP technique at 10 nm technology node

Logic gates	Parameters	Conventional FinFET device	FinFET device with INDEP technique
Inverter	Power dissipation (nW)	59.91	22.49
	Propagation delay (ps)	5.45	8.94
	Power delay product (aJ)	0.33	0.20
NAND	Power dissipation (nW)	63.40	23.27
	Propagation delay (ns)	5.93	5.94
	Power delay product (aJ)	376.40	138.27
NOR	Power dissipation (nW)	74.32	25.20
	Propagation delay (ns)	4.03	5.94
	Power delay product (aJ)	299.98	149.75

**Fig. 4** Waveform of SG (short gate) FinFET inverter at 10 nm technology node

that FinFET can be proven as one of the best technologies for CMOS digital logic design.

The power dissipation and power delay product reduces by 62.46% and 39.39%, respectively, in case of short gate FinFET INDEP inverter as compared to short gate FinFET inverter at 10 nm technology node. The decrease in power dissipation and power delay product is due to the increase in number of OFF transistors from power supply to ground which increases the resistance of path from V_{DD} (supply voltage) to ground. In case of INDEP FinFET NAND gate, the power dissipation gets decreased by 63.28%, while PDP in INDEP FinFET NAND reduces by 63.26%. However, in INDEP NOR FinFET gate, power dissipation reduces by 66.08% in comparison with FinFET NOR gate at 10 nm technology node. In addition to this,

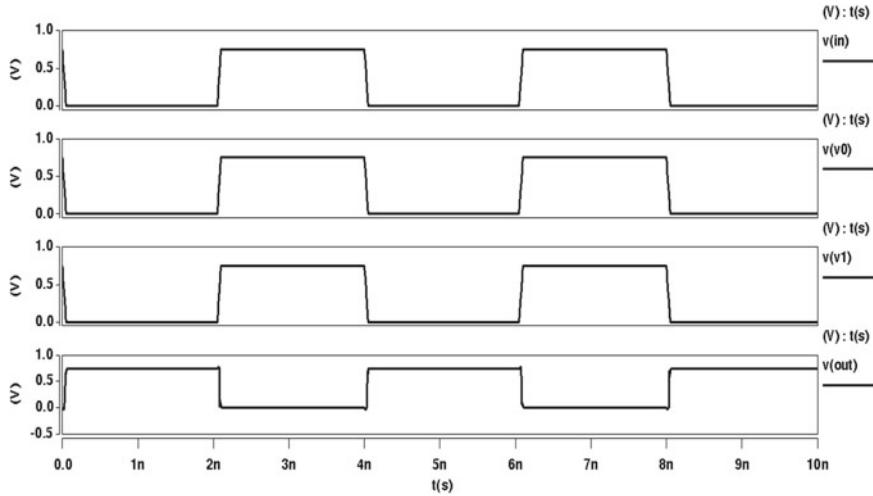


Fig. 5 Waveform of SG (short gate) INDEP FinFET inverter at 10 nm technology node

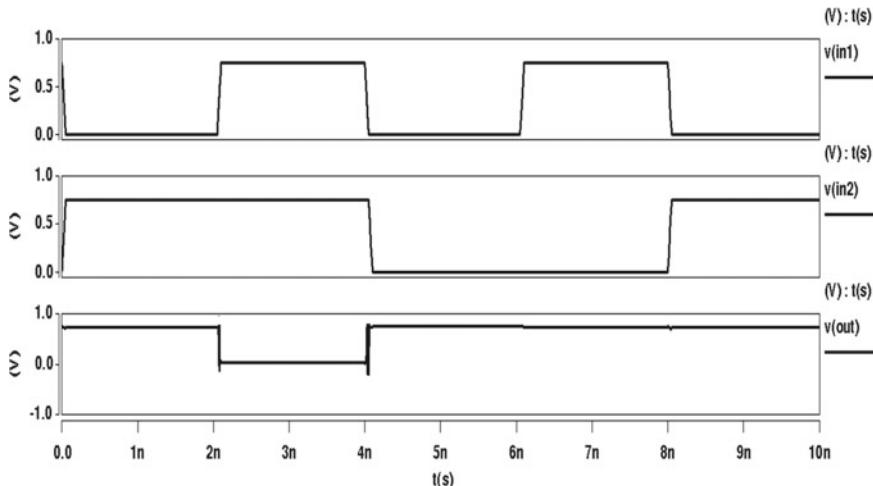


Fig. 6 Waveform of SG FinFET NAND gate at 10 nm technology node

PDP in case of INDEP NOR FinFET reduces by 50.06% in comparison with FinFET NOR gate. The comparative analysis of power dissipation for inverter, NAND and NOR FinFET circuits is represented in Fig. 10, which clearly depicts that INDEP FinFET circuits have better performance in comparison with one without technique at 10 nm technology node.

The performance of logic gates is affected in deep submicron region. Therefore, PVT variations like channel length, supply voltage and temperature in FinFET logic

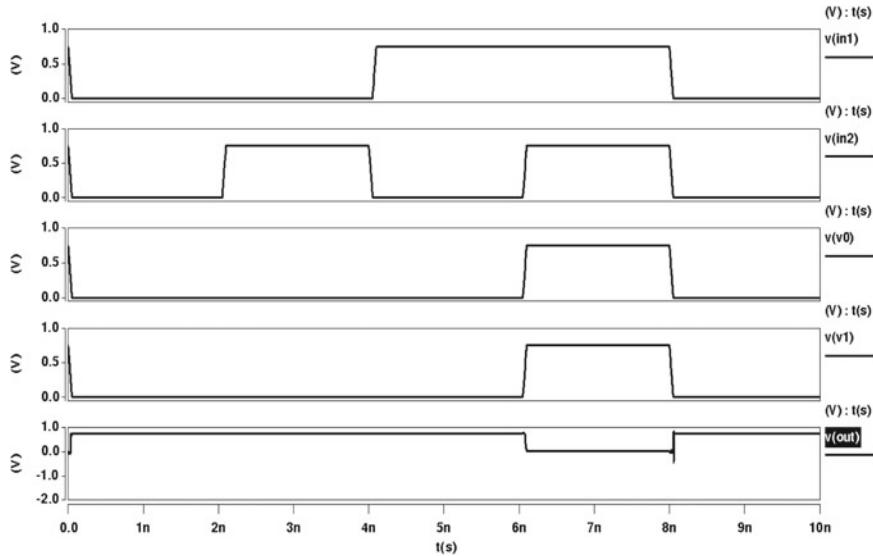


Fig. 7 Waveform of SG INDEP FinFET NAND gate at 10 nm technology node

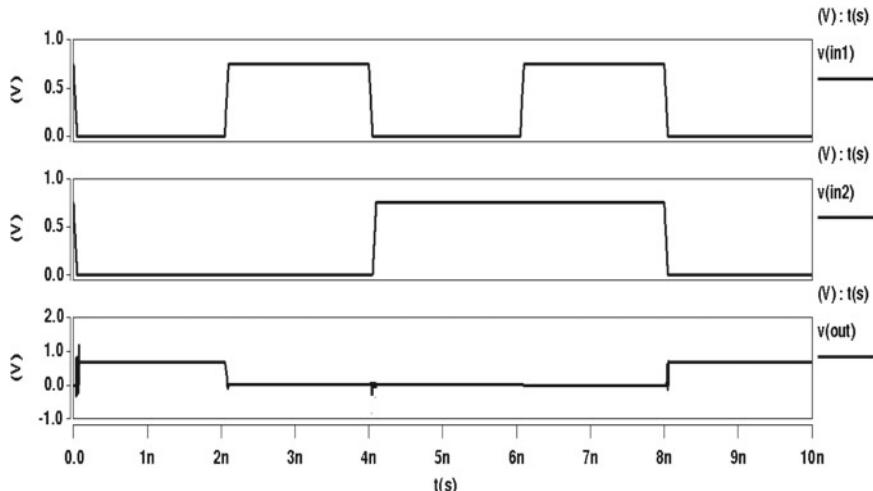


Fig. 8 Waveform of SG FinFET NOR gate at 10 nm technology node

gates and INDEP FinFET logic gates are analyzed. This can be done by reliability analysis using Monte Carlo approach. The reliability of FinFET logic gates is tested by Monte Carlo analysis. This analysis gives the estimate of uncertainties of the logic circuits. The statistical parameters are obtained by Monte Carlo approach. The simulations are performed for $\pm 10\%$ variation under 3σ Gaussian distribution for

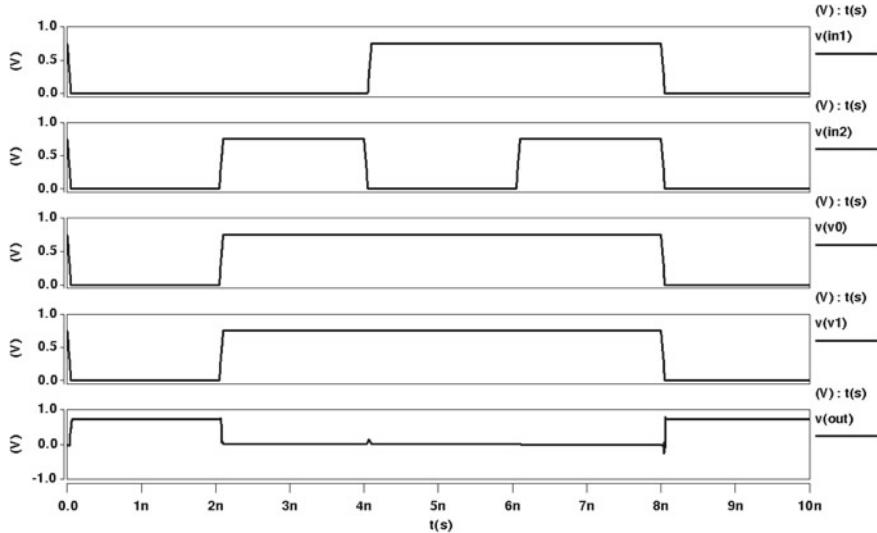


Fig. 9 Waveform of SG INDEP FinFET NOR gate at 10 nm technology node

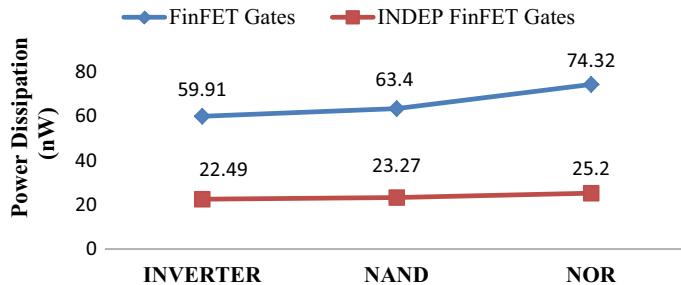


Fig. 10 Comparative analysis of power dissipation for FinFET gates with and without INDEP technique

PVT parameters for power delay product for 50 samples as shown in Table 3. The reduction in standard deviation values clearly depicts that INDEP FinFET logic gates are more immune to process parameter variations than the FinFET logic gates.

4 Conclusion and Future Work

The analysis of logic gates clearly depicts that performance parameters are improved in INDEP FinFET logic gates to a large extent. Simulation results clearly show the low leakage and PDP capabilities of FinFET INDEP approach. Therefore, FinFET with INDEP technique can be proved as one of the best leakage reduction mechanisms

Table 3 Statistical Monte Carlo simulation data of power delay products for different digital logic gates

Metric	Gate	Conventional FinFET gates		INDEP FinFET		% change	
		Mean μ (aJ)	SD σ (aJ)	Mean μ (aJ)	SD σ (aJ)	$\Delta\mu$	$\Delta\sigma$
Power delay product	Inverter	90.49	56.4	20.21	37.86	77.66	32.87
	NAND	169.02	93.6	133.89	14.45	20.78	84.56
	NOR	284.96	55.07	116.32	98.89	59.18	-79.57

especially in lower technology nodes. In addition to this, PVT variations can also be improved to a large extent. As a part of future work, FinFET logic devices can be used to design different memories at lower technology nodes for 5G communication networks. The fin-like fin height, fin length and fin width of FinFET can also be varied which can play a crucial role in the design of logic circuits.

References

1. Bhoj AN, Jha NK (2013) Design of logic gates and flip-flops in high-performance FinFET technology. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 21(11):1975–1988
2. Nowak EJ, Aller I, Ludwig T, Kim K, Joshi RV, Chuang CT, Bernstein K, Puri R (2004) Turning silicon on its edge [double gate CMOS/FinFET technology]. *IEEE Circuits Devices Mag* 20(1):20–31
3. Henderson CL (2013) Failure analysis techniques for a 3D world. *Microelectron Reliab* 53(9–11):1171–1178
4. Russ C (2008) ESD issues in advanced CMOS bulk and FinFET technologies: processing, protection devices and circuit strategies. *Microelectron Reliab* 48(8–9):1403–1411
5. Guo X, Verma V, Gonzalez-Guerrero P, Mosanu S, Stan MR (2017) Back to the future: digital circuit design in the FinFET era. *J Low Power Electron* 13(3):338–355
6. Mishra P, Muttreja A, Jha NK (2011) FinFET circuit design. In: Nanoelectronic circuit design. Springer, New York, NY, pp 23–54
7. Hanchate N, Ranganathan N (2004) LECTOR: a technique for leakage reduction in CMOS circuits. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 12(2):196–205
8. Chun JW, Chen CR (2010) A novel leakage power reduction technique for CMOS circuit design. In: 2010 international SoC design conference. IEEE, pp 119–122
9. Roy K, Mukhopadhyay S, Mahmoodi-Meimand H (2003) Leakage current mechanisms and leakage reduction techniques in deep-sub micrometer CMOS circuits. *Proc IEEE* 91(2):305–327
10. Sharma VK, Pattanaik M, Raj B (2014) ONOFIC approach: low power high speed nanoscale VLSI circuits design. *Int J Electron* 101(1):61–73
11. Sharma VK, Pattanaik M, Raj B (2015) INDEP approach for leakage reduction in nanoscale CMOS circuits. *Int J Electron* 102(2):200–215
12. Mushtaq U, Sharma VK (2021) Performance analysis for reliable nanoscaled FinFET logic circuits. *Analog Integr Circuits Signal Process* 1–12
13. Arulvani M, Ismail MM (2018) Low power FinFET content addressable memory design for 5G communication networks. *Comput Electr Eng* 72:606–613

14. Mushtaq U, Sharma VK (2020) Design and analysis of INDEP FinFET SRAM cell at 7-nm technology. *Int J Numer Model: Electron Netw Devices Fields* 33(5):e2730
15. Turi MA, Delgado-Frias JG (2017) Full-VDD and near-threshold performance of 8T FinFET SRAM cells. *Integration* 57:169–183

SIMPLE-DRR: A New Energy-Efficient Multi-hop Routing Protocol in WBANs for Health Monitoring



Subba Reddy Chavva and Ravi Sankar Sangam

Abstract Wireless body area networks (WBANs) are a growing field. One of the most important applications of WBANs is health monitoring. Body area sensor nodes are attached to the human body to monitor vital parameters. These tiny nodes often have limited power and are tough to replace after the battery runs out. Energy-efficient routing techniques are essential to extend the network's lifetime. SIMPLE is a multi-hop routing protocol that uses little energy. The forwarder node in this protocol is chosen depending on the distance and residual energy of the nodes. However, it has a shorter stability period and a shorter network lifetime. We present a new energy-efficient technique for enhancing nodes and are used to identify the forwarder node in our novel protocol. The received signal strength indicator (RSSI), alive nodes, distance, and residual energy are used to determine the cost function. By comparing our designed protocol to existing state-of-the-art protocols SIMPLE and M-ATTEMPT, we found that it exceeded them in terms of stability, network lifetime, and throughput.

Keywords Energy-efficient · Health monitoring · Multi-hop routing · WBANs

1 Introduction

Due to various reasons in recent times, a huge amount of money spends on health check-ups. Normally, a person nearly spends an average of 20–30% of his/her earnings for general health checks-ups or chronic deceases. Internet of Things (IoT) has to be most impactful and demanding innovation in the health industry. It is often referred to as Internet of Medical Things (IoMT). It is a subcategory of IoT [1].

WBANs are a subgroup of IoMT as well as wireless sensor networks (WSNs). Due to internal architectural changes, the WSN methods and protocols are incom-

S. R. Chavva (✉) · R. S. Sangam

VIT-AP University, Near Vijayawada, Amaravati, Andhra Pradesh 522237, India

e-mail: chavvasubbareddy@gmail.com

patible with the WBANs. Telemedicine, biofeedback, rehabilitation and therapy, ambient-assisted living, and health monitoring (smart health) are some applications in WBANs. Health monitoring is one the main applications widely used in several sectors like military, sports [2, 3].

In human body, nodes are placed to monitor the various internal organ activities in the human body. Sensor nodes are classified into two types: wearable and implanted. The wearable nodes are electromyography (EMG), pulse oximetry, blood pressure (BP), electrocardiography (ECG), and electroencephalography (EEG), and implantable nodes are glucose monitoring and implantable neural stimulators [4–6].

These wireless nodes detect and send information from the human body to the sink node. The data transfer can be done through some route. These routes are fixed or dynamically change based on the importance of data. There are many routing protocols developed in WBANs for health monitoring. The most important protocols are M-ATTEMPT and SIMPLE [7–9].

M-ATTEMPT directly transfers critical and on-demand information to sink node. The nodes are put in the skin generate heat and detect a hot spot. The data transfer takes place via a path away from the hot spot. Sensor nodes with multi-hop routing have a lower temperature than sensor nodes without multi-hop routing. SIMPLE is another energy-efficient protocol. In this protocol, sensed data is being transferred to a base station by using a forwarder or parent node. The forwarder node is chosen depending on the sensor nodes' distance and residual energy. These protocols have less stable, and the network serves fewer times [10, 11]. The main objectives of the proposed protocol are summarising as follows:

- To developing an energy-efficient protocol.
- To increase stability and network lifetime in WBANs.

The remaining sections of the paper are organised as follows. Literature review is included in Sect. 2 and ref The system model is discussed in detail in Sect. 3. A description of the proposed technique can be found in Sect. 4. The findings of the simulation and the debate that follows are covered in Sect. 5. Results that are comparable to one another are shown in Sect. 6. Conclusion and outlines plan for further research in Sect. 7.

2 Literature Review

There are some state-of-art protocols that are reviewed in this section.

DARE protocol is discussed by A. Tauqir et al. In a hospital ward, this protocol was observed. A hospital has eight patients. Each patient is equipped with wireless sensor nodes that monitor a variety of characteristics. A sink node has been moved to a different location in a network. Sensor nodes with decreasing energy transfer data to relay nodes located in the human body's centre waist. The relay nodes after aggregation of data transfer to sink node. DARE gives more network lifetime but less stability period [12].

RE-ATTEMPT is discussed by A. Ahmad et al. In this protocol, the wireless sensor nodes are fixed based on their energy levels. Data transfer to sink node via minimum hop count route from the sensor nodes using multi-hop and single-hop routings. This protocol extends the network's life by reducing the stability period [13].

iM-SIMPLE discussed by N. Javaid et al. It is an extension of the SIMPLE protocol. In SIMPLE protocol, mobility of sensor nodes is not considered. This protocol, an integer linear program method, decreases energy consumption and increases throughput of a network. This network supports mobility. This protocol has higher path loss [14].

PERA protocol is discussed by G. Ahmed et al. This protocol is priority-based. This protocol is priority-based. The emergency data is first priority, and on-demand is second priority. The first and second priority nodes send information to sink node. Third priority for remaining nodes transfers data within their time slots. In this protocol, for routing, they used two nonlinear programming-based optimization models. Additionally, they also used energy levels, reception power, source rate, topology, and number of nodes design parameters. This protocol significantly improves the throughput and network's lifetime [15].

EMRP is discussed by H. Esmaeili and B. M. Bidgoli. This method used evolutionary method (multi-objective function). The multi-objective function prioritises high energy, low energy consumption, low path loss, and short communication distance for forwarder nodes. The sink node receives information from forwarder node using multi-hop routing. This method provides a high throughput, more network lifetime, and a more stability, but at the cost of increased path loss [16].

ERRS is discussed by F. Ullah et al. This protocol has two steps. In the first step, forwarder node selection is based on minimum weight factor using distance and residual energy. After data transmission, to sink node, the forwarder node reaches some threshold value and terminates the transmission for load balancing among sensor nodes. In the second step, rotation of forwarder node based on the previous weight factor which is the function of packet handling rate (PHR) and residual energy (RE). It gives a better stability period, network lifetime, and throughput [17].

3 System Model

Network model 3.1, assumptions 3.2, energy model 3.3, and path loss model 3.4 are described in this section.

3.1 Network Model

Each wireless sensor node is located in a unique location on the human body. Glucose and ECG nodes possess important data and transfer directly to sink. Remaining nodes

are participating in a selection of forwarder node. Sink node allocates time slots for sensor nodes to transfer data using TDMA protocol.

3.2 Assumptions

Nodes are initially having an equal amount of energy except the sink node. The sink node has unlimited processing power and energy. Glucose and ECG nodes are not participating in a selection of the parent node. The remaining nodes take part in multi-hop routing and choose the forwarder node.

3.3 Energy Model

We employ a first-order radio model to send and receive data. While human body moment there is some signal loss [18, 19]. The final Eqs. 1 and 2 are for transmitting and receiving and are shown in the below:

$$E_{\text{tx}}(b, d) = E_{\text{tx-ele}} \times b + E_a \times n \times b \times d^n \quad (1)$$

$$E_{\text{rx}}(b) = E_{\text{rx-ele}} \times b \quad (2)$$

E_{tx} and E_{rx} denote the energy consumed during data transmission and reception, respectively. $E_{\text{tx-ele}}$ and $E_{\text{rx-ele}}$ are the energy costs associated with transmitting and receiving a single bit. The packet size is b , and the energy required for the amplifier circuit is E_a . The path loss coefficient n has a value of 4, and d is distance. Nordic nRF 2401A dataset was used to measure transceivers [20].

3.4 Path Loss Model

The path loss model is shown as follows in (3).

$$\text{PL}(\text{fre}, e) = 10 \log_{10}((4\pi \times d \times \text{fre})^2)/c + 10n \log_{10}(d/d_0) + X\sigma \quad (3)$$

PL is the power obtained, d_0 is the reference distance with value 10, fre is frequency with value 2.4 GHz, c is speed of light in free space, X is Gaussian random variable, and σ is standard deviation [8, 21].

4 Proposed Methodology

The proposed protocol (SIMPLE-DRR) consists of three phases that are network set-up (Sect. 4.1), routing (Sect. 4.2), and scheduling (Sect. 4.3). We employed 8 sensor nodes put on the human body in this protocol, with each node having an initial energy of 0.5 J.

4.1 Network Setup

All of the wireless nodes are implanted on the human body at the appropriate points on the body. At the waist location, a sink is placed. The sink is aware of the positions and remaining energy of all nodes, as well as the distance between the sink and the nodes. Two nodes, namely, the glucosenode and the ECG node, contain critical information and send their sensed data straight to the sink. The remaining nodes are participating in the multi-hop routing procedure, which is a time-consuming process.

4.2 Routing

The wireless nodes are send information to sink using route paths. The proposed protocol, we use a multi-hop routing process for transferring information to sink. The forwarder node is used for multi-hop routing. For selection of forwarder node, we calculate cost function ($CF(i)$) (4) of i th sensor node using distance to sink node ($D(i)$), remaining residual energy ($RE(i)$), RSSI of node ($RSSI(i)$), and number of alive nodes in every round ($AN(i)$).

$$CF(i) = \frac{RE(i) * RSSI(i) * AN(i)}{D(i)} \quad (4)$$

The $CF(i)$ calculates every round of each sensor node. The minimum $CF(i)$ of sensor node acts as forwarder node. All sensor nodes communicate with the sink via forwarder. The forwarder node aggregates collected information from nodes and transfers to sink.

4.3 Scheduling

Nodes were assigned for sensor-to-sensor data transfer based on sink scheduling. TDMA protocol is used to effectively transfer sensor nodes within their time slots. We can avoid collisions among sensor node transfer data within their time slots.

5 Results and Discussions

The proposed protocol was developed in MATLAB R 2017a software running on Windows 10 operating system. We have considered four performance metrics to measure the efficiency of our protocol, namely network lifetime (Sect. 5.1), throughput (Sect. 5.2), residual energy (Sect. 5.3), and path loss (Sect. 5.4).

$$C\% = \frac{A - B}{A} * 100 \quad (5)$$

Here,

A = old protocol value, B = new protocol value, and C = final output value in terms of percentage.

Using (5), we measured parameters or metrics in terms of average stability period, network lifetime, throughput, residual energy, and path loss.

5.1 Network Lifetime

The network performance up to the first node die called stability and last node die called network lifetime. Average network lifetime of WBANs is depicted in

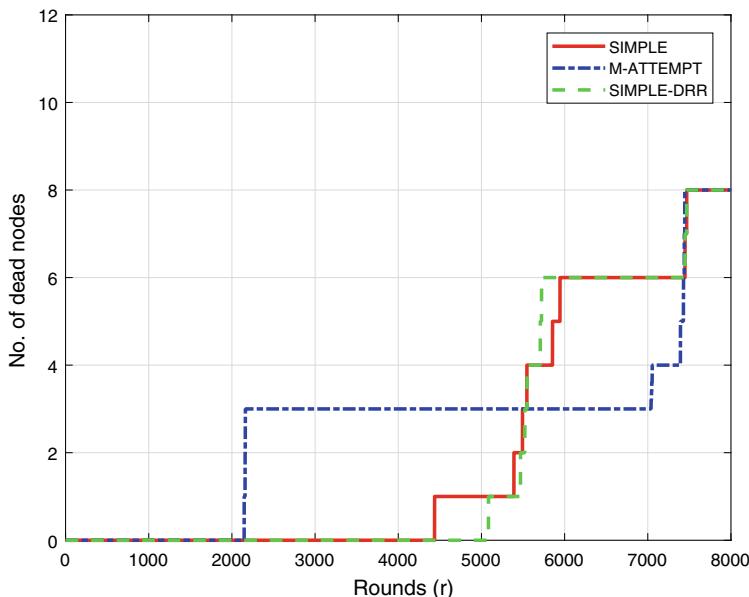


Fig. 1 Network lifetime

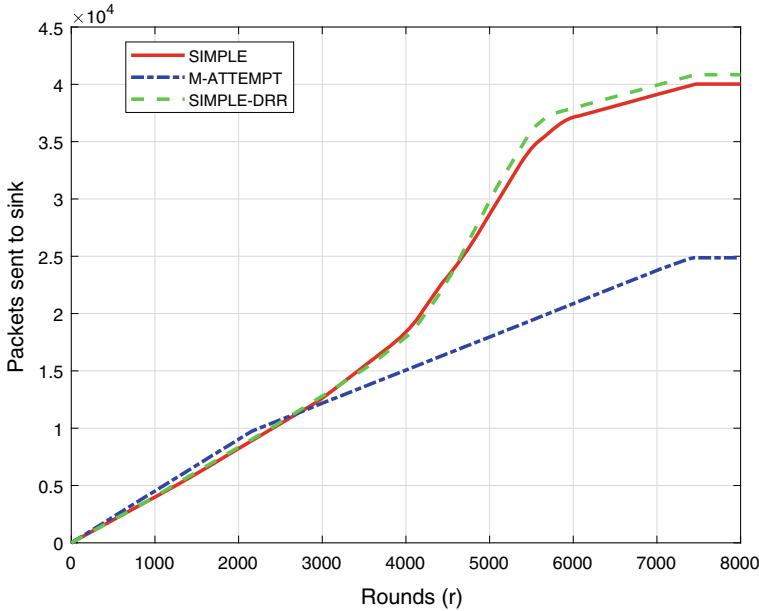


Fig. 2 Packets sent to sink

Fig. 1. This network gives a high network lifetime and stability. The stability period and average network lifetime values are calculated using (5). When compared to SIMPLE, our suggested protocol (SIMPLE-DRR) has a 14.5% stability and 2.36% average network lifetime is more. As a result, our protocol outperforms the SIMPLE and M-ATTEMPT protocols in terms of stability and network lifetime.

5.2 Throughput

Figures 2 and 3 show throughput of a network. Figure 2 is the average number of packets sent to the sink node and Fig. 3 is the no. of packets received at the sink node. By comparing with SIMPLE, our protocol gives an average of 2% more throughput. Thus, our protocol outperforms M-ATTEMPT and SIMPLE in terms of throughput.

5.3 Residual Energy

Figure 4 shows network's residual energy. By comparing with M-ATTEMPT, our protocol gives better average residual energy, and with SIMPLE it is almost (0.02%

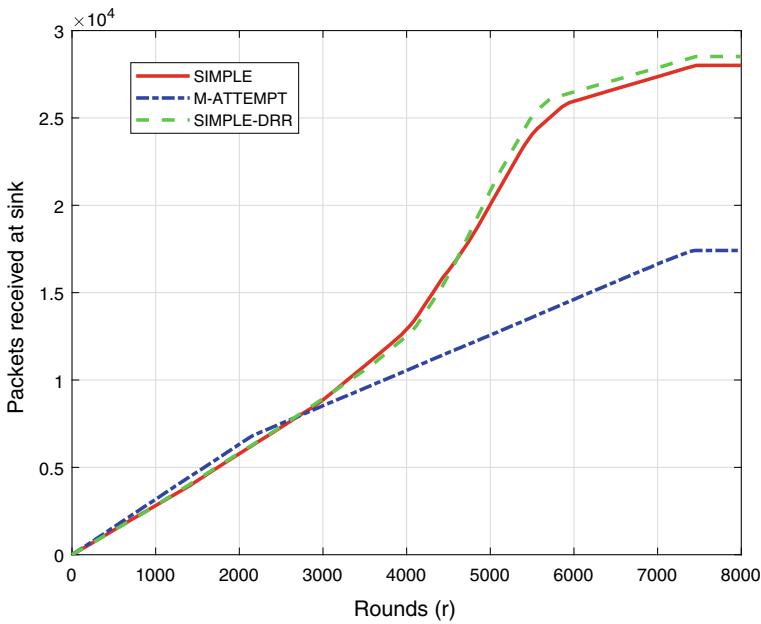


Fig. 3 Packets received at sink

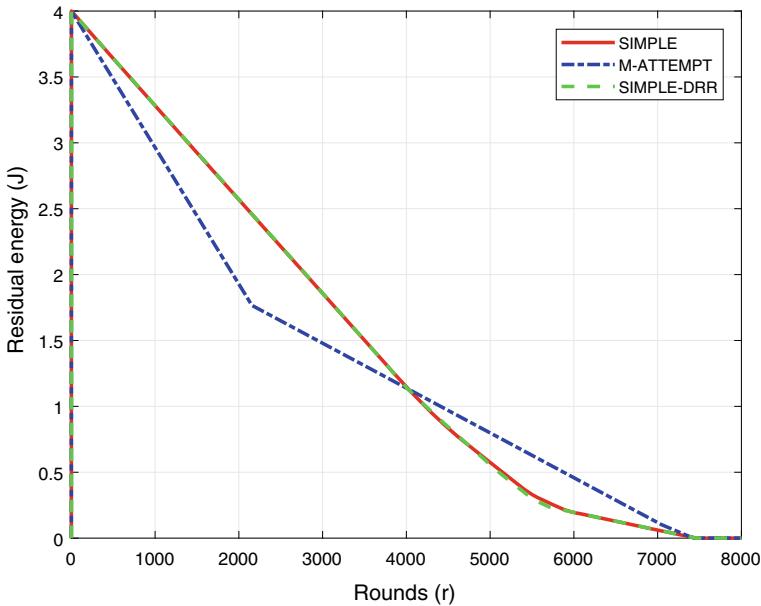


Fig. 4 Residual energy

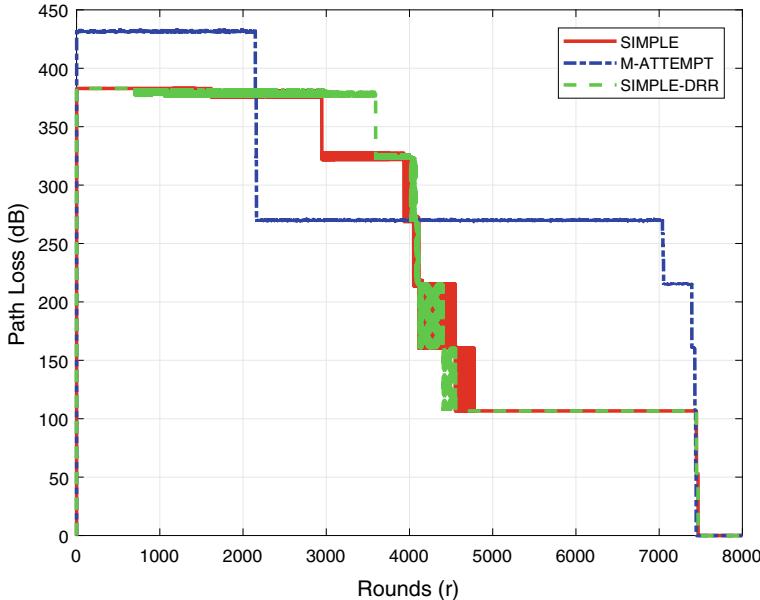


Fig. 5 Path loss

(less) equal to average residual energy. Therefore, our protocol gives average performance in terms of residual energy than M-ATTEMPT and SIMPLE protocols.

5.4 Path Loss

Figure 5 shows path loss of a network. The difference between the averages sent packets to the sink and receiving packets at the sink. To calculate path loss, we used the 3.38 path loss coefficient and 2.4 GHz frequency. By comparing with SIMPLE, our proposed protocol has 1.4% more path loss and with M-ATTEMPT gives less path loss. Therefore, our protocol gives better performance in terms of path loss than the M-ATTEMPT protocol.

6 Comparative Results

Table 1 shows stability of a network using different protocols. In the M-ATTEMPT protocol, first node dies within 3000 rounds, and in the SIMPLE protocol, first node dies within 4700 rounds. But our proposed protocol first node dies after 5000

Table 1 Dead nodes versus rounds

Protocol	No. of dead nodes at diff. no. of rounds							
	1000	2000	3000	4000	5000	6000	7000	8000
M-ATTEMPT	0	0	3	3	3	3	3	8
SIMPLE	0	0	0	0	1	2	4	8
SIMPLE-DRR	0	0	0	0	0	6	7	8

rounds. By comparing with other existing state-art protocols, the proposed protocol (SIMPLE-DRR) has a more stable period.

7 Conclusion and Future Work

Routing plays an important role in WBANs for health monitoring for improving network lifetime and stability. The lifetime of network is very important for patient health monitoring with sensor nodes. We proposed a routing protocol, called SIMPLE-DRR, for improving the stability of WBAN. In our proposed protocol, the cost function is defined based on RSSI, distance residual energy and no. of alive nodes. The cost function's minimum values work as a forwarder node. The sensor nodes, the forwarder node, and the sink node all communicate with one other. This protocol enhances a network's stability period, network lifetime, and throughput.

References

1. Joyia GJ, Liaqat RM, Farooq A, Rehman S (2017) Internet of Medical Things (IOMT): applications, benefits and future challenges in healthcare domain. *J Commun* 12(4):240–247. <https://doi.org/10.12720/jcm.12.4.240-247>
2. Latre B, Braem B, Moerman I, Blondia C, Demeester P (2011) A survey on wireless body area networks. *Wirel Netw* 17(1):1–18. <https://doi.org/10.1007/s11276-010-0252-4>
3. Movassaghi S, Abolhasan M, Lipman J, Smith D, Jamalipour A (2014) Wireless body area networks: a survey. *IEEE Commun Surv Tutor* 16(3):1658–1686. <https://doi.org/10.1109/SURV.2013.121313.00064>
4. Cavallari R, Martelli F, Rosini R, Buratti C, Verdone R (2014) A survey on wireless body area networks: technologies and design challenges. *IEEE Commun Surv Tutor* 16(3):1635–1657. <https://doi.org/10.1109/SURV.2014.012214.00007>
5. Ghamari M, Janko B, Sherratt RS, Harwin W, Piechockic R, Soltanpur C (2016) A survey on wireless body area networks for ehealthcare systems in residential environments. *Sensors* 16(6):831. <https://doi.org/10.3390/s16060831>
6. Crosby GV, Ghosh T, Murimi R, Chin CA (2012) Wireless body area networks for healthcare: a survey. *Int J Ad Hoc Sens Ubiquitous Comput* 3(3):1. <https://doi.org/10.5121/ijasuc.2012.3301>

7. Javaid N, Abbas Z, Fareed M, Khan ZA, Alrajeh N (2013) M-attempt: a new energy-efficient routing protocol for wireless body area sensor networks. *Procedia Comput Sci* 19:224–231. <https://doi.org/10.1016/j.procs.2013.06.033>
8. Nadeem Q, Javaid N, Mohammad SN, Khan M, Sarfraz S, Gull M (2013) Simple: stable increased-throughput multi-hop protocol for link efficiency in wireless body area networks. In: 2013 eighth international conference on broadband and wireless computing, communication and applications. IEEE, pp 221–226. <https://doi.org/10.1109/BWCCA.2013.42>
9. ul Huque MTI, Munasinghe KS, Abolhasan M, Jamalipour A (2013) Ear-ban: energy efficient adaptive routing in wireless body area networks. In: 2013, 7th international conference on signal processing and communication systems (ICSPCS). IEEE, pp 1–10. <https://doi.org/10.1109/ICSPCS.2013.6723991>
10. Chavva SR, Sangam RS (2019) An energy-efficient multi-hop routing protocol for health monitoring in wireless body area networks. *Netw Model Anal Health Inform Bioinform* 8(1):1–10. <https://doi.org/10.1007/s13721-019-0201-9>
11. Qureshi KN, Din S, Jeon G, Piccialli F (2020) Link quality and energy utilization based preferable next hop selection routing for wireless body area networks. *Comput Commun* 149:382–392. <https://doi.org/10.1016/j.comcom.2019.10.030>
12. Tauqir A, Javaid N, Akram S, Rao A, Mohammad SN (2013) Distance aware relaying energy-efficient: dare to monitor patients in multi-hop body area sensor networks. In: 2013 eighth international conference on broadband and wireless computing, communication and applications. IEEE, pp 206–213. <https://doi.org/10.1109/BWCCA.2013.40>
13. Ahmad A, Javaid N, Qasim U, Ishfaq M, Khan ZA, Alghamdi TA (2014) Re-attempt: a new energy-efficient routing protocol for wireless body area sensor networks. *Int J Distrib Sens Netw* 10(4):464010. <https://doi.org/10.1155/2014/464010>
14. Javaid N, Ahmad A, Nadeem Q, Imran M, Haider N (2015) iM-SIMPLE: improved stable increased-throughput multi-hop link efficient routing protocol for wireless body area networks. *Comput Hum Behav* 51:1003–1011. <https://doi.org/10.1016/j.chb.2014.10.005>
15. Ahmed G, Jianhua Z, Fareed MMS (2017) PERA: priority-based energy-efficient routing algorithm for WBANs. *Wirel Pers Commun* 96(3):4737–4753. <https://doi.org/10.1007/s11277-017-4415-9>
16. Esmaeili H, Bidgoli BM (2018) Emrp: evolutionary-based multi-hop routing protocol for wireless body area networks. *AEU—Int J Electron Commun* 93:63–74. <https://doi.org/10.1007/s11277-017-4415-9>
17. Ullah F, Khan MZ, Faisal M, Rehman HU, Abbas S, Mubarek FS (2021) An energy efficient and reliable routing scheme to enhance the stability period in wireless body area networks. *Comput Commun* 165:20–32. <https://doi.org/10.1016/j.comcom.2020.10.017>
18. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd annual Hawaii international conference on system sciences. IEEE, p 10. <https://doi.org/10.1109/HICSS.2000.926982>
19. Chavva SR, Sangam RS (2021) LeHE-MRP: leveraging health monitoring by enhancing throughput of multi-hop routing protocol in WBANs. *J Med Eng Technol* 1–6. <https://doi.org/10.1080/03091902.2021.1921066>
20. Chip S (2007) 2.4 GHz transceiver product specification. Nordic Semiconductor ASA
21. Javaid N, Khan NA, Shakir M, Khan MA, Bouk SH, Khan ZA (2013) Ubiquitous healthcare in wireless body area networks—a survey. arXiv preprint [arXiv:1303.2062](https://arxiv.org/abs/1303.2062)

Design of Class AB and Class C Amplifiers



Tejendra Kumar Singh, Deepti Kakkar, and Sukwinder Singh

Abstract This paper proposes a Class AB and Class C design at 7 GHz using the Advanced Design System (ADS) tool. For this design, Cree CG2H80015D has been used. Stability factor calculation has been carried out to ensure that the amplifier is unconditionally stable. To calculate the output impedance of the active device, load pull has been undertaken. To match the load impedance and source impedance, matching network has been designed. For Class AB, 60% PAE has been achieved, and for Class C, 72% PAE has been achieved.

Keywords Power amplifier (PA) · Class AB · Class C · Matching network

1 Introduction

Power amplifiers are an indispensable unit in any transmitter architecture. In any situation where power has to be transmitted over a channel to long distances, or a case which needs a boost in the operating power level, a power amplifier is absolutely necessary [1]. The power amplifiers are often divided into two groups. A linear power amplifier is the first, and a nonlinear power amplifier is the second. On the basis of efficiency, linearity, and circuit, this can be further broken. On the basic of conduction angle, PA can be divided into Class A, Class B, Class C, and Class AB. Class A has the highest linearity, whereas Class C has the highest efficiency. Other classes include D, E, F, G, and H.

T. K. Singh (✉) · D. Kakkar · S. Singh

Department of Electronics and Communication Engineering, Dr B R Ambedkar National Institute of Technology, Jalandhar, India

D. Kakkar

e-mail: kakkard@nitj.ac.in

S. Singh

e-mail: Sukwinders@nitj.ac.in

1.1 Class AB

This class can be considered as a compromise between Classes A and B in terms of linearity and efficiency—something in between bias point lies in between the bias points of Classes A and B [2], i.e. somewhere between the middle of the load line and the deep cut-off region. This improves the efficiency of this class as the angle of conduction is decrease, but still it is lower than 78.5%, as the conduction angle is between 180° and 360° . In Class AB amplifiers, the cross-over distortion is reduced, and hence, there is an improvement in linearity as compared to Class B. In many push-pull configurations, both the transistors are applied a gate bias such that they do not go to cut-off at the time of half-cycle switch-over. This avoids cross-over distortion.

1.2 Class C

This is another modification of amplifier biasing where below the pinch-off, the bias point is kept. Thus, the transistor, by nature, remains OFF for more than 50% of the time and thus consumes very little power ideally. It conducts less than 50% of the time and corresponds to a conduction angle of less than 180° . It has a higher efficiency than Classes A, B, and AB amplifiers but has the worst linearity.

2 Design

This section deals with the design of the Class AB and Class C PA in Keysight's Advanced Design System (ADS) environment [3]. It elaborates sequentially the steps one should undergo to design a PA that has a maximum PAE with good linearity at the design frequency.

2.1 Choice of Active Device

The active device chosen here is a 15 W GaN HEMT from Cree with the model no. CG2H80015D. It is suitable for up to 8 GHz frequency and has 28 V operation; at 4 GHz, the typical tiny signal gain is 17 dB, while at 8 GHz, it is 12 dB.

2.2 DC and Load Line Analysis

It is crucial to determine the bias point at which the transistor would be operating. This is done by performing the DC I-V tests on the device. The GaN HEMT model is tested on the FET curve tracer test bench of ADS. The gate voltage bias voltage is chosen as -3 V for Class AB and -5 V for Class C.

2.3 Stability and Bias Analysis

One of the most crucial characteristics is stability for amplification to take place. Oscillations (instability) are possible if the real portion of the input or output port impedance is negative. For a given band of frequencies, it means $|inl| > 1$ or $|outl| > 1$. Because Γ_{in} and Γ_{out} depends upon the source and the load matching networks, the stability of the amplifier depends upon ΓS and ΓL as presented by the matching networks. So, stability can be obtained if the following two conditions are satisfied simultaneously.

$$|\Gamma_{in}| < 1 \quad \text{and} \quad |\Gamma_{out}| < 1$$

If the above conditions are true for the entire frequency range of operation, then the PA is said to be unconditionally stable [4]. If these hold true for a particular band of frequencies, the amplifier is said to be conditionally stable when this happens. The $K - \Delta$ test is an important method for determining the stability of an amplifier. The Rollet's stability factor K is determined at the design frequency using a set of S -parameters. Δ is another parameter that is chosen to simplify derivations—a mathematical entity having dependency on S -parameters. The input and output impedances of an amplifier are considered to be unconditionally stable if they yield a positive real part for all passive load and source impedances. In terms of $K - \Delta$ analysis, the condition $K > 1$ and $|\Delta| < 1$ should hold true for the entire frequency range of operation.

2.4 Analysis of the Load Pull

Load pull analysis can be said to be the heart of PA design and is shown in Fig. 1. This is the most important step towards designing a high-efficiency PA and should be done with utmost care. This technique determines the optimum value of the load impedance that should be provided to the amplifiers output in order to have maximized PAE, optimum power delivered, gain, and linearity. The load reflection coefficient ΓL is defined as $\Gamma L = ZL - Z_{out}/ZL + Z_{out}$. The input impedance as seen from the load side towards the amplifier, Z_{out} is fixed and cannot be altered.

Use with S_params or Sparsm_wNoise Schematic Templates

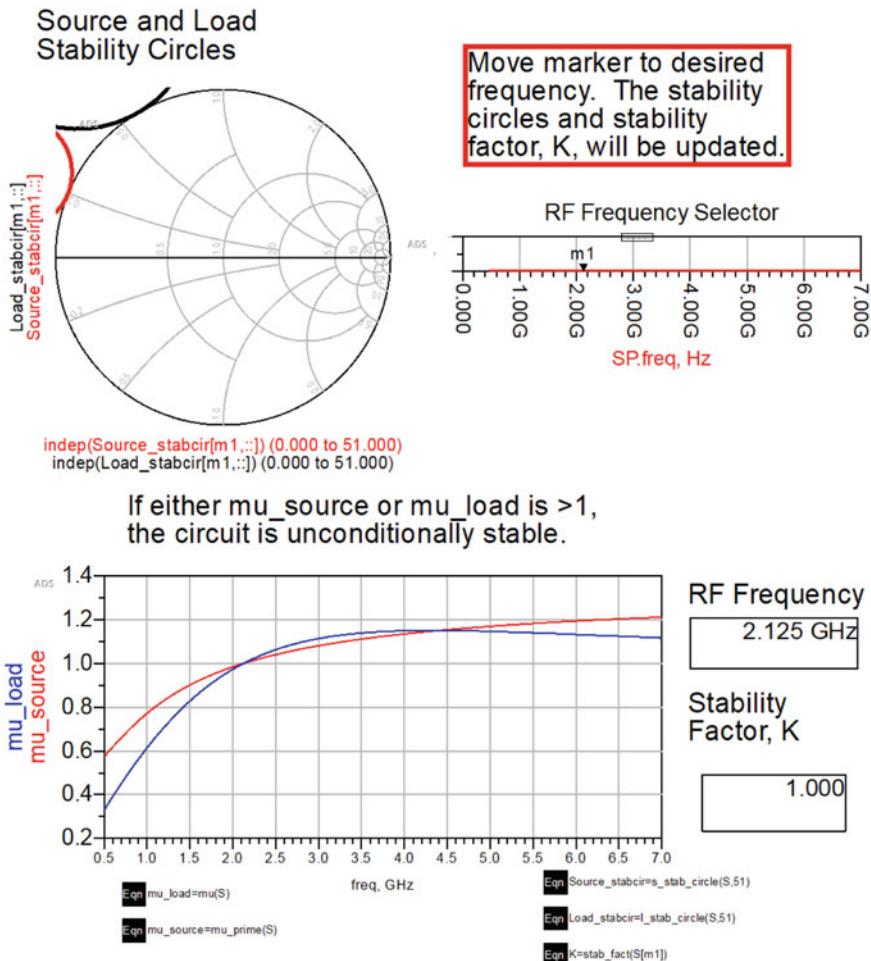


Fig. 1 Stability analysis

Thus, it is evident that the load reflection coefficient Γ_L becomes a function of the load impedance, Z_L [5–7]. Any variation in the load impedance would vary the load reflection coefficient and that is the basic philosophy behind the load pull technique. Basically, the simulator varies the load impedance over a selected area on the Smith chart and plots contours having various values of PAE and P_{del} .

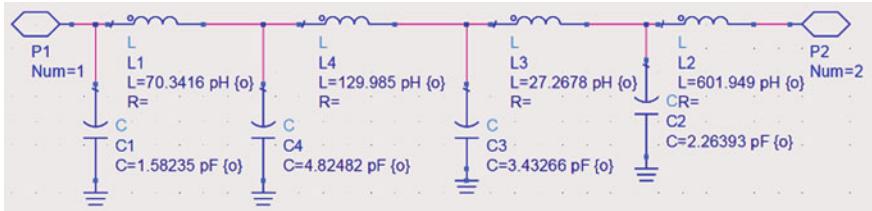


Fig. 2 Output matching network of class AB amplifier

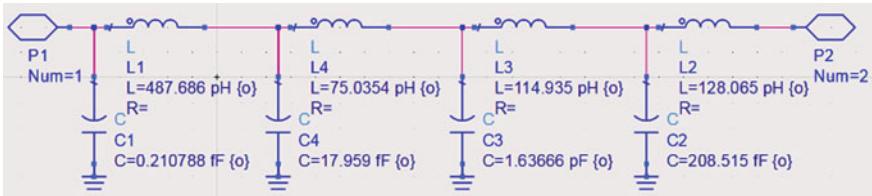


Fig. 3 Output matching network of Class C amplifier

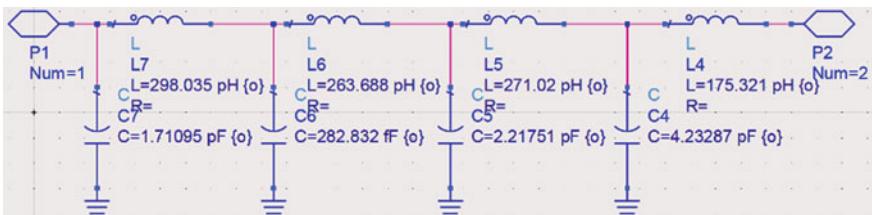


Fig. 4 Input matching network of Class AB amplifier

2.5 Matching Network

1. *Output matching network:* matching topologies exist in literature that match a complex impedance to real load. In this work, a simple wide-band matching has been used which works on the principle of admittance cancellation. The output matching network for a Class AB amplifier is shown in Fig. 2, and the output matching for Class C amplifier is shown in Fig. 3. Upon designing the network, it is connected to the drain of the amplifier.
2. *Input matching network:* following the creation of the output matching network, now attention has to be focused to make an input matching network. For this, the source pull technique is used. Source pulling is similar to load pull except that in this case, the load impedance is fixed and the source impedance is swept across a selected area on the Smith chart. The optimum point is found out again, and the input matching network is designed based on that impedance point. Figure 4 shows the input matching network for Class AB amplifier, and Fig. 5

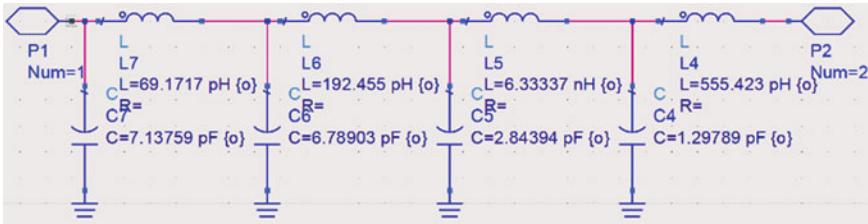


Fig. 5 Input matching network of Class C amplifier

shows the input matching for Class C amplifier.

3 Results and Discussion

For Class AB amplifier, Fig. 6 shows the output power spectrum. The fundamental power obtained is around 14.044 W at the design frequency 7 GHz, for an input power of 0.876 W. Simulated PAE is 60% PAE, drain efficiency is 64%, and large signal gain is 11.48 dB.

For Class C amplifier, Fig. 7 shows the output power spectrum. The fundamental power obtained is around 4.988 W at the design frequency 7 GHz for an input power of 0.605 W. Simulated PAE is 63%, drain efficiency is 72%, and large signal gain is 7 dB.

4 Conclusion

The main aim of this work is to design two amplifiers—one working in Class AB and another in Class C. The designed Class AB amplifier provides 14.044 W at the design frequency 7 GHz, for an input power of 0.876 W. PAE for Class AB is 60%, and drain efficiency is 64%. The designed Class C amplifier provides 4.988 W at the design frequency 7 GHz, for an input power of 0.605 W. PAE is 63%, and drain efficiency is 72%.

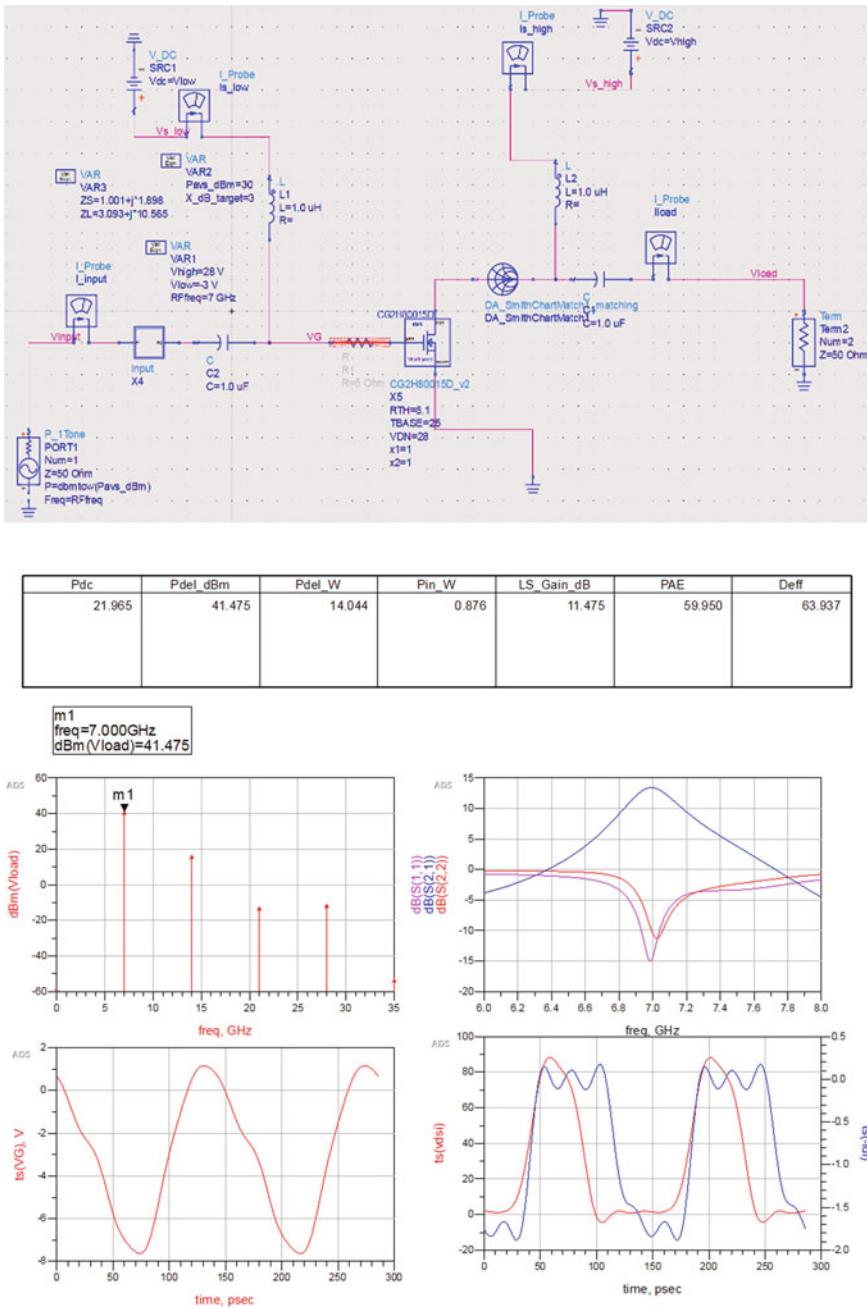


Fig. 6 Schematic and results of Class AB PA

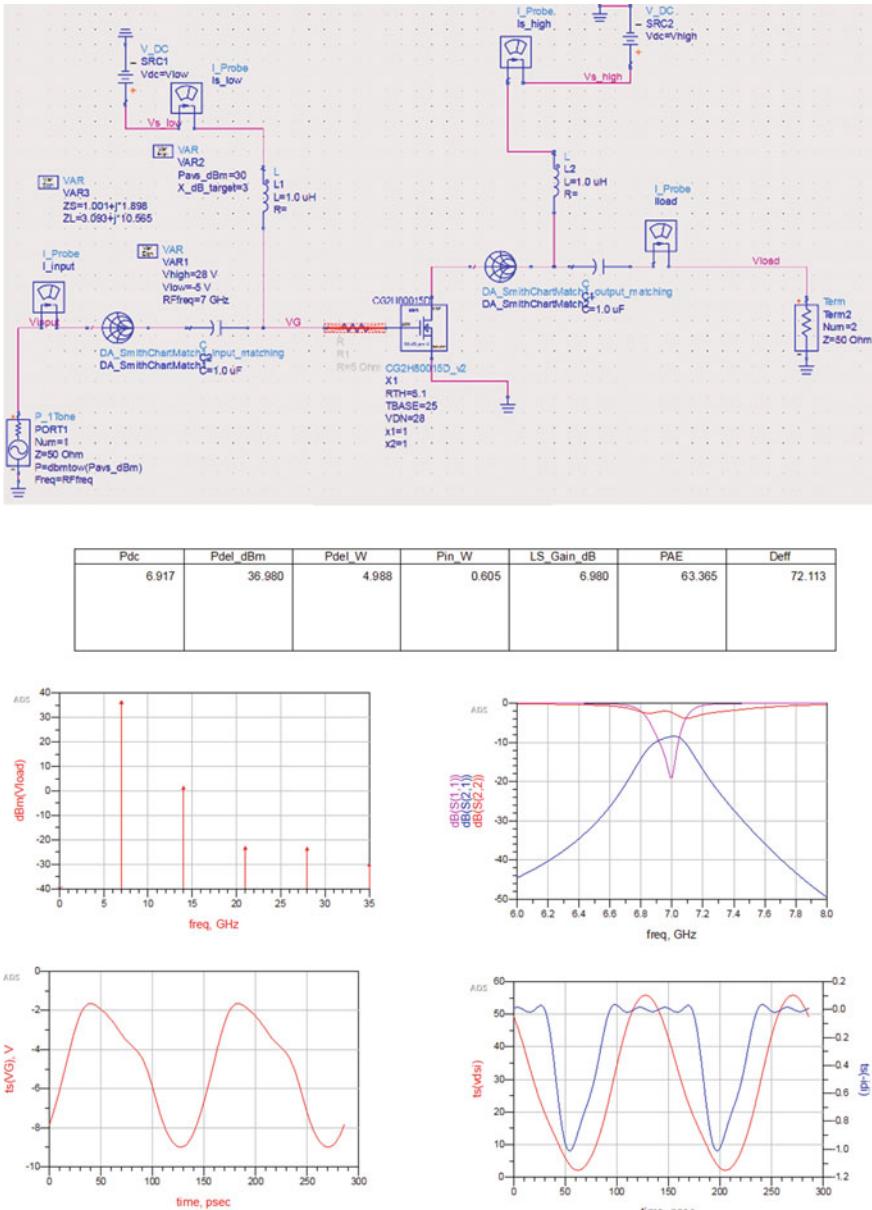


Fig. 7 Schematic and results of Class C PA

Acknowledgements For their involvement and support in our effort, we would like to thank the VLSI and Nanoelectronics Design Laboratory, Department of Electronics and Communication, Dr. B R Ambedkar NIT Jalandhar. The DST-SERB Project's (ECR/2017/000922) constant encouragement has been much appreciated. We would like to convey our gratitude and appreciation to Dr. B R Ambedkar NIT Jalandhar for providing us with the laboratory facilities that allowed us to accomplish this study in a dynamic research environment.

References

1. Cripps SC (2002) Advanced techniques in RF power amplifier design. Artech House, Norwood, MA
2. Virdee BS, Virdee AS, Banyamin BY (2014) Broadband microwave amplifiers. Artech House Inc., Boston
3. Yeom K-W (2015) Microwave circuit design—a practical approach using ADS. Prentice Hall, New York
4. Grebenikov A (2005) RF and microwave power amplifier design. McGraw-Hill, New York
5. Ghannouchi FM, Hashmi MS (2012) Load pull techniques with applications to power amplifier design. Springer, Netherlands
6. Kenington PB (2000) High linearity RF amplifier design. Artech House, Norwood, MA
7. Cripps SC (2006) RF power amplifiers for wireless communications. Artech House Inc., Norwood, MA

LightGBM Model for Credit Card Fraud Discovery



Appala Srinivasu Muttipati, Sangeeta Viswanadham, Radha Dharavathu, and Jayalakshmi Nema

Abstract The exponential growth of e-commerce and online-based payment options has created an empirical universe of financial fraud, with credit card fraud being the most prevalent. For several years, many researchers have developed a variety of data mining-based methods to address this issue. To detect credit card fraud, there has recently been a lot of interest in using machine learning algorithms instead of data mining techniques. In the digital space of financial transactions, on-going work is being conducted to put in a conceptual difference between fraud identification and predicting likely fraudulent opportunities. This paper extends the fraud detection technique and proposes a LightGBM-based detection algorithm. The dataset is a credit card dataset for credit card transactions in Europe. Our approach outperformed other traditional approaches such as random forest, AdaBoost, and XGBoost in this experiment. Furthermore, it demonstrates the value of feature engineering in terms of feature selection and performance tuning.

Keywords Credit card fraud · Machine learning · Feature selection · LightGBM

A. S. Muttipati (✉)

Department of Computer Science and Applications, KL Deemed to be University, Vaddeswaram, Guntur, Andhra Pradesh, India

e-mail: srinivasu.mutti@gmail.com

S. Viswanadham

Department of Computer Science and Engineering, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh, India

R. Dharavathu · J. Nema

Department of Computer Science and Engineering, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India

J. Nema

Department of Computer Science and Engineering, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India

1 Introduction

Nowadays, society is growing globally in all areas, and one of the areas is e-commerce. Due to the increase in e-commerce possibilities in making online payments and as they are easier to use, e-commerce business gained user confidence. This confidence leads to increase in number of users. The online transactions have given a drastic rise in revenue generation. Increase in the user's revenue generation has paved a path to be vulnerable to fraudulent behavior. Credit card fraud is one of the acclaimed problems in the present world. In 2016, there happened to be a benchmark increase in credit card fraud up to 92% compared to the 2012 count. The credit card may happen in one of the following ways: (1) application fraud, (2) stolen or lost cards, (3) account taken over, (4) card counterfeit. The stolen or lost card and account takeover are major problems and are named as card not present (CNP) fraud. In CNP, the cardholder is cheated by stealing the card's sensitive information like CVV, card No and using it remotely. It leads to the transfer of a large amount or the purchase of costly items before the cardholder discovers. As the availability of Internet is increasing in the world, people are showing interest in purchasing things online rather than offline. Due to this, the growth of e-commerce sites is increasing, and thereby the chance of credit card fraud. To solve credit card fraud, we have to find out algorithms that may either avoid or reduce credit card fraud.

1.1 Related Work

Reference [1] have suggested some ensemble models for detecting credit card fraud. Models like random forest, logistic regression, CatBoost have shown better results. The results when compared, random forest and CatBoost have outperformed and could create ROC curve and area under curve. References [2–4] have done performance comparison of naive Bayes, K-nearest neighbor, and logistic regression models in the binary classification of imbalanced credit card fraud. KNN has outperformed the competition based on all of the evaluation metrics. To identify fraudulent transactions in European credit card data, traditional algorithms such as decision tree, support vector machine (SVM) [5], least square regression, naive Bayes classifier, K-nearest neighbors (KNN), and gradient boosting (GB) have proven useful. KNN and outlier detection approaches were suggested [6] and are effective in fraud detection. They can help reduce false alarm rates and improve fraud detection rates. In an experiment, the author has tested and compared the KNN algorithm with other classical algorithms, and KNN performed well [7]. Random forest uses random tree-based and CART-based methods to train the behavioral features of standard and non-standard transactions [8–10]. Despite the fact that random forest obtained results on a small dataset, it faces the issue of imbalanced data. The focus of future work will be on resolving datasets that are imbalanced.

1.2 Our Contribution

This paper suggested a LightGBM-based credit card fraud detection algorithm. The dataset is organized based on the sequential transactions executed using credit cards by European credit cardholders. The dataset encloses a total of 284,315 transactions and is a complex dataset containing 30 variables like the difference between transaction times, transaction amount. In our work, data preprocessing to eradicate some irregular data is of the first importance. It is of great significance since some irregular data can lead to worst performance. LightGBM is executed as our twofold order. LightGBM is one of the tree-boosting framework models utilized by many data scientists to chronicle cutting-edge results to solve many machine learning issues, likewise executed other traditional models in this work like random forest, AdaBoost, and XGBoost. Experiment shows LightGBM performs better compared to other models.

2 Proposed Methodology

The proposed approach uses a three-step procedure which is stated below:

Step 1: Attaining the dataset from repository. The dataset is organized based on the sequential transactions executed using credit card by European credit cardholder. The dataset encloses a total of 284,315 transactions and is a complex dataset which containing 30 variables like difference between transaction times, transaction amount. It also contains 28 other attributes which are kept anonymous in order to protect the identity of the customer. It also contains a column with binary values ‘0’ directs non-fraudulent transaction and ‘1’ directs fraudulent transactions. One thing we can observe in the dataset is it is highly skewed. It is because the dataset is sway toward the genuine class. We can observe this as out of the 284315 transactions, only 492 are not genuine. So, only 0.172% fraudulent transactions are present when compared to whole number of transactions.

Step 2: Dataset splitting. The dataset is divided into two sets, (1) training and test set and (2) training and validation set using cross-validation. Cross-validation is a technique for evaluating a machine learning model and testing its performance. It helps in comparing and selecting an appropriate model for the precise extrapolative modeling problem. The dataset splitting can be carried out by the following steps:

1. To split the dataset into two segments: one segment for training set and other segment for testing
2. To train the model on the training set
3. To validate the model on the test set
4. Repeat Steps 1–3 until k -fold has assisted as the test set.

Step 3: The Creation of Machine Learning Models. Machine learning is categorized into four: supervised, unsupervised, semi-supervised, and reinforcement learn-

ing. The deliberated machine learning algorithms are ensemble models and gradient boosting algorithms.

3 LightGBM-Based Fraud Detection Model

This section will momentarily present our model and offers the parameters of our model. Compared with XGBoost and other traditional models, LightGBM embraces numerous enhancements like gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). Utilizing GOSS keeps all the instances with large gradients and performs arbitrary sampling on the occurrence with small gradients. In order to compensate the influence to the data distribution, when computing the information

Fig. 1 Credit card transactions time density plot

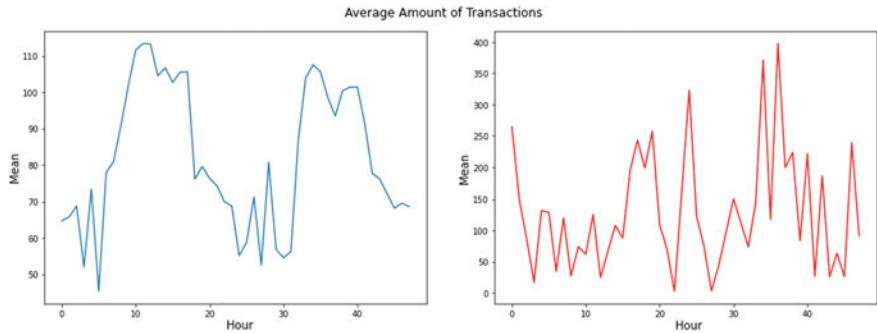
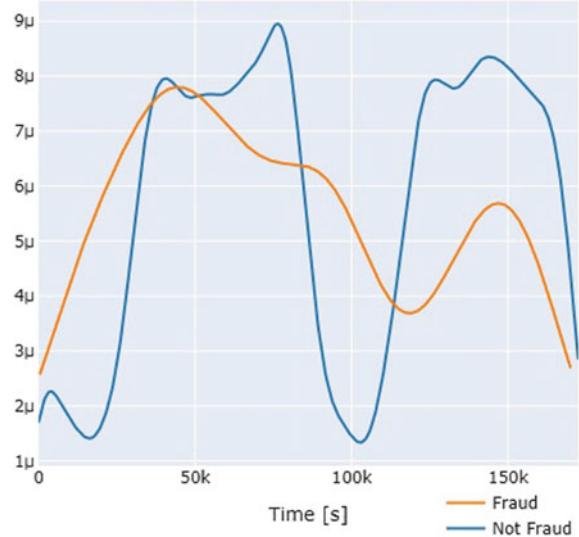


Fig. 2 Average amount of transaction over hour

Table 1 LightGBM parameters

Parameter	Parameter decryption	Values
n_estimators	Number of estimators	2000
learning_rate	Rate of learning	0.05
Num_leaves	Number of leaves < 2	7
Max_depth	Tree maximum depth	4
Min_child_samples	Minimum number of data needs in a chils	100
Max_bin	Number of bucketed bin for feature values	100
Subsample	Subsample ratio of the training instance	0.9
Boosting_type	Type of boosting	gbdt

gain, GOSS introduces a constant multiplier for the data instances with small gradients [11]. With EFB, the model's special features are to reduce the number of features and subsequently improve forecast speed. Through these optimizations, LightGBM beats the large portion of other machine learning algorithms in speed and accuracy. In view of the limits of LightGBM, we applied this model to our exploratory work (Figs. 1 and 2).

To accomplish a superior value of our model, we utilized framework search to tune the parameters of our models. Practically speaking, it is helpful in improving the score around 1 or 2%. We implemented it to the some key parameters like learning rate, completed as of not long ago. Important features for implementation are further selected using feature selection process. To give better detail, Table 1 runs down the parameters of our model, and different parameters which do not show in this table are default parameters.

4 Experimental Analysis

In this session, the experiment was performed on Windows 7 operating system and the open-source software environment. The Jupyter notebook environment is used to develop and run our model. Various libraries are utilized such as NumPy, Pandas, Matplotlib, Seaborn, Sklearn, and imblearm.

Here, AUC-ROC score proves to be the better model. This score value is actually is the area under ROC curve, which is also known as receiver operating characteristic curve value. The curve is plotted by using true positive rate (TPR) against the false positive rate (FPR) at different threshold settings. The formula of TPR and FPR are defined as follows:

Table 2 Performance of various models

Model	AUC value	Accuracy values
Random forest	0.96	0.99
AdaBoost	0.87	0.99
XGBoost	0.90	0.99
LightGBM	0.94	0.99

Table 3 Fivefold cross-validation of LightGBM model

Five folds	Training_AUC_value	Valid_AUC_value
Fold 1	0.967	0.994
Fold 2	0.977	0.962
Fold 3	0.981	0.948
Fold 4	0.970	0.987
Fold 5	0.972	0.993

$$\text{True Positive Rate} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{False Positive Rate} = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}$$

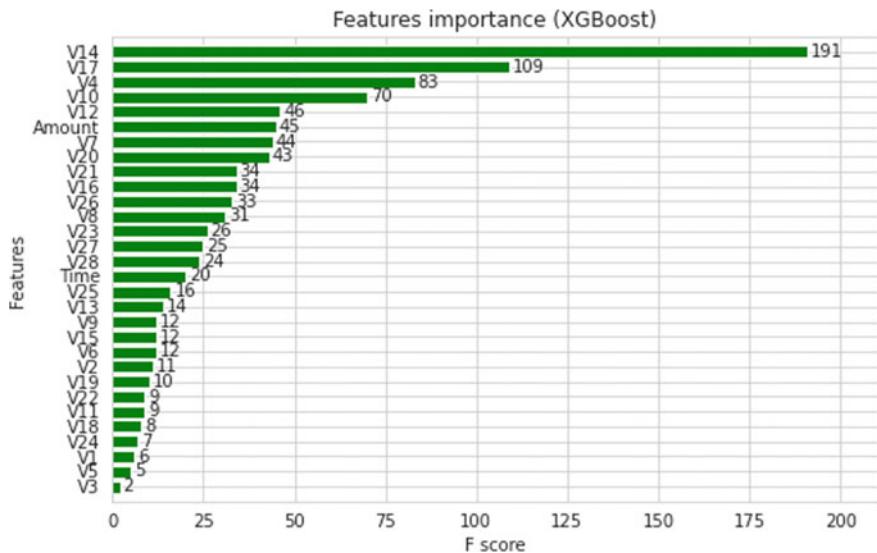
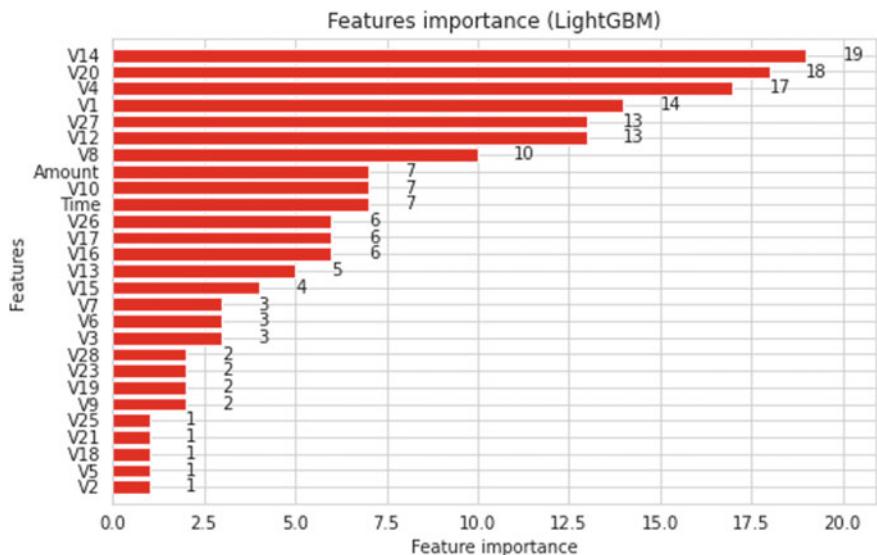
In addition to AUC-ROC value, we also provide the accuracy value of different models. In Table 2, it compared our model with other three models.

From Table 2, it is easy to find out that our LightGBM-based model outperforms the other models on both AUC-ROC value (Table 3).

Tree-based algorithms like LightGBM or XGBoost are not difficult to yield the feature significance of each feature. In Figs. 3 and 4, it shows the significant features in diminishing request. The feature significance charts give us direction on the most proficient method to implement. We can pick portions of significant features as indicated by the diagram.

5 Conclusion

This paper presents a LightGBM model to recognize fraudulent transactions. Here, we utilized both train-validation set split and cross-validation to calculate the model efficiency to forecast ‘class’ value (i.e., discovering if a transaction was fraudulent or not). In this preliminary work, comparison of various machine learning models based on metrics is presented along with identification of significant features.

**Fig. 3** Top important features of XGBoost**Fig. 4** Top important features of LightGBM

References

1. Awoyemi JO, Adetunmbi AO, Oluwadare SA (2017) Credit card fraud detection using machine learning techniques: a comparative analysis. In: 2017 international conference on computing networking and informatics (ICCNI)
2. Dhankhad S, Mohammed E, Far B (2018) Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In: 2018 IEEE international conference on information reuse and integration (IRI)
3. Dormadula VN, Geetha S (2019) Credit card fraud detection using machine learning algorithms. *Procedia Comput Sci* 165
4. Godi B, Viswanadham S, Muttipati AS, Prakash Samantray O, Gadiraju SR (2020) E-healthcare monitoring system using IoT with machine learning approaches. In: 2020 international conference on computer science, engineering and applications (ICCSEA)
5. Hema G, Muttipati AS (2021) Machine learning methods for discovering credit card fraud. *Int Res J Comput Sci* 8(1):1–6
6. Kaithekuzhical LK, Jeet Ch (2019) Detection and prediction of credit card fraud transactions using machine learning. *Int J Eng Sci Res Technol* 8(3):199–208
7. Malini N, Pushpa M (2017) Analysis on credit card fraud identification techniques based on KNN and outlier detection. In: 2017 third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB)
8. Sailusha R, Gnaneswar V, Ramesh R, Rao GR (2020) Credit card fraud detection using machine learning. In: 2020 4th international conference on intelligent computing and control systems (ICICCS)
9. Varmedja D, Karanovic M, Sladojevic S, Arsenovic M, Anderla A (2019) Credit card fraud detection—machine learning methods. In: 2019 18th international symposium INFOTEH-JAHORINA (INFOTEH)
10. Muttipati AS, Sangeeta V, Radhika S, Brahmajirao KN (2021) Recognizing credit card fraud using machine learning methods. *Turk J Comput Math Educ* 12(12):3271–3278
11. Ge D, Gu J, Chang S, Cai J (2020) Credit card fraud detection using Lightgbm model. In: 2020 international conference on E-commerce and internet technology (ECIT)

PV System-Based Switched Capacitor DC–DC Converter for High Voltage Gain Using Fuzzy-PID Controller



Balachandra Pattanaik and Yohannes Bekuma Bakare

Abstract In this paper, the maximum voltage gain switched capacitor-oriented DC–DC converter is proposed in this solar PV system. The system efficiency is decreased in grid connection with solar system due to the transformer, weight and cost. The switched capacitor-based DC–DC converter is proposed in this paper to obtain high voltage as well as avoid the transformer utilisation. In this converter, the improved voltage is regenerated using switched capacitors. The power switches in the converter system is controlled by using the fuzzy-PID controller. The DC bus bar voltage is improved by using low amount of semiconductor devices, switched capacitor and elements with low losses. The results are achieved through using the platform of MATLAB/Simulink.

Keywords DC–DC converter · High gain · PV system · High efficiency

1 Introduction

The DC–DC converter of smooth strength networks is the handiest and attractive device. In the transmission device for solar or wind energy, the voltage standard of generation is too far away from grid voltage. A transformer is primarily used after the voltage source inverter (VSI) level to step up the AC voltage to full fill the grid-connection criteria for efficient synchronisation [1–11]. Especially, in poor environmental conditions, the transformer expands the volume and misfortunes of the framework, mass in size and diminishes generally effectiveness [1, 2, 12]. Many studies have concentrated on replacing the transformer in recent decades and developed multilevel cascaded converters. It does, however, raise system numbers, management performance and quality penalties. The converters consist of power semiconductor elements

B. Pattanaik (✉) · Y. B. Bakare
Electrical and Computer Engineering, Engineering and Technology, Wollega University, Post Box Number: 395, Nekemte, Ethiopia
e-mail: balapk1971@wollegauiversity.edu.et

Y. B. Bakare
e-mail: yohannesb@wollegauiversity.edu.et

for purposes with higher and lower power [4–11, 13]. The converter topologies are utilised and function like DC–DC, AC–DC, AC–AC and more to increase the power provided from either the sources [14]. The traditional buck converter circuit, due to its low price and simple form, has been used to increase DC link voltage in which the load is associated also with converter system [3–5]. Furthermore, the voltage gain for high-step applications is not sufficient [15]. There are additional difficulties with converters, including such reduced efficiency, huge capacitors and a huge number of passive components [7–11, 13, 15–17]. In comparison with SEPIC, ZETA and Cuk converters, the landsman converter is superior; then, it reduces operational components and overcomes the previously described problems [6–8]. Regularly utilised in medium force applications, the exchanged inductor (SI) type converters diminish the current wave and alleviate the issues related with the throbbing current [14–20]. The switching devices are regulated utilising a control system [9] to achieve significant voltage gains from supplied voltage. Traditional systems are managed and maintained by PWM, PI and PID to give gate pulses with the proper operating cycle for the switching devices. The system performance is accomplished by means of a fuzzy-PID control for the suggested converter, which improves the voltage of the DC connection and decreases the adjustment times to retain the high voltage of both the VSC stable [10, 11, 13, 16, 17].

In this proposed system, the high voltage gain converter with regenerative operation of switched capacitor, and this converter system is powered from the solar system. The high-step gain DC–DC converter is suggested in this paper by integrating the switched-capacitor design and the regenerative boost. In this converter system, the transformer is not used which leads to reduce the size of the converter and low cost with high efficiency [18, 19]. The PV system is generating low amount of power which not sufficient for driving load so that proposed converter is used to accomplish high power, and the DC link voltage is maintained at constant through using control method of fuzzy-PID controller.

2 Proposed System

In this system, the combination of switched capacitor with regeneration function converter is used to improve the voltage of the DC link. The DC voltage gain is greatly supported by the switched capacitor-based system. During the conduction state of the power switches, the inductor and switched capacitors are used to regenerate the improved voltage. In the switches off state, the energy of the reactive components is discharged in cascaded form. Because of this performance, the DC link is increased extremely. The circuit diagram of switched converter is shown in Fig. 1. In this converter system, which contains two active powers switches S1 and S2, the converter step up performance is achieved through the inductor L1 and capacitor C1. According to the proposed structure, into the C2 and L2, the improved voltage is regenerated and further stored. The switches S1 and S2 controlled the energy stored in C2 and L2, respectively.

Fig. 1 Circuit diagram of the proposed converter

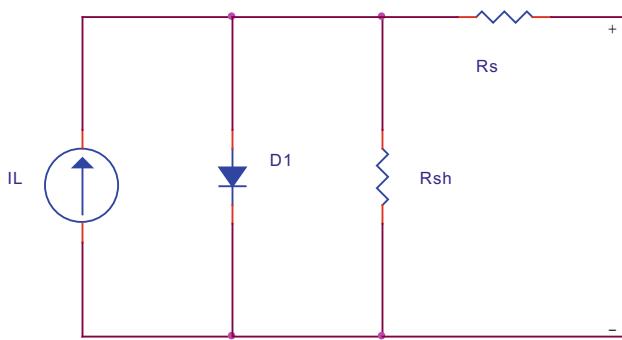
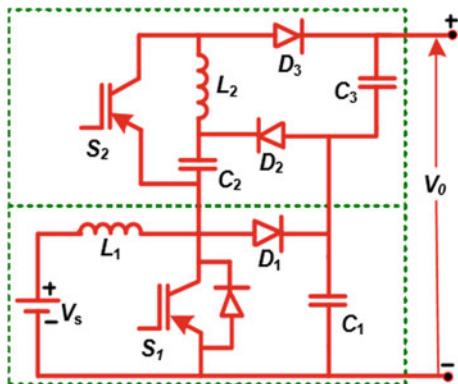


Fig. 2 PV single-diode model

3 PV Source

The photovoltaic system generates DC for both the power grid with sliding mode for both the boost converters (SMC). As illustrated in Fig. 2, to create power out of solar irradiation, a single-diode PV module is presented. The resistance are linked to the diodes by the power supply in series and parallel.

4 Control Method

This article comprises two switching devices (S) regulated with the use of the fuzzy-PID-controller inside the planned PV supplied converter operates. The fuzzy-PID controls for landsman converters powered brushless DC motorised solar system is created using this presented technology. The offered control returns even for the

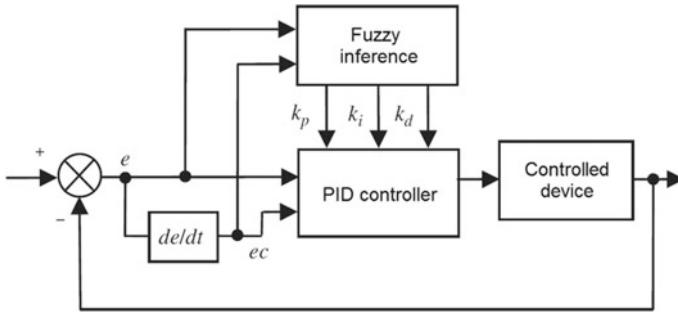


Fig. 3 Fuzzy-PID controller block diagram

fluctuation of the breakpoints to the appropriate set point. The controlling engineering incorporates fluffy rationale just as controllers. A PID regulator as a rule has three credits of relative, necessary and subordinate qualities. The fluctuations may be minimised with this controller architecture and the program's constant state inaccuracy managed. The presented fuzzy-PID regulation is utilised to optimise and supply switching pulses towards the proposed converter main switch, whenever the dynamical situation is both loaded and target value. Variables K_p, K_i and K_d are defined utilising fuzzy logics throughout this scheme. There are two parameters that delta errors and errors and 3 outputs of K_p, K_i and K_d for the developed fuzzy-PID controller. Used with the setup time minimised is indeed the fuzzy-PID regulator. By utilising the developed fuzzy-PID controller, the inverter output is enhanced. The circuit schematic for the proposed DC-DC landsman conversion fuzzy-PID controller architecture is illustrated in Fig. 3.

5 Simulations and Results

In this proposed framework, the high voltage acquire converter with regenerative activity of exchanged capacitor, and this converter framework is fuelled from the planetary group. In this converter system, the transformer is not used which leads to reduce the size of the converter and low cost with high efficiency. The PV system is generating low amount of power which not sufficient for driving load so that proposed converter is used to accomplish high power and the DC link voltage is maintained at constant through using control method of fuzzy-PID controller. The PV system power is shown in Fig. 4. The switched capacitor-based converter is utilised to achieve the high DC bus bar voltage as shown in Fig. 5.

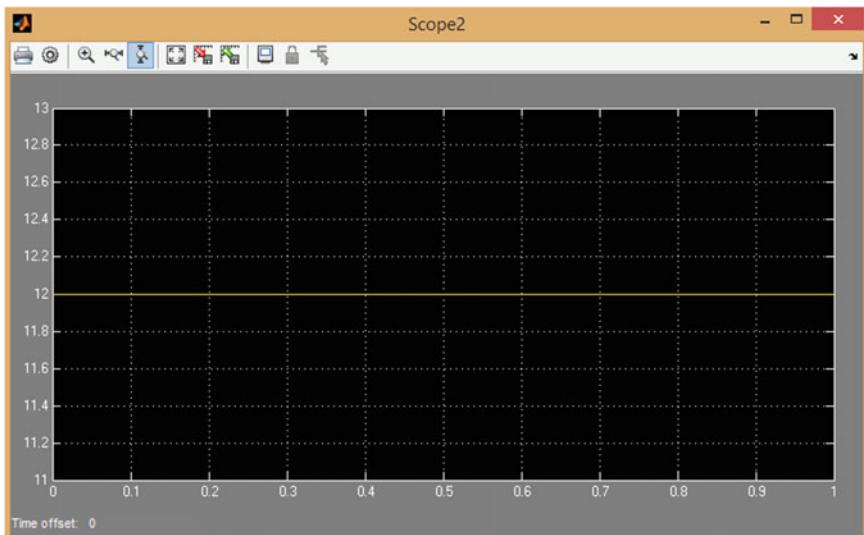


Fig. 4 Input supply of PV system

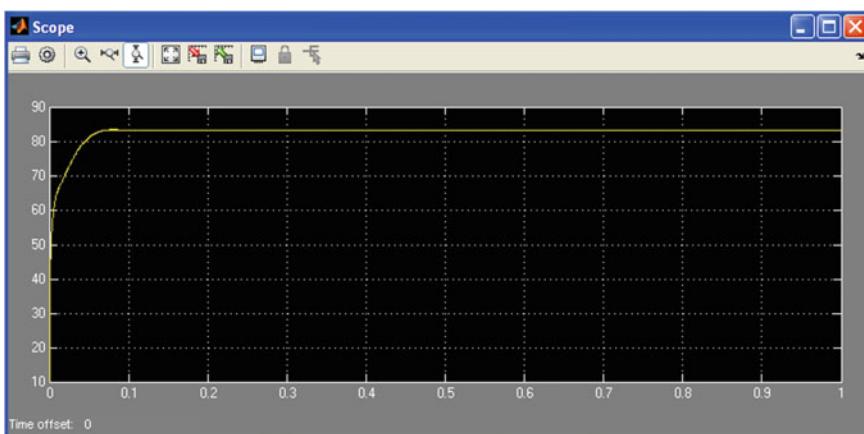


Fig. 5 Proposed converter DC link voltage

6 Conclusion

High power is collected from the photovoltaic system with the power conversion of converter that converts from DC to DC is presented in this article which has the combination of regenerative step up and switched capacitor. The PV system is generating low amount of power which is not sufficient for driving load so that proposed converter is used to accomplish high power. The DC transport voltage is controlled

to, and it is taken care of to load or network associated framework. To control the proposed method, the fuzzy-PID control is used, and it is controlling the duty cycles which directs to converter DC voltage. Compared to traditional technologies, the converting ratio has increased. The electricity extraction out from solar system is performed, and the results are effectively checked with regulated voltage to the network system.

References

1. Kivimäki J, Kolesnik S, Sitbon M, Suntio T, Kuperman A (2017) Design guidelines for multi-loop perturbative maximum power point tracking algorithms. *IEEE Trans Power Electron* 33(2):1284–1293
2. Li X, Wen H, Jiang L, Xiao W, Du Y, Zhao C (2016) An improved MPPT method for PV system with fast-converging speed and zero oscillation. *IEEE Trans Ind Appl* 52(6):5051–5064
3. Pattanaik B (2020) SEPIC converter with PV system BLDC motor control for bidirectional power grid using sliding mode controller. *J Crit Rev* 7(11):3962–3967. ISSN: 2394-5125
4. Darwish A, Alotaibi S, Elgenedy MA (2020) Current-source single-phase module integrated inverters for PV grid-connected applications. *IEEE Access* 8:53082–53096
5. Priyadarshi N, Bhaskar MS, Padmanaban S, Blaabjerg F, Azam F (2020) New CUK–SEPIC converter based photovoltaic power system with hybrid GSA–PSO algorithm employing MPPT for water pumping applications. *IET Power Electron* 13(13):2824–2830
6. Manie N, Pattanaik B (2019) Zeta DC–DC converter based on MPPT technique for BLDC application. *Int J MC Square Sci Res* 11(2):1–12
7. Sureshkumar G, Kannan N, Thomas S, Karthikeyan SP (2019) Matlab/simulink simulations based modified sepic DC to DC converter for renewable energy applications. *J Adv Res Dyn Control Syst* 11(4):285–295
8. Pattanaik B, Babu G (2017) Reliability enhancement of renewable energy using programmable system on chip model. *IJPT* 9(1):28405–28411. ISSN: 0975-766X
9. Pattanaik B et al (2021) Energy storage system with dynamic voltage restorer integrated for wind energy system. *J Phys: Conf Ser* 1964(4):042098
10. Singh AK, Badoni M, Tatte YN (2020) A multifunctional solar PV and grid based on-board converter for electric vehicles. *IEEE Trans Vehic Technol* 69(4):3717–3727
11. Komurcugil H, Biricik S, Guler N (2019) Indirect sliding mode control for DC–DC SEPIC converters. *IEEE Trans Ind Inf* 16(6):4099–4108
12. Ramaprabha R, Balaji K, Raj SB, Logeshwaran VD (2013) Comparison of interleaved boost converter configurations for solar photovoltaic system interface. *J Eng Res [TJER]* 10(2):87–98
13. Pattanaik B et al (2021) Industrial speed control of IM based model predictive controller using zeta converter. *J Phys: Conf Ser* 1964(6):062075
14. Heydari M, Khoramikia H, Fatemi A (2019) High-voltage gain SEPIC-based DC–DC converter without coupled inductor for PV systems. *IET Power Electron* 12(8):2118–2127
15. Pattanaik B et al (2017) Reliability enhancement of renewable energy using programmable system on chip model. *IJPT* 9(1):28405–28411. ISSN: 0975-766X
16. Pattanaik B, Taye A, Umer AA, Henok G (2021) IoT based over current protection control system using for micro grids. In: 2021 6th international conference on inventive computation technologies (ICICT), Coimbatore, India, 2021, pp 235–239. <https://doi.org/10.1109/ICICT50816.2021.9358511>
17. Arunkumari T, Jagadeesh I, Indragandhi V (2019) Design and implementation of modified multilevel sepic converter for PV applications. *Indones J Electr Eng Comput Sci* 14(3):1125–1133

18. Pattanaik B et al (2020) Maximum intermediate power tracking for renewable energy service. Lecture notes on data engineering and communications technologies. Springer, Switzerland, pp 462–468. ISSN: 2367-4512
19. Pattanaik B et al (2021) Prediction of the speed and wind direction using machine learning. *J Phys: Conf Ser* 1964(4):042064
20. Hussain SM, Pattanaik B et al (2021) Sliding mode controller for electric vehicles based on switched boost converter. *J Phys: Conf Ser* 1964(4):042093

Start-up Circuits for Zero Current Reduction in Bandgap Reference Circuits



Anushree, Anushka Dixit, Nandini Singh, and Jasdeep Kaur

Abstract This paper presents a bandgap reference circuit which produces a reference voltage of 685 mV with the temperature coefficient of 7.835 ppm/°C for the temperature range of –40 to 150 °C. The proposed bandgap reference circuit is operating at the supply voltage of 1.8 V with power consumption of 1.74 µW. Further, it includes study of various start-up circuits such as their role, working, applications, advantages, and limitations. These circuits are used alongside the bandgap reference voltage generator circuit for its proper operation. Majorly, small-sized start-up circuits are designed for low voltage applications because these circuits tend to reduced total power consumed by the circuit. In proposed work, variety of start-up circuits such as start-up circuits with feedback mechanism, op-amp-based start-up circuits, inverter-based start-up circuits, switch-based start-up circuits, etc., are discussed.

Keywords Start-up circuit · BGR · Complementary to absolute temperature (CTAT) · Proportional to absolute temperature (PTAT)

1 Introduction

In today's world of electronics, the reference voltage generators have wide range of applications such as D/A converters, A/D converters, analog circuits, voltage regulators, and digital circuits. Similarly, in battery operated portable electronic devices, low voltage and low power applications are widely used. Bandgap reference circuit is one which produces reference voltage or reference current independent of supply voltage, temperature, and process parameters. Device with long length is selected, in order to limit the effect of short channel, high output resistance, achieve high signal gain, lesser variation of threshold voltage, and to improve overall PSRR. Also,

Anushree (✉) · A. Dixit · N. Singh · J. Kaur
Indira Gandhi Delhi Technical University for Women, New Delhi, India

J. Kaur
e-mail: jasdeep@igdtuw.ac.in

bandgap reference circuits are majorly self-biased circuits, i.e., no external supply is provided and that is where the “start-up circuits” came into picture. Start-up circuits can be generic as well as specific with respect to the bandgap reference generator circuit.

In works presented in [1–3], different bandgap reference circuits are presented. In [1], the BGR circuit is formed with a bipolar transistor, current reference circuit which produces current in nano-amperes range and a PTAT voltage generator. A low voltage, current mode-based bandgap reference circuit is presented in [2], which is working in two stable operating points. This can be initially started simply by using a pulse generator circuitry and a power signal. In [3] proposed BGR and sub-BGR circuits, formed with a current reference circuit which produces nano-ampere current, a bipolar transistor, and PTAT voltage generator using only MOSFET and bipolar transistor are discussed.

The paper is organized as follows: literature surveys of various start-up circuits that are used in BGR circuits are presented in Sect. 2. Further, Sect. 3 addresses the designing methodology of the proposed bandgap reference circuit with a start-up circuit along with its simulation results. Finally, Sect. 4 depicts the conclusion of the paper.

2 Literature Survey

2.1 *Role of Start-up Circuits in BGR*

A bandgap voltage reference circuit has two stable operating regions, i.e., no current region and normal operating region. For understanding the role of start-up circuit, Fig. 1 depicts a current mirror-based BGR circuit. Here, if the value of node voltages (node-1 and node-2) is equal, then the current flowing in both the branches of the current mirror will also be equal and hence, PTAT voltage, i.e., proportional to absolute temperature voltage and CTAT voltage, i.e., complement to absolute temperature are generated across R_2 and Q_4 , respectively. The obtained reference voltage is addition of PTAT and CTAT voltage. This region is termed as normal operating region.

The problem arises when the voltages at the node-1 and node-2 are not equal, i.e., at node-1, the voltage is V_{dd} volts whereas at node-2, the voltage is 0 V. Now, this combination will lead to zero current in the branches of the current mirror and no reference output voltage will be generated. Hence, this region is termed as zero current region or no current region. Once BGR enters in this state, it will remain here forever, even while simulating the BGR circuit. Also, it is very difficult to identify whether BGR is operating in the zero current region or normal operating region. As a result of this, BGR produces improper result which leads to further affect the performance of circuits driven by bandgap reference circuit. Hence, in order to draw up the circuit from operating in no current region and make them work in the normal operating region, start-up circuits are used.

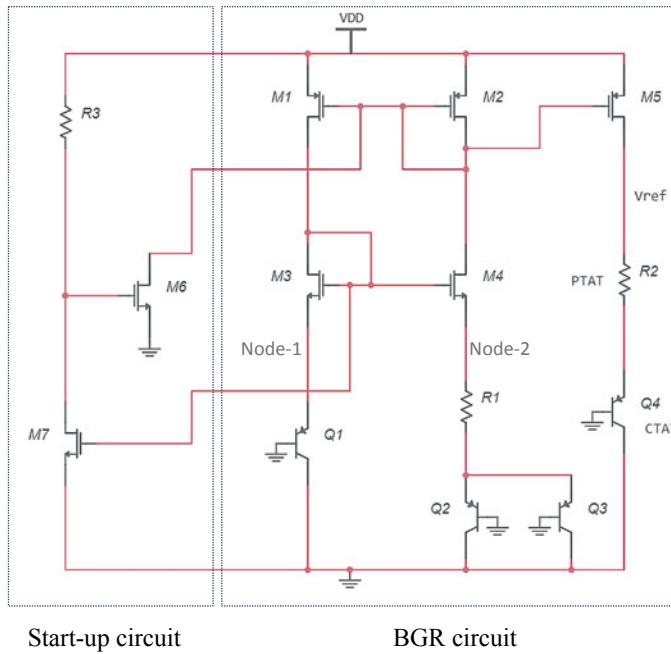


Fig. 1 Schematic of a current mirror-based BGR circuit with the start-up circuit

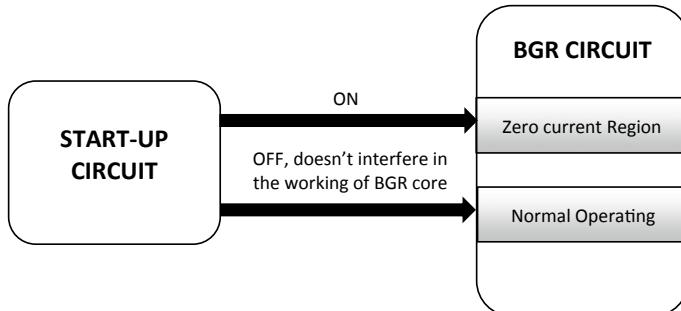


Fig. 2 Block diagram showing the working of a start-up circuit

When the BGR circuit (or BGR core) is operating in zero current region, the start-up circuit is activated and it will supply a constant voltage value to both the nodes of the BGR core, so that it starts working in the normal operating region. However, when BGR circuit is operating in the normal region of operation, the start-up circuit does not interfere. In Fig. 2, block diagram representing the working of start-up circuit is presented.

2.2 Different Start-up Circuits

Start-up Circuit with Feedback Mechanism: In [4], a simple and small-sized start-up circuit for a low voltage bandgap reference circuit is presented that uses a feedback mechanism to generate the start-up voltage. The feedback mechanism generates the start-up voltage if the output of the BGR core is below the expected voltage level. When this expected value has reached the start-up voltage, start-up circuit turns off and start-up operation is finished.

In Fig. 3, M3 and M4 form a current mirror and transistor M2 forms a MOS capacitor between the drain-source-body junction and gate terminal. At the DC steady state, $V = V_{DD}$, as M1 stops charging and it starts acting as open circuit.

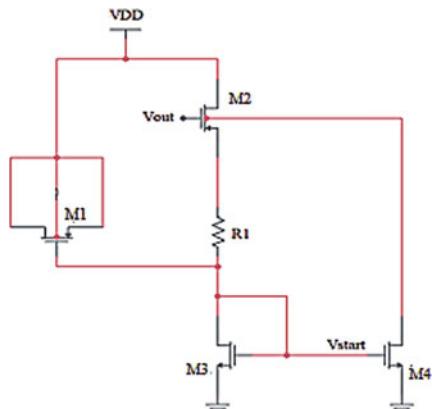
Start-up circuit for low supply voltage operation: In [5], self-biased op-amp-based start-up circuit designed using one NMOS transistor so this is ideal to use in low supply voltage operation.

In Fig. 4, transistors, namely M1, M2, M3, M4, M5, M6, and M7, contribute in the generation of self-biased op-amp circuit and transistor M8 ensures that the op-amp produces stable-biased current until output of BGR core keeps invariable. $V_{in}(-)$ and $V_{in}(+)$ pins will be connected to the BGR core.

Start-up circuit mainly for op-amp-based low power applications: In [6], designed start-up circuit is used especially for op-amp-based BGR core. This start-up circuit consumes zero power at the steady state. Figure 5 represents a start-up circuit, which is mainly used in different low power applications.

In Fig. 5, M1 and M4 form an inverter. When the power is OFF, the output of the inverter would be HIGH and this logic HIGH output will turn ON the transistors,

Fig. 3 Start-up circuit with feedback mechanism



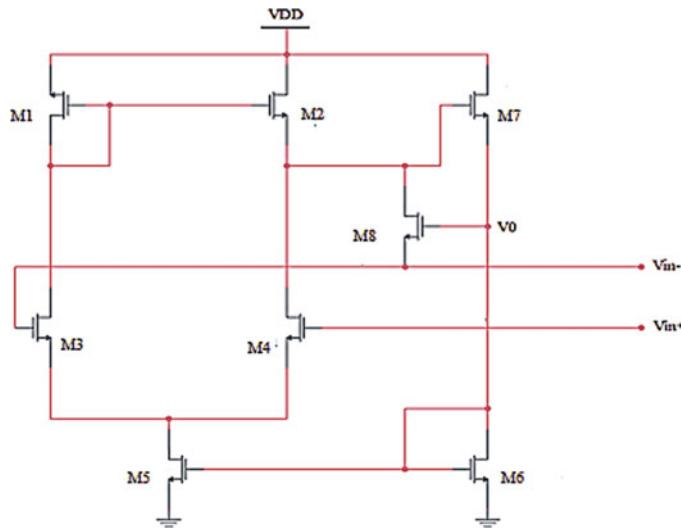
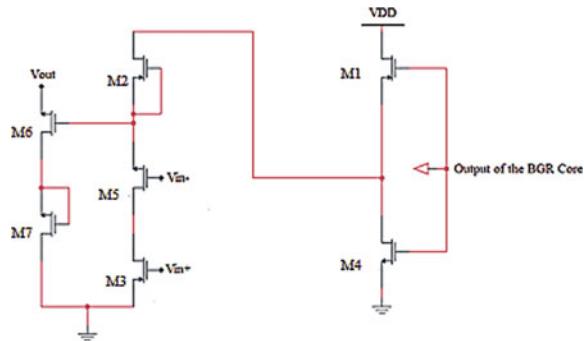


Fig. 4 Start-up circuit for low supply voltage operation

Fig. 5 Start-up circuit for op-amp-based low power applications

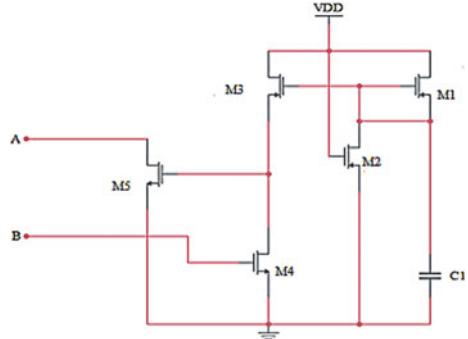


M2, M5, and M3, respectively. Hence, a small amount of current will start flowing into the BGR core. When the power is ON, the output of the inverter would be LOW and the transistors (M2, M5, and M3) will be turned OFF. Also, the transistors, M6 and M7, are placed to ensure that M2, M5, and M3 are successfully turned OFF when the power goes ON. V_{in+} and V_{in-} are connected to op-amp inverting and non-inverting terminals. Here, V_{out} is op-amp's output.

Switch-based start-up circuit: In [7], switch-based start-up circuit is presented where M5 acts as a switch driven by the voltage provided by BGR core. Figure 6 represents switch-based start-up circuit.

When the power is turned ON, the capacitor C1 starts charging by a current, as a result of this transistor M3 of a current mirror circuit charges the gate of transistor M5 to turn it on. Hence, M5 transistor introduces a slight amount of current into

Fig. 6 Switch-based start-up circuit

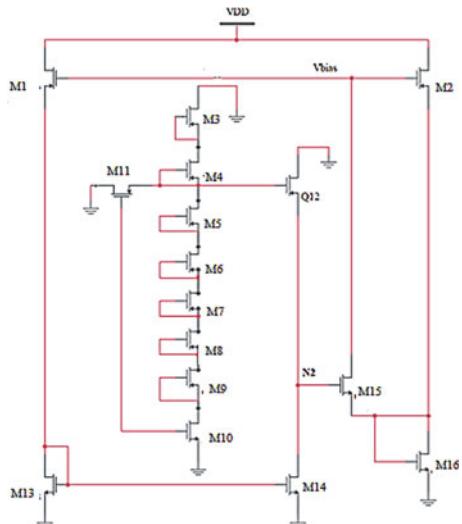


the bandgap core. After start-up operation, M5 transistor starts operating in cut-off region and M4 is turned on. When capacitor C1 is charged to a threshold value below the supply voltage, M1 and M3 will be in cut-off. When power supply is switched off, the transistor M2 discharges capacitor C1. The output from M5 is connected to node A, which is the output voltage terminal of the op-amp present in the BGR core and the gate terminal of M4 is connected to the node B, which is the input terminal of op-amp in the BGR core.

Current mirror-based start-up circuit: Cao et al. [8] indicate a start-up circuit which is used for low power BGR and designed using a current mirror having wide operating voltages with improved power up characteristics.

In Fig. 7, current mirror circuits are formed, one is between M1 and M2, and the other is between M13 and M14. Transistors M3 and M4 are used to enhance the power up characteristics of start-up circuit. Here, V_{DN2} , i.e., voltage produced at the

Fig. 7 Current mirror-based start-up circuit



node N2 is constant. As a result, conditions of the transistor M15 are unchanged and there would be no decrease in the reference voltage produced by BGR core and hence, a stable reference voltage is obtained. Also, 5 PMOS transistors, namely M5, M6, M7, M8, and M9 are connected in series for minimization of the total stand-by current.

Start-up circuit with multiple options for power supplies: Boni [9] presents a start-up circuit which works on multiple power supply circuits. In Fig. 8, start-up circuit with multiple options for power supplies is presented. This circuit comprises 3 PMOS transistors, namely M1, M2, and M3 which can be driven by using either of the two power supplies, i.e., (I) or (II). Figure 9 shows power supply circuits used. In Fig. 9a, an external power on reset signal (POR signal) is required, whereas, in Fig. 9b, no external POR signal is required for pushing a constant current into the bandgap core.

Other techniques: Start-up circuit of [10] consists of a capacitor that controls the transient behavior of pre-regulator so that no abrupt variations can be caused in the load current. He et al. [11] proposed a robust start-up circuit that generates start-up voltage for BGR and for the temperature coefficient calibration according to the operating states of BGR. Also explains problems faced while its designing was performed. Start-up circuit presented in [12] consists of BJT and NMOS transistors

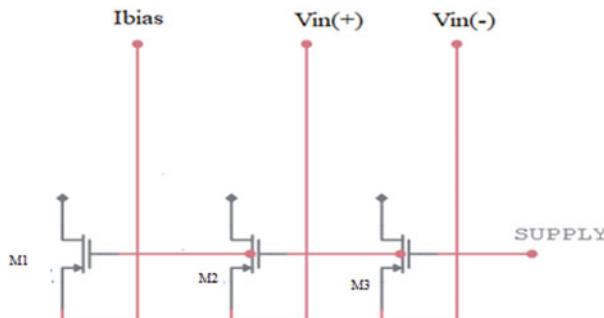
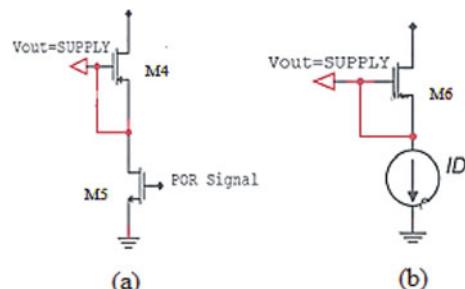


Fig. 8 Start-up circuit with multiple options for power supplies

Fig. 9 Options for power supply circuit



and its start-up time is 8 microseconds. A self-start-up technique is presented in [13] which uses the principle of cross coupled BGR structure for positive start-up loop. In [14], a different self-start-bias configuration has been discussed. This uses soft start-up for BGR which works on the principle of negative feedback.

3 Proposed Bandgap Reference Circuit with a Start-up Circuit

3.1 Designing of Bandgap Reference Circuit

The circuit depicted in Fig. 1 is a general current mirror-based BGR circuit which includes start-up circuit for eliminating zero current problem. The basic function of this circuit is to provide the overall constant reference voltage, i.e., V_{ref} (in the figure) which do not alter its value with the alterations in various parameters, like temperature and supply voltage. M6, M7, and R3 constitute for the start-up circuit. If the node potential between NMOS M3 and M4 is zero, they will not allow any current to pass through them. Similarly, if the node potential between PMOS M1 and M2 is V_{DD} , again will represent zero current condition. Start-up circuit is used such that it will not disturb normal operation of BGR circuit. To ensure the normal operation, a voltage drop is needed in the node potential of M2. Also, an increased node potential is ensured for M4.

Initially, for zero current condition, the node potential between M3 and M4 is zero, resulting into OFF condition of M7. Hence, no current is flowing across M7. For this condition of no current, the potential drop across R3 is zero which further results into ON condition of M6. This will lower down the node potential for M2 and will start working normally. In this way, zero current condition can be eliminated. If the circuit is operating normally, i.e., non-zero node potential between M3 and M4, it will allow current to pass through M7 resulting into voltage drop in node potential between R3 and M7. If this potential drop is lower than the V_{th} , M6 will be in OFF condition.

When start-up circuit is included in proposed bandgap reference (BGR) circuit, M2 and M3 are not working in zero current state any more. Start-up circuit ensures that there is a voltage drop in the gate-voltages of M1 and M2. Similarly, it also avoids the zero voltage state of M3 and M4. Hence, it will start having current flowing in the branches and V_{ref} can be determined by adding the PTAT and CTAT voltages at the output branch. Once the circuit starts working in normal operating region, there is no use of the MOSFETs constituting the start-up circuit.

For the given BGR circuit, the process of generating the reference voltage actually requires the same voltage value at node-1 and at node-2. Here, if the voltage values at both the nodes are same, then the current flowing in both the branches of the current mirror will also be equal and hence, PTAT voltage, i.e., proportional to absolute temperature voltage and CTAT voltage, i.e., complement to absolute temperature

Table 1 Component sizes of proposed BGR circuit

Component	Value
M1, M2 (W/L)	20/5 ($\mu\text{m}/\mu\text{m}$)
M3, M4 (W/L)	20/5 ($\mu\text{m}/\mu\text{m}$)
M5 (W/L)	26/2.8 ($\mu\text{m}/\mu\text{m}$)
M6 (W/L), M7 (W/L)	18/5 ($\mu\text{m}/\mu\text{m}$)
R1	80 ($\text{k}\Omega$)
R2	230 ($\text{k}\Omega$)
R3	54 ($\text{k}\Omega$)
Q1, Q2, Q4	1 * (2 * 2) ($\mu\text{m}/\mu\text{m}$)
Q3	8 * (2 * 2) ($\mu\text{m}/\mu\text{m}$)

will be generated across R_2 and Q_4 , respectively. Hence, the final reference voltage, V_{ref} , which is the result of the sum of PTAT and CTAT voltage will be generated across the drain of M5. In this bandgap reference generator, circuit start-up circuit is connected to obtain the achieved results.

The expression for reference voltage can be represented using Eq. 1.

$$V_{\text{ref}} = (I_{\text{CTAT}} + I_{\text{PTAT}}) \cdot R_N \quad (1)$$

The temperature coefficient (TC) for the proposed bandgap reference circuit can be calculated using the below expression,

$$\text{TC} = (V_{\text{max}} - V_{\text{min}}) / V_{\text{ref}}(T_{\text{max}} - T_{\text{min}}) \quad (2)$$

where V_{max} represents maximum voltage, V_{min} minimum voltage, T_{max} maximum temperature, T_{min} is minimum temperature and V_{ref} is reference voltage produced. The size of the different transistors and the other parameters that are used in the proposed structure of bandgap reference circuit are shown in Table 1.

3.2 Simulation Results

The bandgap reference circuit presented in the paper is designed under 180 nm CMOS technology in OrCAD PSpice software with the inclusion of a start-up circuit. The results obtained after simulation are shown in Figs. 10 and 11. In Fig. 10, it is representing curve drawn between reference voltage of bandgap reference circuit and temperature. Similarly, Fig. 11 is representing the curve drawn between reference voltage (V_{ref}) and the supply voltage (V_{DD}). The proposed bandgap reference circuit produces a reference voltage of 685 mV with the temperature coefficient of 7.835 ppm/ $^{\circ}\text{C}$ over the temperature range of -40 to 150 $^{\circ}\text{C}$. The obtained results

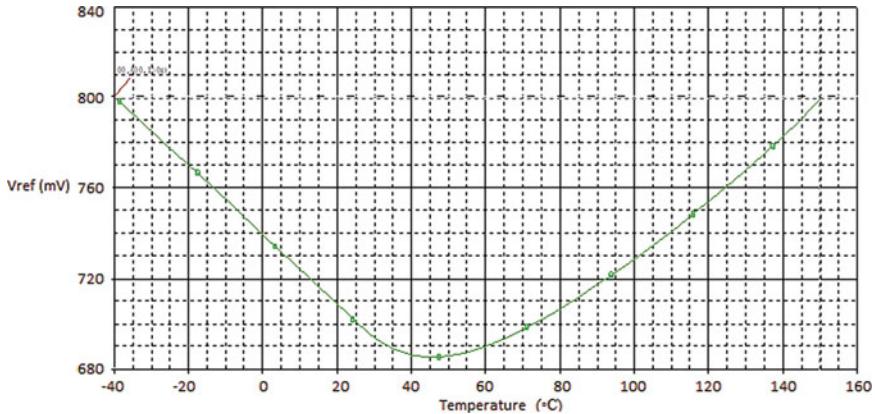


Fig. 10 Reference voltage versus temperature curve

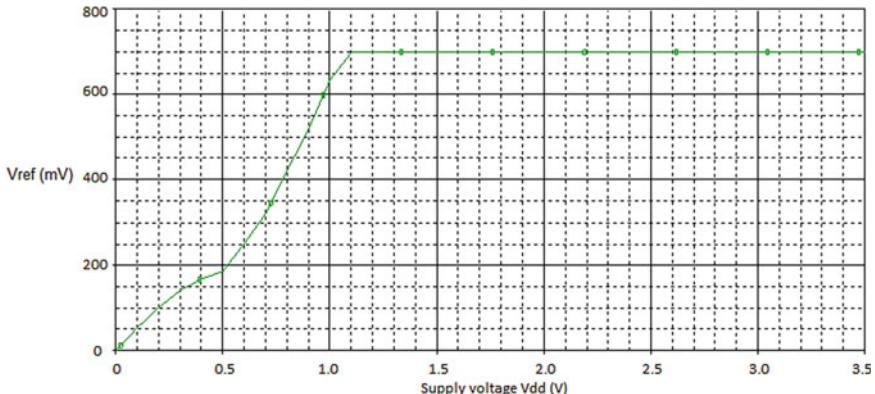


Fig. 11 Reference voltage versus supply voltage curve

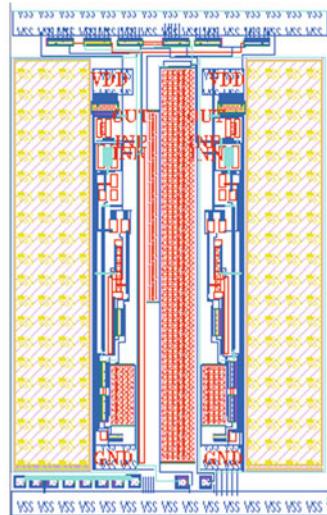
indicate the proposed bandgap reference circuit has good performance with low temperature coefficient and large temperature range.

In Fig. 12, layout of BGR is presented which consumes overall chip area of 0.0267 mm^2 .

4 Conclusion

BGR circuit along with different start-up circuits that can be used with them to eliminate zero current region were presented. The BGR circuit consists of BJTs, MOSFET, and resistors. The experimental results explain that proposed BGR generates reference voltage of 685 mV for the temperature range of -40 to 150 °C with the

Fig. 12 Layout of BGR design



temperature coefficient of 7.835 ppm/ $^{\circ}\text{C}$. Also, power dissipation of the proposed BGR circuit is very less which is of the order of 1.74 μW . Then multiple types of start-up circuits for bandgap reference circuits are discussed. After discussing a wide variety of start-up circuitries, it has been seen that different start-up circuits uses different techniques to generate the start-up voltage for the BGR core. As far as the design is concerned, one can design a simple start-up circuit consisting only a switch or it can be a complex one containing many transistors; it all depends on the requirement of the design. Also, along with the generation of start-up voltage, the start-up circuits perform other tasks as well, for instance, calibration of a temperature coefficient. Furthermore, these circuits aim to reduce the start-up time and the power consumption of the device and most of them tend to have no influence on the BGR core in the steady state. Finally, it can be concluded that for a self-biased bandgap reference circuits, these start-up circuits play a vital role as it ensures that the process of generating the reference voltage is carried out smoothly.

References

1. Hirose T, Ueno K, Kuroki N, Numa M (2010) A CMOS bandgap and sub-bandgap voltage reference circuits for nanowatt power LSIs. In: IEEE Asian solid-state circuits conference, Beijing, China, pp 1–4
2. Yu C, Siek L (2015) An area-efficient current-mode bandgap reference with intrinsic robust start-up behaviour. *IEEE Trans Circuits Syst—II: Express Briefs* 62(10):937–941
3. Osaki Y, Hirose T, Kuroki N, Numa M (2013) 1.2-V supply, 100-nW, 1.09-V bandgap and 0.7-V supply, 52.5-nW, 0.55-V subbandgap reference circuits for nanowatt CMOS LSIs. *IEEE J Solid-State Circuits* 48(6):1530–1538

4. Shrimali H, Liberali V (2014) The start-up circuit for a low voltage bandgap reference. In: 21st IEEE international conference on electronics, circuits and systems (ICECS), pp 92–95
5. Hongprasit S, Sa-Ngiamvibool W, Aurasopon A (2012) Design of bandgap core and startup circuits for all CMOS bandgap voltage reference. *Przeglad Elektrotechniczny (Electr Rev)* 88(4):277–280
6. Hassan AH, El Badry MA, Ismail Y, Mostafa H (2017) A low-power self-startup bandgap circuit for energy efficient applications. In: New generation of CAS (NGCAS), pp 29–32
7. Traversi G, De Canio F, Gaioni L, Manghisoni M, Ratti L, Re V (2015) Design of bandgap reference circuits in a 65 nm CMOS technology for HL-LHC applications. *J Instrum* 10(02):C02004
8. Cao TV, Wisland DT, Lande TS, Moradi F, Kim YH (2008) Novel start-up circuit with enhanced power-up characteristic for bandgap references. In: IEEE international SOC conference, pp 123–126
9. Boni A (2002) Op-amps and startup circuits for CMOS bandgap references with near 1-V supply. *IEEE J Solid-State Circuits* 37(10):1339–1343
10. Kamel AI, Saad A, Siong LS (2016) A high wide band PSRR and fast start-up current mode bandgap reference in 130 nm CMOS technology. In: IEEE international symposium on circuits and systems (ISCAS), pp 506–509
11. He J, Zhao L, Cheng Y (2014) A 0.885 ppm/°C digitally calibrated bandgap voltage reference with robust start-up circuit. In: 12th IEEE international conference on solid-state and integrated circuit technology (ICSICT), pp 1–3
12. Zhou Z, Wang H, Mei X, Chen C, Zhang B (2011) A 1.74 ppm/°C, high PSRR bandgap reference with fast start-up. In: IEEE international conference on microwaves, communications, antennas and electronic systems (COMCAS 2011), pp 1–5
13. Wu C, Goh WL, Yang Y, Chang A, Zhu X, Wang L (2016) A start-up free 200nW bandgap voltage reference. In: 14th IEEE international new circuits and systems conference (NEWCAS), pp 1–4
14. Zhou Z, Wang Y, Shi Y (2018) A bandgap reference using a novel soft self-start bias circuit. In: IEEE Asia Pacific conference on circuits and systems (APCCAS), pp 484–488

Recycle of Plastic Material for Fabrication of Paver Blocks



K. Ananthi, S. Priyadharsini, A. Ijas Ahamed, B. Viney Rangasamy, and R. Prabhu Surya

Abstract The aim of this paper is to develop and fabricate paver blocks from recycled plastic materials. This will reduce the disposal of plastic waste in landfills, which are primarily responsible for environmental contamination. The most often used recyclable plastic items include beverage packaging widely used for water, beer, soft beverages and juice, plastic bags and plastic containers used for food packaging. These recyclable plastic products are reinforced by bricks, and the availability of traditional construction materials in terms of quantity and quality is a hectic task for the builders. Demand for construction materials is growing tremendously day-to-day due to the ever-increasing demands of the housing and habitat sectors. Considering these two issues, this paper deals with the processing and examination of the features for using plastic wastes into paver blocks. Pavement blocks are the best materials for easy laying and finishing on the pathways and streets. The mechanical strength of paver blocks made up of plastic trash and the construction implications for paver blocks containing plastic pollution products are discussed here. The primary goal is the use of plastic materials with limited additions in the field of architecture. This is undoubtedly a cost-effective and environmentally sustainable approach.

Keywords Paver block · Recycled plastic material · Plastic-sand component

1 Introduction

Pavers are widely used throughout construction equipment for residential, manufacturing, and development properties. Some of the major applications are sidewalks, parking lots, porches, archways, green pathways, street designs, squares and residential streets. Pavers become available in a multitude of surface morphology,

K. Ananthi (✉) · A. Ijas Ahamed · B. Viney Rangasamy · R. Prabhu Surya
Sri Krishna College of Engineering and Technology, Coimbatore, India
e-mail: ananthik@skct.ac.in

S. Priyadharsini
Rajalakshmi Institute of Technology, Chennai, India

colours, textures, curves and other exclusive characteristics, rendering concrete slab pavers suitable for limitless design choices. Pavers being selected not just for the distinctive appearance or basic building type, but above all for certain outstanding stress concentration capabilities. Such capabilities allow the pavements suitable for building supplies on commercial sites as well as on low-speed paths. Pavers could be divided into two types, a concrete block paver and a clay brick furnace [1, 2]. Paver blocks find application in making road pavement in some areas with low traffic. Interlocking pavers ensure durability, flexibility, easy repair and maintenance, aesthetics and are very popular amongst public and industrial pavement, loading docks, and industrial storage areas.

The idea of utilizing waste material in the construction field has two-fold purposes, one, environmental concern is considered by the proper disposal of waste and next is the economic impact. Plastic use makes a significant contribution to the volumes of solid waste production, as plastic products are often used only once before they are disposed [3, 4]. The quantity of plastics in solid waste composition is more like plastics have become standard materials in our daily lives because of their properties such as toughness, versatility and light weight. The easiest way to treat plastic waste is to recycle and reuse it. For example, high-density polyethylene waste is used for the manufacture of bags and dustbins [5, 6]. The problem of disposal can be dealt with scientifically. This paper focuses primarily on the manufacture of paver blocks using plastics, with a view to the disposal of plastic waste. This results in reduced cracking or thermal melting, lower moisture rubbing and offers high longevity, almost no impact on production as well as provides eco-friendly design at low cost.

The usage of concrete for making paver blocks is discussed in Sect. 2, the idea of using plastic for making paver blocks is given in Sect. 3, Sect. 4 gives the details of materials used for making plastic paver blocks, Sect. 5 describes the working principle and process involved in proposed method, discussion about the proposed method is given in Sect. 6 and Sect. 7 depicts the conclusion of the work.

2 Concrete Paver Blocks

Concrete is composed of popular cement paste, volcanic aggregates (soil and small stones) and moisture. The moisture used to blend the concrete combines with the asphalt to produce a cured paste. The amount of water that is combined with cement varies from around 22 to 28% of the overall amount of adding water in concrete. Nutrient granules are usually separated into two divisions depending on their molecular weight. Particulates contained in this sponge are known as hard aggregates. Organic sand is often seen as a cement content in a concrete mixture where important weights are broken as a rough aggregate. Crushed stone pieces may also be used as a cement content when divided into particle size distributions that pass through a 4.7 mm sieve. The thoroughly characterized rugged rubble that can be used for concrete elements is 12.5 mm. Although, the chemical composition of the gross total does not really reach one third of the lightest net thickness of the void bricks. Fine

aggregate is a concrete product used for concrete elements. Cement is the costliest material per unit volume of aggregate. Thus, finer and rough granules are combined in such ratios that the concrete mixture is feasible and has a minimal concrete mixture for the desired consistency [7]. Paving blocks made of concrete are the best materials on the footpaths for quick laying, better look and finishing.

When producing concrete paver blocks, material loss is a major issue. In the concrete industry, efficient transport of raw materials is also a key problem. Waste management is important in order to achieve a clean working environment and also to increase production performance, which can be accomplished through simple adaptive cleaning mechanisms. The cement manufacturer requires about 1.5 billion tonnes of cement, 9 billion tonnes of aggregate and 1 billion tonnes of mixing water per year. This allows the concrete sector the biggest consumer of natural resources. Avoiding the waste of such environmental assets should be considered as important in a sustainable concrete production.

3 Paver Blocks with Plastic

In household waste, the composition of plastic would be high. The composition of waste is different in many countries because it is usually influenced by socio-economic characteristics, usage habits and waste management systems. In any way, the amount of plastic in the waste composition is normally high. One of the key components of plastic waste is polyethylene which is followed by polypropylene, polyethylene terephthalate and polystyrene. As recycling waste is strongly encouraged, the idea of recycling waste in the production of paver bricks has been suggested [8, 9]. The large amount of materials needed for building attracts areas for the reuse of waste materials in the construction industry recycling plastics for paver blocks has the benefits of long service life, reduced cost and attractive aesthetics. With paver blocks made from plastics, the environmental benefits are the safe disposal of bulk waste along with the reduction of environmental impacts generated by the burning of plastics. Plastics also help to save resources at home. It is estimated that the use of plastic foam insulation in homes and buildings per year will potentially save close to 60 million barrels of oil compared to other forms of insulation [10].

There are many sources of plastic waste, municipality is one of them. The word municipal solid waste describes those waste materials which are collected by the municipality itself or by the municipality. Thus, the concrete is avoided and the plastic waste is used to make the paver blocks accessible at a lower cost. This helps to reduce the cause of environmental contamination and helps us to recycle plastic waste in a useful form [11]. In [12], authors discussed the application of IoT for management of waste material and waste collection for smart cities.

4 Materials Used for Plastic Paver Blocks

4.1 Waste Plastics

By definition, plastics can be made in various shapes when they are heated in the closest setting. It is used in different ways, such as plastic bags, cups, food and beverage containers and even furniture. They become waste material after use. When these wastes are accumulated, they have harmful effects on human life as well as on plant life. It is also important to dispose of these waste materials properly. It is beneficial to use these wastes in their recycled form. Recycling is an effective way to treat waste in relation to plastics. If left unrecyclable, they would become a significant threat to the atmosphere as a pollutant. They are not easily decomposed and do not allow water to enter the soil, which decreases the amount of the groundwater, even though they are poisonous. It is suggested here that this plastic waste material be used as an effective composition for the manufacture of pavement blocks [13, 14].

Developing plastic waste paver blocks is an efficient way to make good use of plastic waste, which is a significant challenge to the sustainability of the ecological balance [15]. When plastic waste is recycled, it benefits

- To minimize the use of earth-based material as clay for the manufacture of brick resulting in the depletion of land, environmental degradation.
- Reduce the volume of plastic waste on land and water to prevent soil contamination and water pollution.
- Reducing the dumping area of plastic waste.
- Production of cost-effective materials.
- Preventing human health from infectious diseases.

4.2 River Sand

Sand is a natural granular product consisting of naturally occurring mineral particles and a finely divided content. The structure of sediment differs based on specific rock conditions and origins, but silica dioxide (SiO_2) in quartz type is the most important component of sediment in continental freshwater and non-tropical southern waters. The second most commonly used sand is sodium chloride, such as hematite, that has been produced primarily through living organisms, like reef and molluscs, throughout the last billion plus years. Sand is being used in the whole construction process.

5 Working Principle

Initially, plastic bags and polyethylene bags are collected and sorted out. This is a batch operation. Unused waste materials must be disposed of safely. Next, the

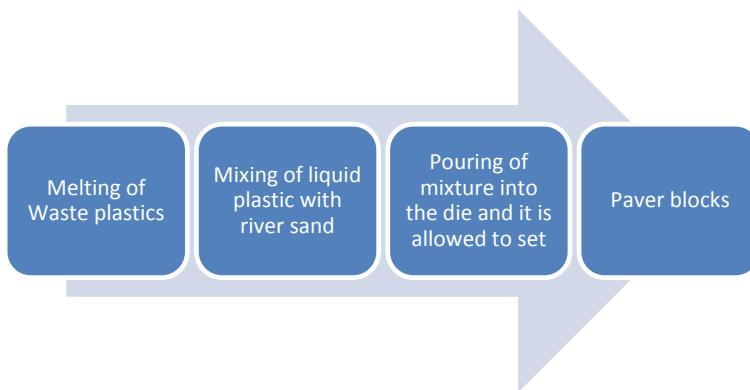


Fig. 1 Process diagram

collected waste bags are washed with water and dried to remove the water from it, after which the plastics are burned through the use of stones and firewood. The stones are arranged to carry the drum, and the firewood is placed in a gap between the stones and is ignited. The drum is put over the above setup and heated to remove the moisture in the drum. The plastic bags are then applied one by one to the drum, and the river sand is added to the plastic as it turns into hot fluid. Using a rod and a trough, the sand is carefully combined until it hardens. The mixture has a very short environment, but on the other hand, the mixing process does not take longer. This mixture is then poured into the brick mould and compacted using steel rods and finished using a trowel on the end. Before inserting the mixture into the container, the sides of the mould are greased with oil to extract the bricks quickly. All these processes are done step by step as shown in Fig. 1.

5.1 Melting

The method of recycling surplus or plastics waste recycling the content into usable items is called plastic recycling. Even as a large bulk of plastics are non-biodegradable, reuse is part of initiatives to reduce the amount of plastic throughout the landfill, particularly the nearly 8 million tonnes of plastic waste entering the planet's ocean each year. The amount of plastic that can be completely recycled instead of recycled or discarded, melting is the best method for recycling the waste plastics. The waste plastic is melted in the drum for 20 min. This converts the waste plastic into a liquid form. This liquid form of plastic has good flowability. The steps to be followed in the melting process are place the plastic in a heat-proof container, preheat the oven to around 149 °C, provide adequate ventilation, watch plastic carefully during melting in order to avoid burning. Melting is followed by the mixing process.

Table 1 Comprehensive strength test and water absorption test for mixtures of sand and plastic

Sand plastic ratio	Comprehensive strength (MPa)	Water absorption %
70:30	13.67	4.6
60:40	11.53	6.8
50:50	8.23	8.8

5.2 Mixing

The strength of the paver depends on the ratio of mixing melted plastic with sand. Here, it is experimented with different ratios of sand and plastic starting with a 50:50 equal mixture of sand and plastic. It is then changed to 60:40 and 70:30 to identify the best fit ratio in terms of strength of melted and mixed material. Continue mixing until all plastic waste is completely melted and there is a consistent black liquid. Also, at extremely high temperatures, plastic lumps can often remain.

The three specimens are taken as plastic and sand in the ratio of 50:50, 60:40, 70:30 and tested for comprehensive strength and water absorption capacity. Comprehensive strength test is conducted for ensuring maximum load the paver block can bear before fracturing. Water absorption tests on paver blocks are conducted to determine durability of the blocks. The results of comprehensive strength test and water absorption for the taken ratios of sand and plastic are shown in Table 1.

A combination of 70:30 fits well for floor tiles to be used in house keep. Usually, the tiles contain more sand than plastic, so that the plastic serves as a bonding agent to hold the sand together.

Stirring and heating shall continue until all lumps are removed and homogeneous paste is obtained. Homogeneous paste is essential to ensure the strength of the material. The time taken for obtaining this consistency is about 20 min. Enough care should be taken not to let the liquid get so hot. If it is overheated, it will not be suitable for making paver blocks. The sand is applied until the appropriate mixture is obtained and continues to be combined in such a way that the plastic, which acts as a binder, is very well mixed with sand and looks like grey liquid cement. The mixture has river sand 70% and plastic 30% which is already found suitable for paver block making.

5.3 Moulding

The shape of the bricks is moulded using a brick moulding box. It is making an enticing transition. Not only does brick moulding produce an exciting transition, but it also fills the void between the neighbouring materials. Usually, brick moulding has a block-like profile, with decorative routing running down the exposed face or edge; depending on the manufacturer, the forms and the exact dimensions vary. The most common material for brick moulding is wood. For outer trim, rot and pest-resistant

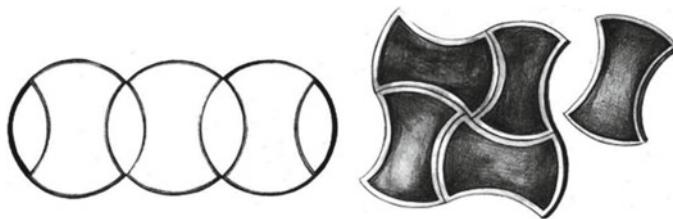


Fig. 2 Moulding samples

plants such as redwood are classic choices. Alternatively, pine brick moulding is offered by suppliers. Press the mixture into the mould and work it until there are no air gaps. A popular replacement for natural materials is plastic mouldings. Plastic mouldings, whilst comparatively costly, are resistant to rot and pest infestation. Using the spade with a metal shaft is used to remove the mixture easily and bring it into the mould. Using a mould mask, the mixture is put in a mould box compartment and rammed into a compact size. The moulding samples are shown in Fig. 2.

5.4 Setting of Blocks

Enable the hot mixture to be set in the mould for a few minutes, shaking the mould periodically to loosen the edges, a rocking motion works well. Keep trying to remove the mould. Remove the mould and leave until the mixture has cooled so that the slab does not crumble. It should harden in around 2 h. Bricks are ejected after proper ramming and left out for around 12 h in the sun. All these processes are done to allow the bricks to harden. Thus, the strong bricks made of recycled plastics are obtained which can be coloured further to use as paving blocks. This helps us to make an eco-friendly output and helps in the recycling of the waste plastic into a useful form.

6 Discussions

This method of paver block making helps in the manufacturing of low-cost paver blocks. Waste plastic is converted into a useful form and also it avoids the usage of cement. The blocks made by this have higher compressive strength and less water absorption. The output of the machine gives an eco-friendly product that does not harm the environment. As the plastic is added to the blocks, it improves the corrosion resistance. This also reduces the weight of the blocks than the normal paver blocks. These blocks can be applied to lay the platform floors and the roads. It can also be used to decorate the house parking floors by adding the colouring agent to the blocks.

It is also applied for the smooth finishing of the floor instead of the sand which is cost-effective. These materials can help to make the engineering components, which are fast and quick. When these blocks are broken or damaged, they can be collected from the users, and again they can be reheated to melt. They are again mixed with the available sand and using moulds bricks can be regenerated. This is the easy and the possible way of reuse of the plastic-sand components. The fabrication photos are shown in Fig. 3.



Fig. 3 Fabrication of process model

7 Conclusion

When sand is mixed with plastic waste, the substance is ideal for the manufacture of paver blocks. Sand and plastic waste paver blocks have been manufactured for hard standing areas. Plastic recycling is taking place at a significant rate in India. The process of turning plastic waste into a valuable resource for paver blocks is the best form of plastic recycling. By this process, the issue of the accumulation of plastic waste can be managed to a large extent. As a future work, sand can be replaced by other materials such as wood powder and casting powder. High quality burners and melting materials of high-density plastics may also be used.

References

1. Chkheiwer A (2017) Improvement of concrete paving blocks properties by mineral additions. *J Univ Babylon Pure Appl Sci* 25(1):157–164
2. Kanawade B, Nawale S (2018). Strength and durability of concrete paver block. NDT test for bridges
3. Saiprasad MK, Nagendra N (2019) Feasibility study on plastic-soil brick as a construction material. *Int J Eng Res Technol* 8(11):89–91
4. Mohan DMS, Vignesh J, Iyyappan P, Suresh C (2018) Utilization of plastic bags in pavement blocks. *Int J Pure Appl Math* 119:1407–1415
5. Ghuge J, Surale S, Patil BM, Bhutekar SB (2019) Utilization of waste plastic in manufacturing of paver blocks. *Int Res J Eng Technol (IRJET)* 06(04). e-ISSN: 2395-0056
6. Pandey AN, Yadav A, Chaudhary D, Nageshwar, Srivastava VK (2019) Waste plastic used in paving block: a research. *Int J Sci Res Rev* 07(03)
7. Caronge MA, Tjaronge MW, Hamada H, Irmawaty R (2017). Effect of water curing duration on strength behaviour of portland composite cement (PCC) mortar. IOP conference series: materials science and engineering. IOP Publishing Ltd. <https://doi.org/10.1088/1757-899X/271/1/012018>
8. Shubbar A, Sadique M, Kot P, Atherton W (2019) Future of clay-based construction materials—a review. *Constr Build Mater* 20–43. <https://doi.org/10.1016/J.CONBUILDMAT.2019.03.206>
9. Arjun Kumar S, Ganesh Babu S, Gowri kumar B, Afrid Sukkur S, Elango KS (2019) Utilisation of waste plastic in manufacturing of paver blocks. *Int J Innov Res Sci Technol* 5(11)
10. Ghuge J, Surale S, Patil BM, Bhutekar SB (2019) Utilization of waste plastic in manufacturing of paver blocks. *Int Res J Eng Technol (IRJET)* 6(4)
11. Jaivignesh SB, Sofi A (2017) Study on mechanical properties of concrete using plastic waste as an aggregate. IOP conference series: earth and environmental science. IOP Publishing Ltd. <https://doi.org/10.1088/1755-1315/80/1/012016>
12. Ali T, Irfan M, Alwadie AS, Glowacz A (2020) IoT-based smart waste bin monitoring and municipal solid waste management system for smart cities. *Arab J Sci Eng* 45:10185–10198
13. Mohan DMS, Vignesh J, Iyyappan P, Suresh C (2018) Utilization of plastic bags in pavement blocks. *Int J Pure Appl Math* 119(15):1407–1415
14. Yeole RC, Varma M (2017) Comparison of mix designs of paver blocks using waste rounded steel aggregates and rubber pad
15. Navia R, Chamy R (2017) The link between waste management, climate change and bioeconomy. *Waste Manag Res* 35(6):561–562. <https://doi.org/10.1177/0734242X17712622>

Handwritten Cursive English Character Recognition Using DAG-CNN



P. V. Bhagyasree, Ajay James, N. D. Bisna, and K. S. Vipin Kumar

Abstract Handwritten character recognition (HCR) is considered as an important part of optical character recognition (OCR). Handwritten character recognition can be used in bank cheques, prescriptions of medicine translations, etc. Handwritten characters, when compared to the printed characters, are very difficult to recognize due to the varying writing styles of individuals. The size, shape, and angles of every letter vary from person to person and even a single person's writing style will be different in two different documents. For character recognition approaches using geometrical features, structural features and statistical features were used traditionally. Nowadays, handwritten character recognition widely uses deep learning approaches. Deep learning techniques can learn the visual features automatically. This helps to avoid limiting the features like in traditional methods. Convolutional neural networks are a deep learning approach and it always gives better accuracy than the conventional techniques. In this paper, a novel deep learning technique, namely directed acyclic graph–convolutional neural network (DAG-CNN), is used for handwritten character recognition.

Keywords Directed acyclic graph (DAG) · Deep learning · Convolutional neural network (CNN) · Rectified linear unit (ReLU) · Handwritten character recognition (HCR) · Optical character recognition (OCR) · Pattern recognition

1 Introduction

In the era of digital revolution where everything is being converted to digital form, all type of documents in which both handwritten and printed forms are getting digitized. This digitization of documents has various applications and it enables to store in cloud, pen drive, and other storage media which helps to carry and preserve several

P. V. Bhagyasree
Christ College of Engineering, Irinjalakuda, India

A. James (✉) · N. D. Bisna · K. S. Vipin Kumar
Government Engineering College, Thrissur, India

documents in digital form rather than in the traditional paper-based documentation system. The old records exist in their original handwritten form that is required to be converted to electronic format for preserving, editing, sharing, and to prevent tampering of the original copy. Optical character recognition has eased this task to a great extent. We can define optical character recognition as a process of recognizing characters from the text image, in order to encode the text in a particular format, which is more convenient to edit. These texts can be either handwritten or printed. In printed format throughout the document, the characters will be uniformly arranged. Irrespective of the writers, same characters will have same appearance. But for handwritten document, according to the writer, the size and the shape of the character vary, which will make it far difficult to process. Existing OCR methods can be classified as conventional methods and deep learning methods. The former is based on handcrafted features, and the latter is based on neural network models. Because of low quality of features, traditional handcrafted feature cannot provide much recognition rate or accuracy. The modern methods of recognition of characters use deep learning techniques and in which convolutional neural networks (CNNs) are the most widely used techniques. CNN is capable of solving the problem of lower quality features. The deep learning techniques differ from conventional methods mainly in the way, how the features are taken. Conventional methods take the hand-crafted features, which mainly depends on knowledge of the programmer, while deep learning techniques learn the features in a self-learning way from data. Even when there are enormous features, deep learning can automatically learn feature representation from data. Deep learning techniques dominate in the field of OCR due to its high-performance and simpler recognition procedure.

In this paper, directed acyclic graph convolutional neural network (DAG-CNN), a cursive English handwritten character recognition using the deep neural network called is explained. The rest of the paper is organized as follows: Sect. 2 gives some related works done in this area. Section 3 explains the proposed method, Sect. 4 discusses the result and then Sect. 5 lists the conclusions.

2 Related Works

Aleskerova and Zhuravlev [1] proposed a technique in which an approach to solve the difficulty to do separate Chinese handwritten characters recognition is done. A hierarchy of multiple neural networks is proposed here with a first-level network selects a second-level network which performs the final recognition. Further, as consistently level organization is prepared for a subset of classes, it tends to be more precise and less complex than the entire order organization. In the methodology, a few classes of Chinese characters are consolidated with comparable outside highlights in a single gathering (bunch) and work with these groups further. Exploratory outcomes on 3755-class and CASIA 200-class written by hand datasets are introduced, which contrast the proposed progressive methodology and the traditional single-network approach.

Another work proposed by Parthiban et al. [2], used recurrent neural network (RNN) for recognition of English handwritten characters. Recurrent neural network was utilized to discover the arrangement of characters in this work. The dataset was made by gathering handwritten information from various people, and the data are split randomly as training and testing dataset. By using RNN algorithm, an accuracy of 90% was obtained.

Offline handwritten Kannada character recognition using convolutional neural networks, proposed by Ramesh et al. [3], arranges and distinguishes the Kannada transcribed characters utilizing the deep learning technique. This technique gives a simple route to the client since there is no pre-processing of information. Those works are taken care of by the neural network, which is the mind of the profound learning model. This lessens the weight on the client making the work seriously encouraging. With a case organization, the model prepared with a decent measure of information can perceive the Kannada handwritten characters. The future executions can be to perceive the words and later perceiving sentences. The acquired exactness compared 93.2 and 78.73% for the two unique sorts of datasets utilized in the work.

K-nearest neighbor (KNN) for classification and hidden Markov model (HMM) for recognition was used by Leena Hepzi et al. [4]. The median filtering method is used for pre-processing. Several assumptions must be made and are considered as one of the disadvantages of using HMM. Here, it is assumed that for every observation, it only depends on the current state and it does not have any dependency with the previous state. It made handling the correlation between the continuous observations becomes difficult. Finally, classification is done using KNN algorithm. An accuracy of 90% is obtained by this method.

A complete OCR system that recognized characters from an input document rather than recognizing just characters was proposed by Shanjana and James [5]. The document was binarized using Otsu's technique. Horizontal projection profile was identified for the entire image, and it was used to segment the lines. The average height was used to separate lines that were too close. The word segmentation was completed on the assumption that distance between words is always higher than those between the characters. Vertical projection profile was used in word segmentation and also used in character segmentation with the addition of connected component analysis. SVM was the classifier used along with radial basis function (RBF) kernel and PCA for reduction of dimensionality. The overall accuracy of 89.7% was reported for this method.

In Ashiquzzaman et al. [6], a technique to improve the classification accuracy of handwritten Bangla compound characters is suggested. This paper suggests a method using a greed layer-wise training approach by using deep convolutional neural network (DCNN). Vanishing gradient and over-fitting are the two main problems identified here. These are tackled by using exponential linear unit (ELU) and dropout, respectively. This method achieved an accuracy of 93.68%.

ResNet-52 [7], GoogleNet [8], VGG-16 [9], and AlexNet [10] are some famous CNN models that have an aggregate of 150, 78, 57, and 27 hyper-boundaries, individually. A terrible decision for hyper-boundaries can cause a huge calculation cost which lead to poor CNN execution. The analyst's skill plays a significant job in

settling on the design of hyper-boundaries. This makes a few inquiries regarding CNN's plan for handwriting recognition. What impact do diverse hyper-boundaries have on CNN execution? How is CNN better in removing particular highlights from handwritten characters? What is the part of plan boundaries in improving CNN execution? To manage future exploration in the handwriting recognition field, it is imperative to address these inquiries. The convolutional neural network gets an insurgency the handwritten acknowledgment field and conveyed the cutting-edge execution in this space [11–15].

3 Proposed Method

3.1 System Architecture

The modules in the proposed system are the following. First module is the data acquisition, then augmentation of dataset, followed by creation of training and testing data, pre-processing of dataset, model creation and training, and last phase, classifying and predicting output. The overall system architecture is shown in Fig. 1.

3.2 DAG Network

Directed acyclic graph (DAG) is a directed graph without cycles. A DAG-CNN is a convolutional neural network in multi-scale which can be sub-divided into several paths. A basic convolutional neural network is linear in structure, and the layers are connected in such a way that, the output of each convolutional layer is connected to the succeeding layer. For the last convolutional layer, the output will be directly fed to the output layer for classification. Because of this linear structure, the accuracy will get reduced as the network goes deeper. This accuracy decrease happens due to the vanishing gradient problem. Deep neural networks address this issue using skip connections. Figure 2 shows the basic CNN architecture, and Fig. 3 shows the DAG-CNN architecture.

Instead of following a linear structure, DAG is connected as a directed acyclic graph. Due to this architecture, it has a unique characteristic. It can have multiple inputs and outputs. Having multiple connections, this structure allows every layer to have a direct connection to the final classifier layer directly using skip connections. This in turn, facilitates all the low, mid, and high level features to contribute to enhance the network performance. At the same time, it leads to a huge number of features, which may lead to a problem called over-fitting. The problem of over-fitting can be avoided by selecting only the optimal layers which enhances performance. A greedy forward selection strategy devised to select those optimal layers. Here, at each convolutional layer, features are extracted and are fed to an addition block.

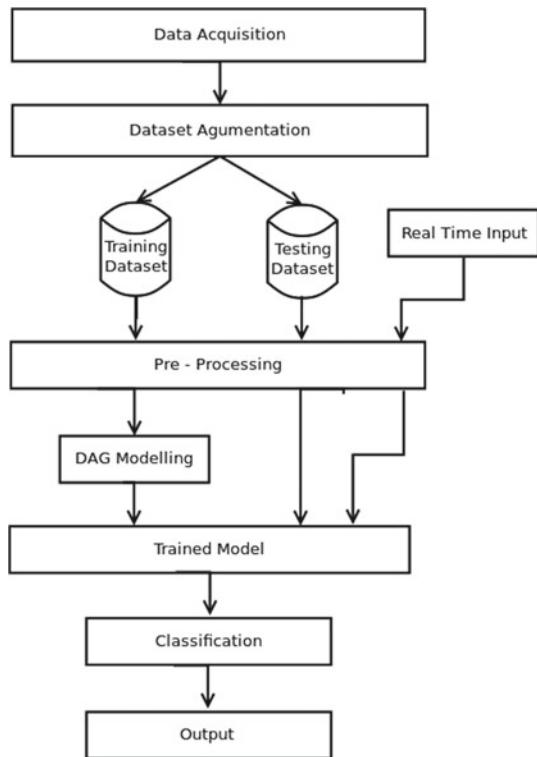
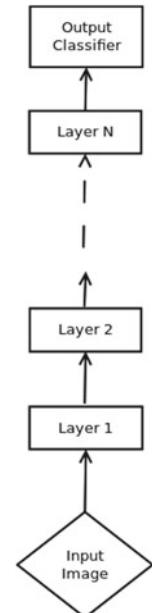
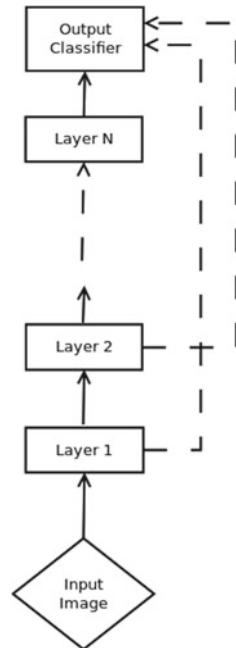
Fig. 1 System architecture**Fig. 2** Basic CNN

Fig. 3 DAG-CNN



In this addition block, an element wise addition of features across the layers will be done. Following this, normal average pooling operation is done. From there for classification and accuracy prediction, the output will be given to softmax module.

3.3 *System Modules*

The proposed system is divided into data acquisition, augmentation of dataset, training and testing datasets creation, dataset pre-processing, model training, classification, and testing modules.

Data Acquisition: Neural network techniques need huge amount of data to train the network properly. Many standard datasets are available for cursive English words and sentences, but no character dataset is available. In this project, character dataset is used as focus is on feature extraction. The data are obtained from various students and other people. Each cursive letter was collected in a A4 sheet which was divided into many grids. One sheet contains 112 ($14 * 8$) grids. Hence, from a sheet 112 samples of a letter can be obtained. So the total number of samples collected was 29,120 ($1120 * 26$).

Dataset Augmentation: Second step is performing dataset augmentation. Affine transformation is a method to increase the existing dataset. Affine transformation

is a linear method, which preserves every points, straight lines, and planes. After an affine transformation, sets of parallel lines remain parallel. Scaling, translation, sheering, and rotation are the four major affine transformation. While collecting the dataset, it was found that some of the samples were too different from the original characters. So the character list is filtered to remove all such mistakes. However, the characters were preserved that were legible to at least some extend so that the system will be efficient to deal with even such characters. After filtering and dataset expansion, a total of 52,000 images were obtained. This is then used to train the network.

Training and Testing Dataset Creation: Next step is to divide the dataset as training and testing data. 80% of data is used to train the network, and remaining 20% was used for testing. Here, 800 images of each character were used for training and tested with 200 images.

Dataset Pre-processing: Dataset pre-processing is done to improve the quality of the image. Pre-processing removes the unwanted elements from the input image and converts it into a required format. The collected and cropped images will be of different size. So all the individual images need to be converted into a standard size. Therefore, the image needs to be cropped, selecting only the character field, and scaled to the required size.

Model Training: Model training is the most important step. First, the structures of CNN need to be defined. The type of layers, count of layers, neurons count per layer, kernel size, stride in every operation, and the classifier type to be used need to be determined. It is not possible to know which configuration of the network will give the best accuracy. So the training dataset needs to be tested against different settings of the network. The detailed architecture of a basic DAG-CNN is as shown in Fig. 4.

Classification: The second last layer of the CNN is a softmax layer, which has softmax function for squashing all the inputs into a value bet and 0 and 1. The sum of all values in a softmax layer will be between 0 and 1. The given input image is classified using the next layer. This classification is based on the values generated by the previous softmax layer. This layer classifies the input image into one of the 26 classes.

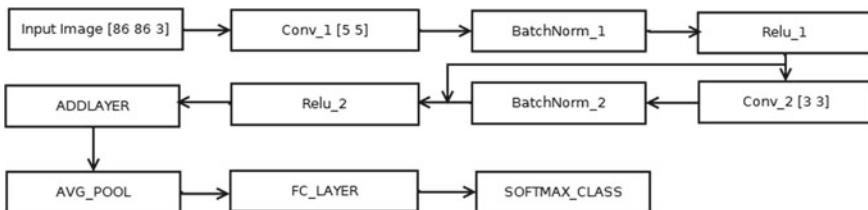


Fig. 4 Detailed DAG architecture

Testing: The entire dataset is first divided into a testing dataset and a training dataset. The split is done so that 80% of the dataset is selected as the training set and the remaining is selected as the testing set. The split is done on a random basis. This training set is tested on the CNN after training. The class labels of testing images are hidden at first. These images are then passed through the CNN which predicts the class of the input image. Thus, accuracy of the CNN can be found by dividing the count of correct prediction by the total count of test images. Additionally, new raw inputs were taken from different individuals and tested against the network. For this, the input images were first binarized and then scaled to 86×86 sizes. This image is then fed into the network and then the class label of the input image is predicted. After the label of the input character image is predicted, it is now required to get the editable format of the input image. A CSV file is made where the label of each character acts as the index and unicode as its value. By referencing this CSV file, The OCR system can now generate the unicode from the predicted label. The final stage of any OCR is the input image in editable format; the system opens a.txt file in notepad with the recognized character.

4 Results and Discussion

4.1 Dataset

In this work, the dataset used is handwritten cursive English alphabets. The dataset contains both capital and small letters. The data are collected from various students and other people. Since the data are obtained from different individuals, it was easy to get various types of handwriting. This work mainly focused on feature extraction and hence characters were collected. After the collection of raw data, it is scanned and the softcopy is made. Scanning is performed at 150 dpi. After that, the each data sheet is cropped to obtain single character images. The cropping is done using MATLAB code. This set of character image is mentioned as raw data in the later sections. Since neural network needs lots of data to train the model, the raw data are spanned by augmentation. Here, augmentation is done using Xn-View tool. This augmented data are mentioned as augmented data in the later sections of this chapter. The size of each character image in these datasets is 86×86 . However, for training and testing of model size of the images is reduced to 28×28 for maintaining the originality of the network. The characters of cursive small are labeled “a” to “z”, and characters of cursive capital is labeled as “A” to “Z”. There is a total of 52,000 images in this 26 classes of each image.

4.2 CNN Versus DAG-CNN

Convolutional neural network defined in this project is a simple neural network with three convolutional layers, three batch normalization, three rectified linear units and one poling, fully-connected, and softmax layer. In this architecture, every convolutional layer's output is fed into the next convolutional layer and so on. Only the final layer is connected directly to the softmax module for classification. DAG-CNN, directed acyclic graph is also a neural network which incorporates more features than basic CNN can include. DAG achieves this property by making its architecture unique. DAG is treated as a multi-scaled CNN. In DAG also there are three convolutional layers, three batch normalization; three rectified linear units and one poling, fully-connected, and softmax layer. But the difference from basic architecture is that, here every layer of convolution can be directly given to softmax module for classification. This is achieved by using a skip connection. Neural network can have many layers, and if all the layers are connected to the classifier using skip connection, there will be hundreds or thousands of features. As the number of features increases beyond a required amount, there will be problem of over-fitting. So to avoid such problems, only those layers which contribute efficient features will be connected to the classifier directly. The best layers are found out by trial and error method. In this work, two datasets are tested against basic CNN, DAG_1, and DAG_2 and the accuracy of each dataset with various epochs and learning rate using different networks is noted as tables.

From Table 1, it is observed that accuracy for epoch 10 is more for learning rate 0.01. From the table, it is clear that the accuracy slightly increases from learning rate 0.004 till 0.01, afterward the accuracy slightly decreases for all the epochs. The comparison is made by taking three different epochs 5, 10, and 15. First, the accuracy increases with increase in learning rate and then after 0.01, it starts decreasing with increase in the learning rate.

From Table 2, it is observed that accuracy for epoch 10 is more for learning rate 0.01. From the table, it is clear that the accuracy slightly increases from learning rate 0.004 till 0.01, afterward the accuracy slightly decreases for all the epochs. The comparison is made by taking three different epochs 5, 10, and 15. First, the accuracy increases with increase in learning rate and then after 0.01, it starts decreasing with increase in learning rate.

From Table 3, it is observed that accuracy for epoch 10 is more for learning rate 0.01. From the table, it is clear that the accuracy slightly increases from learning rate 0.004 till 0.01, afterward the accuracy slightly decreases for all the epochs. The comparison is made by taking three different epochs 5, 10, and 15. First, the accuracy increases with increase in learning rate and then after 0.01, it starts decreasing with increase in learning rate.

From Table 4, it is observed that accuracy for epoch 10 is more for learning rate 0.01. From the table, it is clear that the accuracy slightly increases from learning rate 0.004 till 0.01, afterward the accuracy slightly decreases for all the epochs. The comparison is made by taking three different epochs 5, 10, and 15. First, the accuracy

Table 1 Comparison of accuracy for different learning rate and epoch for cursive small letters using DAG_1

Epoch	Basic learning rate	Accuracy (%)	Training time
5	0.004	93.98	1 min 38 s
	0.006	94.69	1 min 36 s
	0.008	95.07	1 min 33 s
	0.01	95.33	1 min 45 s
	0.02	94.42	1 min 41 s
	0.03	94.38	1 min 38 s
10	0.004	94.23	2 min 45 s
	0.006	94.56	2 min 38 s
	0.008	95.26	2 min 31 s
	0.01	96.17	2 min 52 s
	0.02	95.17	2 min 31 s
	0.03	95.09	2 min 16 s
15	0.004	93.66	2 min 58 s
	0.006	94.26	3 min 56 s
	0.008	95.26	3 min 42 s
	0.01	95.69	4 min 03 s
	0.02	95.33	3 min 59 s
	0.03	95.16	3 min 48 s

Table 2 Comparison of accuracy for different learning rate and epoch for cursive capital letters using DAG_1

Epoch	Basic learning rate	Accuracy (%)	Training time
5	0.004	97.99	1 min 48 s
	0.006	98.02	1 min 38 s
	0.008	98.21	1 min 22 s
	0.01	98.38	1 min 59 s
	0.02	98.12	1 min 45 s
	0.03	98.04	1 min 22 s
10	0.004	98.42	3 min 04 s
	0.006	98.56	3 min 00 s
	0.008	98.54	2 min 57 s
	0.01	99.08	3 min 14 s
	0.02	98.76	3 min 02 s
	0.03	98.44	2 min 47 s
15	0.004	98.36	3 min 54 s
	0.006	98.38	3 min 45 s
	0.008	98.49	3 min 42 s
	0.01	98.52	4 min 12 s
	0.02	98.23	3 min 52 s
	0.03	98.06	3 min 48 s

Table 3 Comparison of accuracy for different learning rate and epoch for cursive small letters using DAG_2

Epoch	Basic learning rate	Accuracy (%)	Training time
5	0.004	93.56	1 min 48 s
	0.006	94.11	1 min 36 s
	0.008	94.82	1 min 22 s
	0.01	95.27	1 min 50 s
	0.02	94.98	1 min 47 s
	0.03	94.44	1 min 38 s
10	0.004	95.29	2 min 54 s
	0.006	96.22	2 min 35 s
	0.008	97.01	2 min 32 s
	0.01	97.31	3 min 15 s
	0.02	96.90	3 min 12 s
	0.03	95.75	3 min 02 s
15	0.004	95.42	4 min 02 s
	0.006	95.56	3 min 48 s
	0.008	95.62	3 min 41 s
	0.01	96.18	3 min 34 s
	0.02	96.08	3 min 22 s
	0.03	95.21	3 min 02 s

Table 4 Comparison of accuracy for different learning rate and epoch for cursive capital letters using DAG_2

Epoch	Basic learning rate	Accuracy (%)	Training time
5	0.004	97.99	1 min 56 s
	0.006	98.02	1 min 32 s
	0.008	98.26	1 min 30 s
	0.01	98.54	1 min 42 s
	0.02	98.23	1 min 40 s
	0.03	98.02	1 min 12 s
10	0.004	98.56	3 min 45 s
	0.006	98.48	3 min 06 s
	0.008	99.06	2 min 56 s
	0.01	99.56	3 min 15 s
	0.02	99.02	3 min 02 s
	0.03	98.58	2 min 45 s
15	0.004	98.24	3 min 29 s
	0.006	98.79	3 min 22 s
	0.008	98.74	3 min 11 s
	0.01	99.37	3 min 55 s
	0.02	98.88	3 min 35 s
	0.03	98.72	3 min 12 s

increases with increase in learning rate and then after 0.01, it starts decreasing with increase in learning rate.

To get the best network configuration, it was necessary to try and find the optimal settings. Different parameter values and different techniques for pooling as well as slight modification of the network were carried out. Max pooling and average pooling were 2 of the popular sampling methods. Both sampling methods were tried on the network. The results are shown in Table 5, while there was increase in accuracy for average pooling over max pooling, there is also a slight increase in training time as well. However, due to the fact that training time does depend on the system which is used, accuracy is the value that really matter and average pooling has the highest accuracy among them.

So the results in Table 6 show that DAG network provides better accuracy than the basic CNN model. Table 7 shows the comparison of various techniques. The observation table shows that deep learning techniques achieve better accuracy than the handcrafted feature extraction methods.

Table 5 Comparison of pooling strategies

Pooling method	Testing accuracy	Training time
Average pooling	99.58	2 min 22 s
Max pooling	99.01	2 min 17 s

Table 6 Comparison of CNN and DAG-CNN for cursive small and capital letters

Dataset	CNN (%)	DAG 1 (%)	DAG 2 (%)
Cursive small letters	90.50	96.17	97.31
Cursive capital letters	97.00	99.08	99.56

Table 7 Observation table

Technique used	Method	Accuracy (%)
DCNN	Deep learning	93
KNN classification	Handcrafted	90
Pixel intensity analysis	Handcrafted	91
HMM	Handcrafted	88.3
Convex Hull algorithm	Handcrafted	85
Basic CNN (small)	Deep learning	90.5
Basic CNN (capital)	Deep learning	97
DAG 1 (small)	Deep learning	96.17
DAG 1 (capital)	Deep learning	99.08
DAG 2 (small)	Deep learning	97.31
DAG 2 (capital)	Deep learning	99.56

5 Conclusion

Handwritten written character recognition assumes a significant part in optical character recognition. It extraordinarily adds to computerization of numerous things like clinical solutions, assessment forms, and so on. Prior high quality component techniques were utilized for character recognition, which was not productive and requires a lot of exertion and time. Extraordinary compared to other option for customary handmade highlights for character recognition was to utilize deep learning strategies. Every one of the features utilized here is machine-created features and delivers a lot of productivity and precision to the outcome. The different deep learning methods utilized are all convolutional neural network (CNN). The CNN delivers a superior precision; however, it is a straight model where just the final layer may be straightforwardly chosen for order. So, in order to utilize the low, center, and undeniable level features for grouping, we apply multi-scaled CNN known as directed acyclic graph convolutional neural network (DAG-CNN).

References

1. Aleskerova N, Zhuravlev A (2020) Handwritten Chinese characters recognition using two-stage hierarchical convolutional neural network. In: 2020 17th international conference on frontiers in handwriting recognition (ICFHR)
2. Parthiban R, Ezhilarasi R, Saravanan D (2020) Optical character recognition for English handwritten text using recurrent neural network. In: 2020 international conference on system, computation, automation and networking (ICSCAN). IEEE
3. Ramesh G, Sharma GN, Manoj Balaji J, Champa HN (2019) Offline Kannada handwritten character recognition using convolutional neural networks. In: 2019 5th IEEE international WIE conference on electrical and computer engineering (WIECON-ECE)
4. Leena Hepzi J, Muthumani I, Selvabharathi S (2017) English cursive handwritten character recognition. *Adv Nat Appl Sci*
5. Shanjana C, James A (2015) Offline recognition of Malayalam handwritten text. In: 8th international conference interdisciplinarity in engineering. INTER-ENG2014. Procedia technology, vol 19, pp 772–779
6. Ashiquzzaman A, Kawsar Tushary A, Dutta S, Mohsin F (2017) An efficient method for improving classification accuracy of handwritten Bangla compound characters using DCNN with dropout and ELU. In: 3rd international conference on research in computational intelligence and communication networks (ICRCISN2017)
7. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016
8. Christian S, Wei L, Yangqing J, Pierre S, Scott R, Dragomir A, Andrew R (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 7–12 June 2015
9. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: Proceedings of the international conference on learning representations, San Diego, CA, USA, 7–9 May 2015
10. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105

11. Tabik S, Alvear-Sandoval RF, Ruiz MM, Sancho-Go'mez JL, Figueiras-Vidal AR, Herrera F (2020) MNIST-NET10: a heterogeneous deep networks fusion based on the degree of certainty to reach 0.1% error rate. Ensembles over and proposal. *Inf Fusion* 62:73–80
12. Lang G, Qingguo L, Mingjie C, Tian Y, Qimei X (2017) Incremental approaches to knowledge reduction based on characteristic matrices. *Int J Mach Learn Cybern* 8:203–222
13. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39:2481–2495
14. He S, Zeng W, Xie K, Yang H, Lai M, Su X (2017) PPNC: privacy preserving scheme for random linear network coding in smart grid. *KSII Trans Internet Inf Syst* 11:1510–1532
15. Sueiras J, Ruiz V, Sanchez A, Velez JF (2018) Offline continuous handwriting recognition using sequence to sequence neural networks. *Neuro Comput* 289:119–128

Automation of Security Scanning Tools Using K8's Environment for Commercial Telecom Product



Pooja and K. S. Shushrutha

Abstract Automation of security scanning tools at container level in Kubernetes Environment has become an essential task in commercial telecom field to mitigate manual interruption and reduce time. In this paper, the focus is on automation of a robustness testing tool at container level on Linux server. This robustness scanning tool finds the defects in the microservices of service enablement platform, which are tested protocols in system under test, and it provides preventive testing, i.e., identifies the issues in advance of the threat, it needs special set of test cases for each tested application, and the robustness of the software is being decided based on the result. Automating the security scanning procedure using robot framework reduces manual interruption, manual hours and reduces the need for monitoring. Special keywords are created in RIDE tool, and proper time stamp has been provided for checking the robustness of the microservices. A case study to demonstrate manual and automated robustness scanning microservice of operation, administration and maintenance (OAM) of LTE architecture is presented. This paper manifests the results of both manual and automation testing. Constructive comparison has been made with these results to prove the efficiency of automated security scanning

Keywords Robot framework · RIDE · System under test (SUT) · Robustness testing · Kubernetes (K8's) · Service enablement platform (SEP) · Virtual machine (VM)

Pooja (✉) · K. S. Shushrutha
RV College of Engineering, Bengaluru, India
e-mail: pooja.lcs19@rvce.edu.in

K. S. Shushrutha
e-mail: shushruthaks@rvce.edu.in

1 Introduction

Security testing is a sort of software testing that identifies vulnerabilities, dangers and hazards in software applications and guards against hostile intruder assaults. The basic purpose of automating security testing is to determine and assess possible vulnerabilities in a system without manual interruption so that attacks can be faced, and the system does not cease working or be exploited [1]. Automation of robustness scanning aids in the detection of any potential security vulnerabilities and anomalies in the system in less time, as well as assisting developers in the resolution of issues through script.

Robustness is the degree to which any system functions accurately in the presence of unexpected inputs or traumatic environmental conditions [2]. This robustness testing tool is a proprietary tool of which the name is not being disclosed in this paper. The need of guaranteeing safe access to services and activities grows [3]. There are different security scanning tools like DoS, NMAP for port scanning and Slowloris which are used in telecom field. There is a pressing need to incorporate security testing into standard testing procedures [4]. At the container level, development on security scanning technologies is as scarce as hen's teeth. There are various types of individual test suites defined for each protocol; automating the complete processing of the tool on K8's environment using RIDE would be a one-time activity to conclude on software robustness, as shown in this paper, and the script written can be triggered any number of times. The test plans are developed inside test suites depending on the kind of tested protocol at the Kubernetes container level, and the results are assessed based on the test plan execution results.

Robot framework is an open-source automation framework hosted on GitHub with pre-defined keywords, and required keywords based on the testcase requirements can also be developed [5]. A few standard libraries are included with robot framework, and they provide automation and testing capabilities to robot framework. K8's is a platform for automating the deployment and control of containerized implementations [6]; automation of security scanning tools at the container level is underutilized. This testing framework is widely used for robustness testing and even for collecting logs after testing because of its reusable script, low cost and high maintenance. Specific keywords are created under the testcases, appropriate timeouts are provided, and documentation is produced in this tool [7].

In this paper, robustness testing of the microservices that are deployed in SEP on multinode K8's cluster developed manually has been automated using RIDE. The automation of this process at container level does not require manual interruption and saves time. Sect. 1 of the paper introduces the robustness testing tool, the robot framework, and the relevance of automation at the container level. In Sect. 2 brief review on the security scanning, robot framework and robustness scanning tool, the execution of manual framework and automated framework with server's connectivity from SUT to VM is presented. The complete manual execution is automated, from the SEP deployment, which includes microservices, to the robustness testing process. In Sect. 3, a case study demonstrating the robustness scanning of a microservice is

detailed. The results of manual and automation execution are compared in Sect. 4 of the paper. Conclusion and future scope are presented in brief in Sect. 5.

2 Implementation of Security Scanning of Softwares/Microservices in Sut

2.1 Brief Introduction to Robustness Testing Tool

Robustness testing tool tests all interfaces of the systems that are used to provide unpredicted inputs in application protocols and networking [8]. In robustness testing technique, large amount of spontaneously generated and on purpose contorted protocol messages are sent to the microservices, as a pretested protocol. Robustness tester is a tested protocols dependent tool, and this dependency requires a special set of testcases for every tested microservice. A marked mini-simulation protocol is developed to generate testcases to achieve above requirements in the robustness testing process. Testcases are generated robotically by grouping anomalies and valid elements. Generated testcases will be documented inside robustness tester test suites. The test plans are created inside test suites are based on the type of tested protocol at Kubernetes container level, these test plans are executed, and results are analyzed.

Automation of robustness scanning has prerequisites as the setup/SUT should be ready with SEP deployment and PODs/microservices to be taken for security testing need to be tested. The steps undertaken are described as below:

1. Tool cluster is the SUT under which microservices are deployed, these are the tested protocols, and all three nodes of tools cluster are connected internally using calico network.
2. Virtual machine on which robustness scanning tool is installed contains test suites which comes with the license agreement, test plans are created inside these test suites, and all required configurations are entered inside these test plans.
3. New keywords written based on the protocol, pre-existing libraries in RIDE tool and inbuilt keywords are used for writing automation script. In RIDE, a preparation file which includes login credentials of servers must be set before executing the script.

2.2 Flow of Execution

SUT is the targeted server, tested protocols refer to the microservices which are part of SEP, calico network is used for internal nodes connection, and Multus SRIOV deployed on top of Kubernetes is used for external connectivity, to have secured connection between SUT (multinode cluster) and VM. The IP and VLANs assigned

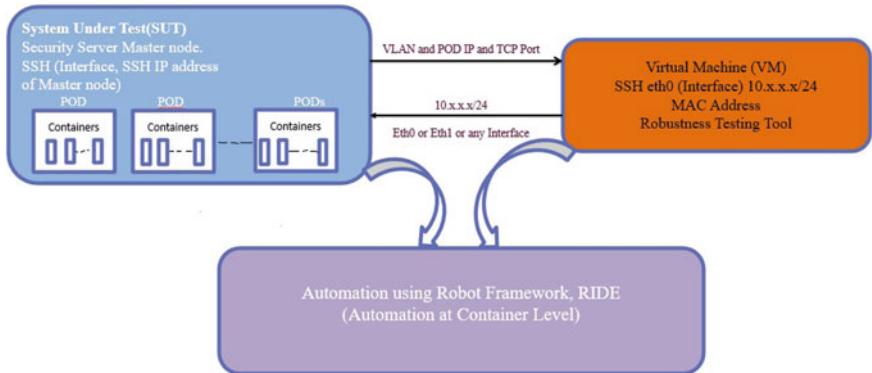


Fig. 1 Flow of execution and connection establishment

to the server interfaces which are up on servers are used to have connectivity between SUT and VM, the ping is checked between both servers, and successful ping indicates that both setups are connected as shown in Fig. 1. The process of robustness scanning on tool is executed as below:

1. The test plans are created manually on robustness scanning tool on required suite.
2. This tool has a feature to check interoperability, and this check is between the tested protocols on SUT and robustness scanning tool.
3. If interoperability check is successful, then test cases execution is done, and results are analyzed.

2.3 Manual Execution

As discussed in the sections above, robustness testing of the tested protocols that are called as microservices, is manually executed first, and robustness tester has an engine that simulates the peers of a tested system and drives any protocol implementation through state machines to conduct a variety of systematic attacks when the target is at its weakest state. This simulation capability enables testing of complex protocol dialogues, it can generate lacks random testcases which are sent to the target to check the robustness of the microservice.

1. The interoperability check is done to see if the connection between the microservice and robustness testing tool is established.
2. If this check is failed, then the re-check of the connectivity, pinging and the configurations made are checked, also the suite that is used for that testcase is a valid suite or not matters.

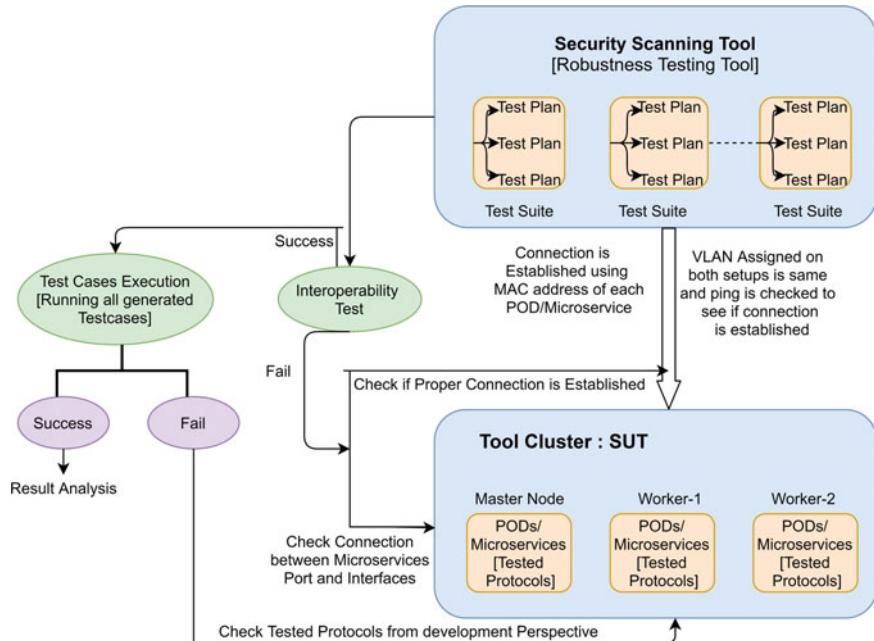


Fig. 2 Flow of manual execution of complete process

3. If the interoperability check is successful, then testcases execution is done, in this execution, if lacks testcases generated are passing, then the microservice is secure and less prone to the vulnerabilities, and the result analysis is done.
4. If any testcase in execution has failed, then the microservice or target server is analyzed based on development perspective. This execution flow is shown in the Fig. 2.

The result of robustness testing tool can be collected after manual execution and after automating. The results are in xml, html format or any txt format.

Manual Execution Report

The manual execution result can be seen in robustness scanning tool as well as in VM on Linux platform tool. This result is in xml format, the logs can be analyzed, it shows the number of testcases loaded, executed and failed, also, it shows the timeout and in detailed report can be seen in logs, and such logs are found in the directory that has been given in the command while execution. The manual execution result is shown in Fig. 3.

```

<?xml version="1.0"?>
<test-run-data>
    <suite>HTTP-Server</suite>
    <state>Finished</state>
    <case-index>720278</case-index>
    <elapsed-time>7200432</elapsed-time>
    <remaining-time>0</remaining-time>
    <executed-cases>720279</executed-cases>
    <remaining-cases>5271</remaining-cases>
    <passed-cases>720279</passed-cases>
    <failed-cases>0</failed-cases>
    <passed-snmp-cases>0</passed-snmp-cases>
    <failed-snmp-cases>0</failed-snmp-cases>
    <failed-snmp-traps>0</failed-snmp-traps>
    <passed-safe-guard-checks>1</passed-safe-guard-checks>
    <warning-safe-guard-checks>0</warning-safe-guard-checks>
    <failed-safe-guard-checks>0</failed-safe-guard-checks>
    <failed-isa-secure-cases>0</failed-isa-secure-cases>
    <failed-syslog-messages>0</failed-syslog-messages>
    <other-cases>0</other-cases>
    <case-rate-per-second>100.032745</case-rate-per-second>
    <total-cases>725550</total-cases>
    <completion>1000</completion>
</test-run-data>

```

Fig. 3 Manual execution, xml format test plan report

2.4 Automation of Test Plan Execution of Security Scanning Tool

The automation using RIDE, for robot framework test data RIDE, is a test copy-holder. The first task taken for automation is SEP deployment, this is done as per the deployment procedure, some pre-defined keywords like switch connection (to switch from one server to the other, it basically login to the other SSH), built-in-comment (used to for commenting any row), run keyword and return status (which is used to run the keyword that we have created and return the status or required result for that keyword), and many more such inbuilt keywords are used. The arguments required for testcases are given in argument section, suites also contain suite setup, for and setup requirements of suite, test case has a test setup in which login details are given, and then the timeout is set for each and every testcase and suites above that. Then, in robustness scanning process, a keyword is written to first load and open the tool using command line. After loading tool, the keyword for loading of suite, logging into test plan, opening and loading a test plan inside that particular suite using command line, checking interoperability to the final robustness testing process is written, and the log collection keyword is written to collect the results and logs of the execution. A sample screenshot of robot framework is shown in Fig. 4.

Documentation	Test Case Purpose: Purpose of this test cases is to perform robustness scan for <Suite name> protocol to test testplan interface robustness using external port [1234] exposed for outside traffic. *Author* xyz			
Setup	Security Test Setup			
Teardown	Security Test Teardown			
Timeout	50 minutes			
Template				
Tags	<input type="button" value="Security"/> <input type="button" value="<Add New>"/>			
1 Run Keyword And Ignore Error	SUT log collection	Before_test		
2 Start robustness scan	Testplan name			
3 Log	Generate html report after 1st execution			
4 Generate report	Testplan name	<code>\$(TEST_DIR)</code>	<code>\$(html_path)</code>	
5 Copy logs	run-time-info.xml			
6 Rerun failed tests				
7 Log	Generate html report after re-execution of failed TCs			
8 Generate report	Testplan name	<code>\$(rerun_path)</code>	<code>\$(rerun_path)</code>	

Fig. 4 Robustness testing automation script

```

20210430 07:13:13 Test run: 724593/725550 (99%) cases run, 109.9 cases/s
20210430 07:13:14 Test run: 724595/725550 (99%) cases run, 109.9 cases/s
20210430 07:13:15 Test run: 724847/725550 (99%) cases run, 109.9 cases/s
20210430 07:13:16 Test run: 724978/725550 (99%) cases run, 109.9 cases/s
20210430 07:13:17 Test run: 724980/725550 (99%) cases run, 109.9 cases/s
20210430 07:13:18 Test run: 725377/725550 (99%) cases run, 110 cases/s
20210430 07:13:19 Test run: 725550/725550 (100%) cases run, 110 cases/s
20210430 07:13:22 MESSAGE:Suite HTTP-Server executed.
20210430 07:13:22 MESSAGE:Test run verdict:PASS
20210430 07:13:22 MESSAGE:Executed cases:725550 Passed:725550
20210430 07:13:24 MESSAGE:Testplan executed.
20210430 07:13:24 Exit with code
+ KEYWORD BuiltIn.Comment Log, ${time out}
+ KEYWORD BuiltIn.Should Contain ${[REDACTED]_output}, Testplan executed

```

Fig. 5 Automation script log report of robustness testing tool from RIDE

Automation Report

The robot framework report can be analyzed in the log collection keyword that is written, this shows the detailed elapsed time of each testcase, and other details as manual result. If all testcases passed, then it indicates that the microservice tested is robust, and there are not threats. The robot log report is shown in Fig. 5.

Oamterm-[REDACTED]_Server_V4			
Settings >>			
1	Run Keyword And Ignore Error	SUT log collection	Before_test
2	Start robustness scan	Working_OAMT-[REDACTED]_Ser	
3	Log	Generate [REDACTED] report after 1st execution	
4	Generate report	Working_OAMT-[REDACTED]_Ser \${TEST_DIR}	\${html_path}
5	Copy logs	run-time-info.xml	
6	Rerun failed tests		
7	Log	Generate html report after re-execution of failed TCs	
8	Generate report	Working-[REDACTED]_Server \${rerun_path}	\${rerun_path}
9	Run Keyword And Ignore Error	SUT log collection	After_test
10			

Fig. 6 Automation script of OAM testcase

3 Case Study

Security scanning tools are used to analyze threats in real time using various common mechanisms. But this is the first time, automation on this tool has been exploited on a container level by enhancing the performance in a simple, quick and efficient manner. This case study covers a specific testcase scenario in which OAM is deployed as a group of containers (POD) of SEP deployment package, OAM is the treaty for positioning, monitoring and overhauling Ethernet metropolitan area network (MANs) and Ethernet WANs among numerous PODs. The test case in Kubernetes environment is networked as follows: The IPv4 IP and VLAN are allocated to OAM; robustness testing tool is loaded on a VM, and the peer IPv4 IP and same VLAN of OAM is assigned to an interface. ‘Ping’ is used to connect between a microservice and a virtual machine, and for security scanning of this microservice, a test plan is produced in the robustness scanning tool; this test plan is created manually under a specific suite in robustness scanning tool licence. During the execution of automation script shown in Fig. 6, it first enters basic configurations such as the destination IP address, port, VLAN, MAC address, subnet mask and gateway IP into the test plan, after which the script runs an interoperability test whose report (which can be seen manually) snapshot is shown in Fig. 7, with the source and destination IP addresses hidden on Fig. 7.

The test run tool then generates lakhs of testcases to assess the robustness of the microservice. Figure 5 depicts the results of the test run. The whole robustness scanning is performed in a containerized Kubernetes environment, with RIDE being used for the first time.

Log

```

2021-03-08 13:21:34.697 TEST CASE #0
2021-03-08 13:21:34.700 TCP.SYNACK-Reset valid
2021-03-08 13:21:39.785 Network device: eth0
2021-03-08 13:21:39.785 IPv4: [REDACTED]
2021-03-08 13:21:39.785 Ethernet: 5c:69:1d:00:00:00
2021-03-08 13:21:39.827 IPv4 1 [REDACTED] --> [REDACTED] TCP SYN 36 [view]
2021-03-08 13:21:39.845 IPv4 2 [REDACTED] <- 2 [REDACTED] TCP SYN-ACK 36 [view]
2021-03-08 13:21:39.849 IPv4 2 [REDACTED] --> 2 [REDACTED] TCP RST 32 [view]
2021-03-08 13:21:39.849 Test case #0 pass
2021-03-08 13:21:39.849 Test case #0 completed

```

Fig. 7 Interoperability test report of OAM

4 Result Analysis

For a given microservice in both cases, i.e., automation and manual, the execution time taken to run the testcases generated by the robustness scanning tool after an interoperability test is same. In automation execution, script is written for one microservice, and it can be used for any other similar security scan of microservices. This saves a lot of time and eliminates manual interruptions by eliminating the requirement for continuous process monitoring. After the automation, script is generated. A single step procedure is required to load test plans, test interoperability, execute testcases and record the results in logs. The amount of time it takes an operator to obtain a software security result has been reduced, and the script may be reused. Manual execution, on the other hand, necessitates redoing the entire operation, increasing the risk of human mistake and lengthening the time required for security screening.

5 Conclusion and Future Scope

In this paper, the full execution of security scanning tool's mechanism and the architecture of the SUT is presented. The automation of manual execution at container level is developed, and this saves time and prevents disruptions in the operation. To activate the testcases manually, it takes a long time, which requires continual monitoring. The automation script can be reused and is available for project management. The HTML format of the report released will provide stakeholders with a clear picture.

The robot framework is an excellent platform for automating such tools, test plans are developed manually here and then taken for automation, and thus, test plan creation can be attempted in the future utilizing RIDE.

References

1. Rangnau T, Buijtenen RV, Fransen F, Turkmen F (2020) Continuous security testing: a case study on integrating dynamic security testing tools in CI/CD pipelines. In: IEEE 24th international enterprise distributed object computing conference (EDOC). IEEE, pp 145–154
2. Patki P, Gotkhindikar A, Mane S (2018) Intelligent fuzz testing framework for finding hidden vulnerabilities in automotive environment. In: IEEE fourth international conference on computing communication control and automation (ICCUBEA). IEEE, pp 1–4
3. Jayakody JADCA, Perera AKA, Perera GLAKN (2019) Web-application security evaluation as a service with cloud native environment support. In: International conference on advancements in computing (ICAC). IEEE, pp 357–362
4. Sinchana K, Sinchana C, Gururaj HL, Sunil Kumar BR (2019) Performance evaluation and analysis of various network security tools. In: International conference on communication and electronics systems (ICCES).). IEEE, pp 644–650
5. Wotawa F (2016) On the automation of security testing. In: International conference on software security and assurance (ICSSA). IEEE, pp 11–16
6. Chang CC, Yang, Yen EH, Jeng JY (2017) A kubernetes-based monitoring platform for dynamic cloud resource provisioning. In: IEEE global communications conference. IEEE, pp. 1–6
7. Amir KC, Goulart A, Kantola R (2016) Keyword-driven security test automation of customer edge switching (CES) architecture. In: 8th international workshop on resilient networks design and modeling (RNDM). IEEE, pp. 216–223
8. Alzahrani A, Alqazzaz A, Zhu Y, Fu H, Almashfi N (2017) Web application security tools analysis. In: IEEE 3rd international conference on big data security on cloud (bigdatasecurity), ieee international conference on high performance and smart computing (hpsc), and ieee international conference on intelligent data and security (ids). IEEE, pp 237–242

Rad-Hard Model SOI FinTFET for Spacecraft Application



Ajay Kumar Dharmireddy, Sreenivasa Rao Ijjada, K. V. Gayathri, K. Srilatha, K. Sahithi, M. Sushma, and K. Madhavi

Abstract Radiation-hardened model silicon on insulator fin gate tunnel field effect transistor (SOI FinTFET) devices are used for spacecraft applications to protect from radiation effects. In this, radiation environment revealed that the radiation-induced effect degradation of device performance and also trap charge carriers should not be neglected. Enormous simulations were carried out to examine the production of electron and holes in the gate oxide and to predict the characteristics of device from sub-threshold to inversion region. The gamma radiation model of Sentaurus Technology Computer-Aided Design (TCAD) was used to determine the radiation properties of SOI FinTFET.

Keywords Radiation hardened · Spacecraft applications · Sub threshold region · SOI FinTFET · Gamma radiation · Sentaurus TCAD

1 Introduction

The universe is a huge widespread open space that would be hold on to massive quantities of radiation. Man is always searching for other planets' information and also observe our earth condition with help of satellites. Satellites are used to a clearer sense of information about the universe and solar system. A numeral of automated and worked space satellites have been developed around the world in recent decades. Satellites with millions of electronic circuits are launched into orbit to scan and process the information collected. Most of these electronic devices are intolerant and regularly failed by the effect of cosmic radiation. Due to the widespread development is required to design of radiation tolerant device. The “radiation hardening [1]” of a device refers to the process of fabricating devices that are radiation-resistant. The

A. K. Dharmireddy (✉) · K. V. Gayathri · K. Srilatha · K. Sahithi · M. Sushma · K. Madhavi
SIR C. R. Reddy College of Engineering, Eluru, Andhra Pradesh 534007, India

S. R. Ijjada
GITAM Institute of Technology, Vishakhapatnam, Andhra Pradesh 530016, India

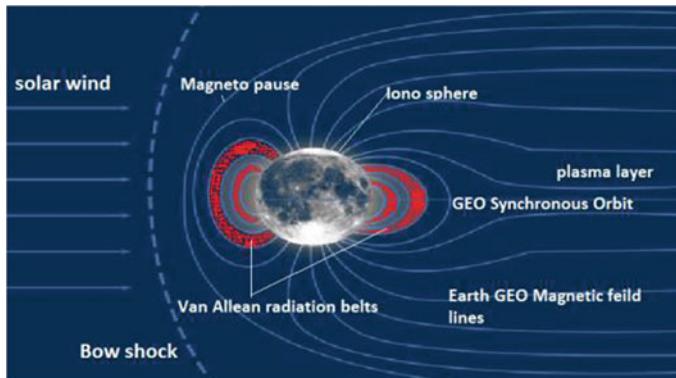


Fig. 1 Strong earth magnetosphere trapped high concentrated energetic protons to form Van Allen belt

term “Rad-hard devices” refers to devices that have been “hardened” to the effects of radiation.

The impacts of space environmental [2] influences such as trapping and ambient radiation, as well as solar and galactic cosmic rays, on spacecraft electronics are explored [3]. Space radiation effect [4] is the major reason of malfunction, compatibility issues, and potential failure of electronic device. Figure 1 shows the gravitational perturbations formed when charged particles from the cosmic rays interact with the magnetism and expand it sun-ward. The trapping particles are found in the Van Allen radiation belts, which become two doughnut-shaped magnetic circles that orbit the Earth. A reasonably constant distribution of protons with vitalities more than 10 meV are being located in the inner belt. Electrons with vitalities up to 10 meV are mostly located in the outer belt.

The two main effects of radiation on CMOS electronics are as follows is shown in Fig. 2. First one, single event effect [5] occurred by heavy ion hit the single event of electron in semiconductor device causes different forms: single event transient or upset (SET’s or SEU’s) [6] is occurred when the transient pulse change the logic operation of device will cause soft error. The single-event functional interrupt (SEFI) is an extreme SEU that bases the system to enter a test mode, pause, or unknown condition due to a SEU in the device’s control circuitry. Single event latch-up (SEL’s) occurred when potentially damaging device based on produced high current in beyond the requirement level will cause hard error or permanent damage of the circuit. SEL strongly depends on temperature sensitive. SEGR is a disruptive burnout caused by the creation of a charge transport dielectric breakdown in the gate oxide.

The second one, total ionizing dose (TID) [7] is an ionization mechanism bases long-term destruction of electrons in valence bands get allowing them to join the conduction bands due to the cumulative energy stored in a material. The unit of TID is “rads” or “Grays.” Trapped electrons, trapped protons, and solar protons are major causes of TID distribution in the space region. Distinctive effects also include parametric failures, or change in system parameters such as leakage current, threshold

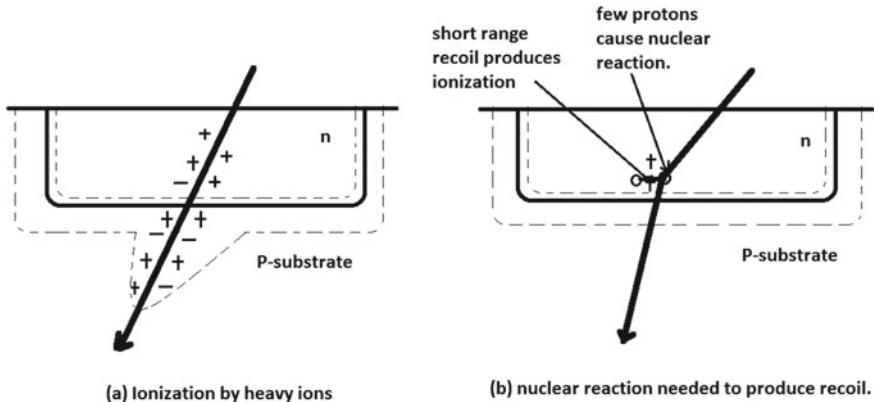


Fig. 2 Space radiation effects on semiconductor electronic device

voltage shift, etc. Shielding, derating, and moderate circuit architecture will all help to minimize TID on satellite systems.

Based on Kane's method calculate tunneling current (I_D) which is done by assimilating the BTB generation rate of carriers (G_{BTB}) on the device volume.

In this model we using BTBT approach.

$$I_{\text{BTBT}} = q \int G_{\text{BTB}} dV = q \int \mathcal{A} \frac{|E_F|^{\mathcal{G}}}{E_g^{1/2}} \exp\left(-\mathcal{B} \frac{E_g^{3/2}}{|E_F|}\right) dV \quad (1)$$

where

$$\mathcal{A} = \frac{q^2 \sqrt{2m_{\text{tunel}}}}{h^2 \sqrt{E_g}}, \quad \mathcal{B} = \frac{\pi^2 E_g^{3/2} \sqrt{m_{\text{tunel}}/2}}{qh}, \quad m_{\text{tunel}} = \frac{m_0 m_e m_h}{m_e + m_h};$$

where \mathcal{A} and \mathcal{B} are defined by material-dependent Kane parameters, band gap materials and effective electron (or) hole mass. E.g. is the energy band gap. Regional and average electrical fields (E_{avg}) are indicated in the E_F component. Here, m_0 is the portion of the mass electron and m_e, m_h are the effective mass of electrons and holes.

2 Architecture and Specifications of Proposed Device

A cross section view of SOI-based FinTFET [8–10] device is shown in Fig. 3a. Table 1 lists show the device parameters used during the simulation. The gate oxide thickness is taken as 5 nm to quantify the effect of radioactivity [11, 12] on the SiO₂. The proposed device specifications are: the gate material aluminum (AU) is used, gate length is 100 nm, and gate thickness is 5 nm. The total silicon bar length is 200 nm,

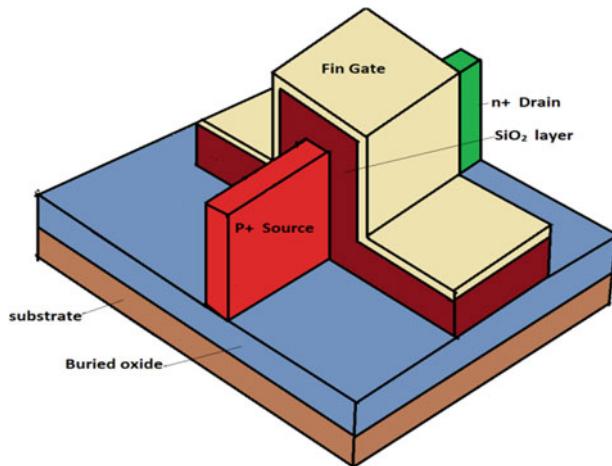


Fig. 3 Schematic diagram of SOI FinTFET structure

Table 1 Device specifications

Region	Length (L) (nm)	Thickness (t) (nm)	Doping concentration (cm^{-3})	Material
Gate	100	5	—	Au
P+ source	50	20	1.0×10^{19}	Si
P channel	100	20	1.0×10^{15}	Si
N+ drain	50	20	5.0×10^{18}	Si
Gate oxide	100	5	—	SiO_2
Buried oxide	200	100	—	SiO_2

and in this, p-type source and drain length 50 nm, thickness is 20 nm, and source-side doping concentration is $1.0 \times 10^{19} \text{ cm}^{-3}$ and drain-side doping concentration is $5.0 \times 10^{18} \text{ cm}^{-3}$. p-type channel length is 100 nm, thickness is 20 nm, and doping concentration is $1.0 \times 10^{15} \text{ cm}^{-3}$. Silicon oxide material used in the gate oxide and buried oxide.

Gate oxide length and buried oxide length, thickness are 100 nm and 5 nm and 200 nm and 100 nm. The Sentaurus TCAD simulation tool was carried out for finite element simulation program. It can solve nonlinear Poisson's equations and charge carrier continuity equations. Sentaurus TCAD models, i.e., SRH recombination model, gamma radiation model, nonlocal BTBT model, drift-diffusion model, and radiative recombination model were used to predict the device's electrical characteristics such as electron and hole trapping (EHT) dynamics, drain current, and threshold voltage transition.

3 Results and Discussion

In order to assess the impact of various forms of radiations, we have developed a SOI FinTFET device in gamma radiation, alpha particles, and heavy ions. In this description, we used efficiently measured radiation models on the SOI FinTFET on the device as well regional imitation of different types of radiation.

Figure 4a appearsances the radiation dose consequence on the FinTFET energy band at $V_{GS} = 0.4$ V and $V_{DS} = 0.6$ V. Since SOI FinTFET is a P-I-N architecture, the influence of holes trapped in the Si/SiO₂ interface is mainly focused on the source-channel tunnel interface due to the radiation dose. As result, the energy band goes down as the radiation increase.

Figure 4b displays a comparison of SOI FinTFET drain current characteristics to analyze the performance of output current due radiation of gamma rays. As dosage increases from 50 krad to 100 Mrad in the SiO₂, some major variations in the output current can be seen in the sub-threshold region as analogized with ON-state. As result, it can be determined that SOI FinTFET with stand radiation in the active region, allowing it to be used as a radiation hard-end device in integrated circuits (ICs) with difficult radiation environment. The same radiation was occurred on SOI Fin TFETs for comparative purposes and drain current was attained.

Figure 5a illustrates the effect of various radiation doses on FinTFET surface potential as a function of channel length. The generation of electron–hole pairs (EHPs) in the gate oxide region and Si/SiO₂ interface is observed as the dose capacity improve from 100 krad to 100 Mrad in SiO₂. The potential curve of the system will be changed due to these EHPs. Figure 5b shows the transient difference in the radiation production with a dose of 50 krad (Si). The attained results show that as rising the dose duration proportionally increase the radiation generation this is due to the formation of EHPs in the FinTFET caused by gamma radiation. Due to the accumulation of holes, the radiation production is higher at the source-channel interface as compared to the drain-channel interface

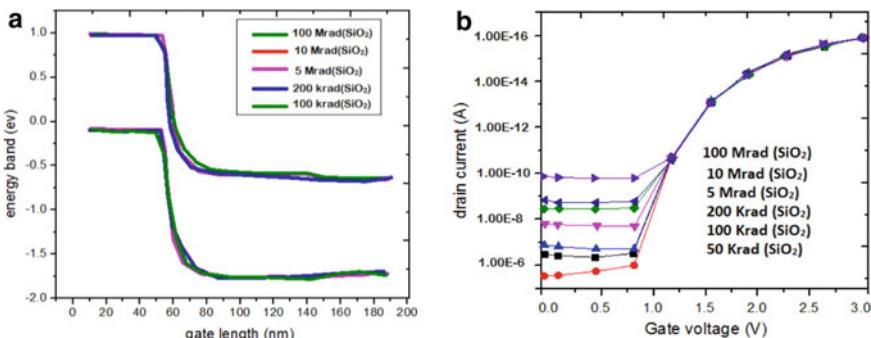


Fig. 4 **a** Variation in the SOI Fin Tunnel FETs energy band with varying doses around the channel range. **b** Drain current transition characteristics of SOI-based FinTFETs in the presence of gamma radiation

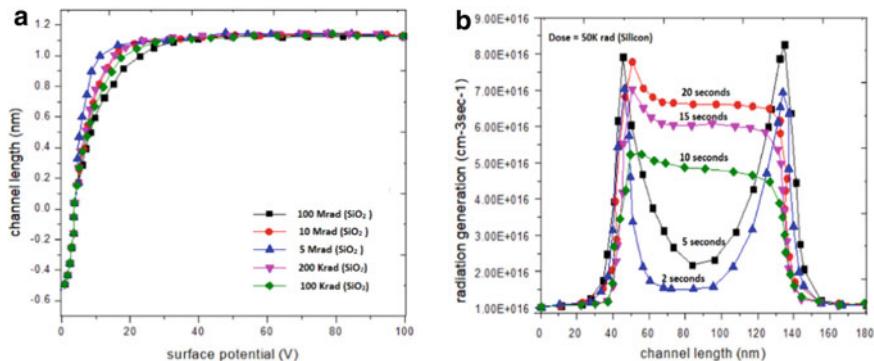


Fig. 5 **a** Radiation's effect on FinTFET surface potential at various ionizing doses. **b** Depicts the radiation production in FinTFET for a dose of 50 krad (Si) as a function of channel duration with various dose period

4 Conclusion

In this paper, TCAD simulation-based study of the TID effect on SOI FinTFET and its potential application as a dosimeter done. With the variance of ionizing radiation, electrical parameters such as drain current, trap position, V_{th} , $I_{\text{on}}/I_{\text{off}}$ ratio, and radiation production intensity are derived in the simulation. The change in drain current in sub-threshold region is strongly dependent on the dose of gamma radiation. However, owing to ionizing radiation, the variation of drain current in the on-state region is very small. Whenever it gets to withstanding extreme radiation, SOI FinTFET device is supplementary radiation tolerant device as compared with SOI MOSFET in order to verify radioactive functionality.

References

1. Kerns SE, Shafer BD, Rockett LR Jr, Primore JS, Berndt DF, van Vonno N, Barber FE (1988) The design of radiation-hardened ICs for space: a compendium of approaches. Proc IEEE 76:1470–1509
2. Connerney JEP, Adriani A, Allegrini F, Bagenal F, Bolton SJ, Bonfond B (2017) Jupiter's magnetosphere and aurorae observed by the Juno spacecraft during its first polar orbits. Science 356:826–832. <https://doi.org/10.1126/science.aam5928>
3. Baker DN, Erickson PJ, Fennell JF, Foster JC, Jaynes AN, Verronen PT (2017) Space weather effects in the Earth's radiation belts. Space Sci Rev 214:60–68. <https://doi.org/10.1007/s11214-017-0452-7>
4. Miura-Mattausch M, Kikuchihiara H et al (2020) Compact modeling of radiation effects in thin-layer SOI-MOSFETs. In: International conference on simulation of semiconductor processes and devices (SISPAD), pp 319–322. <https://doi.org/10.23919/SISPAD49475.2020.9241636>
5. Sayyah R, Macleod TC, Ho FD, Radiation-hardened electronics and ferroelectric memory for space flight systems. Ferroelectrics 413:170–175. <https://doi.org/10.1080/00150193.2011.554145>

6. Maurer RH, Fraeman ME, Martin MN, Roth DR (2008) Harsh environments: space radiation environment, effects and mitigation. *J Hopkins APL Tech Dig* 28:17–29
7. Xi K, Bi J, Chu J, Xu G, Li B, Wang H, Liu M, Sandip M (2020) Total ionization dose effects of N-type tunnel field effect transistor (TFET) with ultra-shallow pocket junction. *Appl Phys A* 126:440–447. <https://doi.org/10.1007/s00339-020-03622-2>
8. Ajaykumar D, Ijjada SR, Tejomurthy PHS (2019) Performance analysis of tri-gate SOI FINFET structure with various fin heights using TCAD simulation. *J Adv Res Dyn Control Syst* 11:1291–1298
9. Dharmireddy AK, Sharma A, Babu MS, Ijjada SR (2020) SS < 30 mV/dec; hybrid tunnel FET 3D analytical model for IoT applications. *Mater Today Proc*. <https://doi.org/10.1016/j.matpr.2020.09.367>
10. Ajaykumar D, Avinash S, Krishna Allu MV, Lavanya K, Anjani Devi N, Jyothsna U, Ijjada SR (2021) 3D analytical modeling of surface potential and threshold voltage model of Dm Fintfet with dual hetero gate oxide structure. *Turk J Comput Math Educ* 12:1245–1259. <https://doi.org/10.17762/turcomat.v12i2.1180>
11. Garg R, Jayakumar N, Khatri SP, Choi G (2006) A design approach for radiation-hard digital electronics. In: Design automation conference, pp 773–778. <https://doi.org/10.1145/1146909.1147105>
12. Roy K, Mukhopadhyay S, Mahmoodi-Meimand H (2003) Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. In: Proc IEEE 91(2)

Direct Synthesis Method-Based PID Controller for Critically Damped Time Delay Systems



E. Govinda Kumar, R. Vinoth Kumar, and D. Deivasigamani

Abstract In this work, design a proportional-integral-derivative (PID) controller using direct synthesis method for the control of critical damped second order plus time delay (CDSOPTD) systems. The different categories of CDSOPDT systems are considered based on the ratio of time delay to system time constant. CDSOPTD systems performance is analyzed using PID controller. The gains of PID controller are obtained using direct synthesis method with tuning parameter α ranging from 0.1 to 1. The various PID parameters are used for the control CDSOPTD systems, and the simulation results are obtained. For low value of α , the closed-loop responses exhibit that, the unstable behavior, oscillations in a controlled variable, and peak overshoot in responses. In direct synthesis method, the higher value of tuning parameter α is used for PID controller tuning, and it is delivered better closed-loop performance for the control of CDSOPTD systems.

Keywords PID controller · Direct synthesis method · Critical damped system · Second order system · Controller tuning

1 Introduction

An electrical, biomedical, aerospace and chemical engineering applications are operated based on the concept of process control with different forms controller. The controller is needed to achieve a required closed-loop performance which is designed based on simple in nature and robust control. In laboratory, equipments and process plants are controlled using proportional-integral-derivative (P-I-D)/proportional-integral (P-I) controllers. The P-I/P-I-D controller contains the following advantages,

E. Govinda Kumar

Department of Electronics and Instrumentation Engineering, Karpagam College of Engineering, Coimbatore, Tamil Nadu 641032, India
e-mail: govindakumar@kce.ac.in

R. Vinoth Kumar (✉) · D. Deivasigamani

Department of Electronics and Instrumentation Engineering, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu 641032, India

robustness, adequate closed-loop performance and simple structures [1]. Also, P-I-D controllers are easy to use for any control applications based on their easiness, and functionality of the controllers is very strong [2]. The control loops of many process plants are conveyed; superior closed-loop performance with the controller parameters of P-I-D controller is properly tuned and designed [3]. The Cohen–Coon and Ziegler–Nichols (ZN) approaches are regularly used to fit the P-I/P-I-D controller settings and attain the superior closed-loop performance. These controller parameters are applied for the control of non-minimum and minimum phase systems, integrating system, unstable and stable systems. It delivers better set-point tracking and robustness for the process parameters uncertainty [4]. In recently, the controller parameters of PID are tuned by simple and analytic approach, and it is used for the control of non-linear process [5]. The process transfer functions are necessary for design P-I/P-I-D controller, and mostly, first-order with time delay model is used for tune the controller parameters [6, 7]. Although, the response of stable process contains with and without overshoot responses, and this type of system is represented by critical damped and under-damped second-order plus time delay model, respectively [7, 8]. An under-damped second-order plus time delay system is controlled using Z-N approach-based P-I/P-I-D controller [9]. A second-order plus time delay (SOPTD) system is controlled by an internal model control (IMC) structure is combined with PID controller. This combination control technique is conveyed superior set-point tracking and achieve the improved closed-loop performance in terms of parameters uncertainty [10]. The controller setting of PID controller for a critical damped second-order plus time delay (CDSOPTD) system is obtained by an on-line tuning technique [11].

Design of PID controller using equating coefficient method (ECM) technique is delivered better closed-loop performance of CDSOPTD and set-point tracking than the IMC method-based tuning [12]. The time/frequency domain performance criterion is used for the design and tuning of P-I-D controller. The closed-loop performance criterion is not directly related with closed-loop dynamic performance. Henceforth, P-I-D controller is designed based on the desired closed-loop performance using direct synthesis (DS) approach. The P-I-D controller settings are assessed analytically by using DS method, and the desired closed-loop responses are coordinated with set-point variable [13]. This approach is used to design P-I/P-I-D and applied for the control of many systems such as stable system, integrating and unstable system with and without time delay systems [13, 14]. A higher order system is approximated by FOPTD or SOPTD model using time delay or measurement delay (L) [15]. Recently, a tuning-free PID control settings is introduced for the control of critically damped SOPTD stable systems [16]. In this work, CDSOPTD system is controlled using PID controller, and the parameters settings are tuned with direct synthesis method. The different categories of CDSOPTD systems are considered for further analyzing the controller performance.

The paper is structured as follows. The model of CDSOPTD systems are discussed in Sect. 2. Designing approach of PID controller based on direct synthesis method for CDSOPTD system is presented in Sect. 3. Section 4 contains the responses

of different categories of CDSOPTD systems and the controller performances are analyses. Conclusions are drawn in Sect. 5.

2 Critical Damped SOPTD System

Mostly, second-order linear system is considered as processes dynamics of process industry, and it is represented by,

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\delta\omega_n s + \omega_n^2} \quad (1)$$

where ω_n and δ are natural frequency and damping ratio, respectively. The variables $R(s)$ and $C(s)$ are the Laplace transform of input and output variables, respectively. An under-damped, critically damped and over-damped second-order system responses are obtained based on the value of damping ratio. Equation (2) is represented as differential equation of second-order plus dead time system,

$$a^2 \frac{d^2 c(t)}{dt^2} + 2\delta a \frac{dc(t)}{dt} + c(t) = K_p r(t - L) \quad (2)$$

where $r(t)$ and $y(t)$ are output and input variables, respectively. The four system parameters are the steady-state gain K_p , damping ratio δ , second-order system time constant a and dead time L . The damping factor $\delta = 1$ for critically damped system, and it is given by,

$$a^2 \frac{d^2 c(t)}{dt^2} + 2a \frac{dc(t)}{dt} + c(t) = K_p r(t - L) \quad (3)$$

By applying Laplace transform of the above equation and the transfer function model of critical damped, SOPDTD model is represented by,

$$\frac{C(s)}{R(s)} = \frac{K_p e^{-Ls}}{as^2 + 2as + 1} \quad (4)$$

The simplified form of Eq. (4) is,

$$\frac{C(s)}{R(s)} = \frac{K_p e^{-Ls}}{(as + 1)^2} \quad (5)$$

Using the above form of CDSOPTD, models are used to tune the PID controller settings. The parameters time constant a and dead time L are used to obtain the different categories of CDSOPTD models in this work.

3 PID Controller Design with Direct Synthesis Method

The transfer function form of a unity feedback critical damped SOPTD system is,

$$\frac{C(s)}{R(s)} = \frac{G_c(s) G_p(s)}{1 + G_c(s) G_p(s)} \quad (6)$$

where $G_c(s)$ and $G_p(s)$ are controller and critical damped SOPDT system transfer functions, respectively. The series form of PID controller structure is given by Eq. (7), and it is considered for controller parameter settings.

$$G_c(s) = k_c \left(1 + \frac{1}{\tau_I s} \right) (1 + \tau_D s) \quad (7)$$

where k_c is the proportional gain, τ_I is the integral constant and τ_D is derivative constant. The parallel form of PID controller settings is obtained by using following equations. The proportional gain is represented as,

$$K_C = \frac{k_c(\tau_I + \tau_D)}{\tau_I} \quad (8)$$

The integral time constant is characterized by,

$$T_I = \tau_I + \tau_D \quad (9)$$

The derivative time constant is represented by,

$$T_D = \frac{\tau_I \tau_D}{\tau_I + \tau_D} \quad (10)$$

The characteristics equation of unity feedback system is,

$$1 + G_c(s) G_p(s) = 0 \quad (11)$$

The above equation can be written as,

$$1 + k_c \left(1 + \frac{1}{\tau_I s} \right) (1 + \tau_D s) \frac{K_p e^{-L s}}{(a_1 s + 1)^2} = 0 \quad (12)$$

The dead time $e^{-L s}$ is approximated using Pade's approximation, and it is denoted as,

$$e^{-L s} = \frac{1 - L s / 2}{1 + L s / 2} \quad (13)$$

The time constant of critical damped SOPDT is related to τ_I of the PID controller and 25% of τ_I is considered as τ_D of PID controller. For the simplification of derivative time τ_D is considered as,

$$\tau_D = \frac{1 + \alpha \tau s}{1 + \tau s} \quad (14)$$

where α is the tuning parameter, and the value of α is considered between 0.1 and 1 in the high frequency region. Equation (14) as one with a range of low to high frequencies. Substituting Eq. (13) in Eqs. (12) and (15) is represented after cancelling the terms τ_I and τ_D .

$$1 + \frac{K_p k_c}{\tau_I s} \frac{(1 - 0.5Ls)}{(1 + 0.5Ls)} \alpha = 0 \quad (15)$$

The parameter α as tuning parameter of direct synthesis method. Equation (16) is represented as characteristics equation of a critical damped SOPTD system.

$$\tau_{cl}^2 s^2 + 2\tau_{cl}s + 1 = 0 \quad (16)$$

where τ_{cl} is the closed-loop time constant of a system. The desired form of characteristics equation of second-order system is achieved by the value of damping ratio is 0.707.

$$\tau_{cl}^2 s^2 + 1.414\tau_{cl}s + 1 = 0 \quad (17)$$

The following relations are derived by the comparison of s^2 and s coefficients in Eqs. (12) and (17).

$$\tau_{cl}^2 = \frac{0.5L\tau_I}{\alpha K_p k_c} \quad (18)$$

$$0.5L + 2\tau_{cl} = \frac{\tau_I}{\alpha K_p k_c} \quad (19)$$

The value τ_{cl} is determined by deploying the above equations which represented as,

$$\tau_{cl}^2 - L\tau_{cl} - 0.25L^2 = 0 \quad (20)$$

By solving Eq. (17) and find the value of τ_{cl} which is used in Eq. (18), k_c is obtained. The derivative time τ_D is determined based on the tuning parameter α .

4 Closed-Loop Responses of CDSOPTD Systems

The closed-loop responses of a critical damped SOPTD system is obtained with a unit step change as reference input. The controller settings of PID controller are tuned by direct synthesis method. In this work, consider the different categories of CDSOPTD systems which are categorized based on the system parameters a and L . Using DS method, the setting of PID controller is tuned with different categories of critical damped SOPDT system models. These models are chosen based on the values of ratio between L and a which are ranging from 0.1 to 1.

All the closed responses of CDSOPTD systems with a unit step input are obtained using MATLAB. The PID parameters of CDSOPTD systems are calculated using the direct synthesis method with the different values of α . The ratio between L and a for critical damped SOPTD is 0.17 is considered in first case, and it is given below (CASE I).

$$\frac{C(s)}{R(s)} = \frac{1 e^{-2.64s}}{(14.9s + 1)^2} \quad (21)$$

The controller settings k_c , τ_I and τ_D are tuned with the different values of α which is varies from 0.1 to 1, and the controller settings are given in Table 1. The servo response of above system is shown in Fig. 1. The observations of different responses are noted, and it is given in Table 1.

The second case of CDSOPTD system is also similar as above, the ratio of L/a is 0.5 is considered, and it is represented by (CASE II),

$$\frac{C(s)}{R(s)} = \frac{1 e^{-9.88s}}{(19.76s + 1)^2} \quad (22)$$

Using direct synthesis method, the controller setting of the above system is tuned and given in Table 2.

Table 1 Closed-loop observation of CASE I system with ratio L/a is 0.17

Tuning parameter α	PID controller parameters			Observations
	K_C	K_I	K_D	
0.1, 0.2	21.28, 11.61	1.299, 0.649	28.84	Unstable response
0.3	8.38	0.433		Responses are oscillatory (Fig. 1a)
0.4–0.7	6.77, 5.81, 5.16, 0.469	0.325, 0.26, 0.216, 0.185	28.83	Stable response with peak overshoot (Fig. 1b)
0.8, 0.9	4.35, 4.09	0.162, 0.144	28.83	Stable response with peak overshoot is reduced (Fig. 1b)

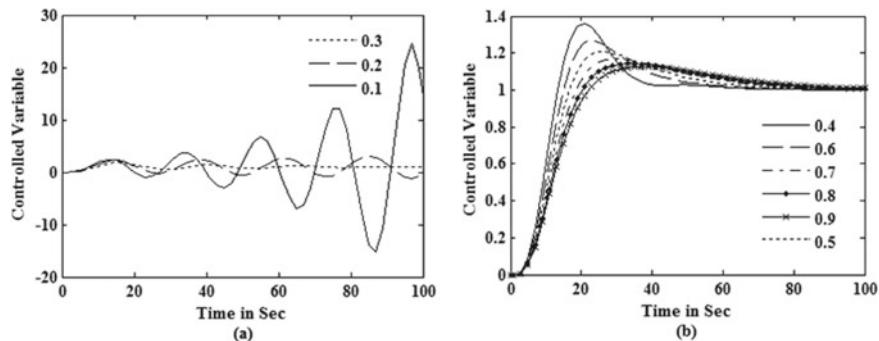


Fig. 1 Servo responses of CDSOPTD system with ratio L/a is 0.17

Table 2 Closed-loop observation of CASE II system with ratio L/a is 0.5

Tuning parameter α	PID controller parameters			Observations
	K_C	K_I	K_D	
0.1, 0.2	7.546, 4.116	0.347, 0.174	13.553	Unstable response and
0.3	2.973	0.116	13.553	Oscillatory closed-loop responses (Fig. 2a)
0.4–0.7	2.401, 2.058, 1.829, 1.666	0.087, 0.069, 1.829, 0.05	13.553	Stable responses with peak overshoot (Fig. 2b)
0.8, 0.9	1.544, 1.448	0.043, 0.039	13.553	Stable response with peak overshoot is decreased (Fig. 2b)

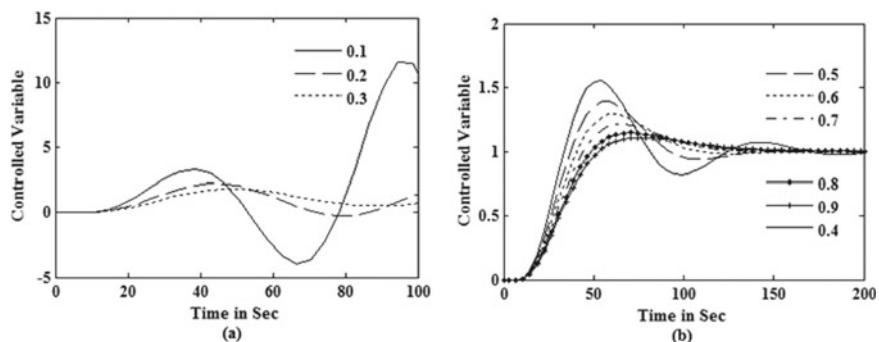


Fig. 2 Closed-loop responses of CDSOPTD system with ratio L/a is 0.5

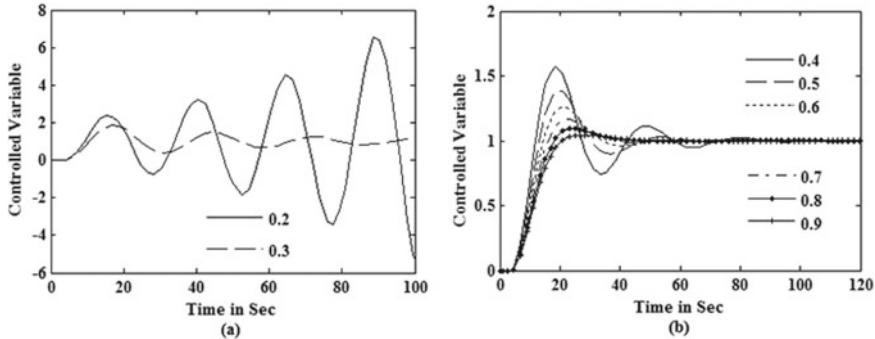


Fig. 3 Servo responses of CDSOPTD system with ratio L/a is 1

Similarly, parameter α is changing from 0.1 to 1, and the corresponding closed-loop responses are shown in Fig. 2. It shows that the closed-loop responses and the responses are obtained for the control of CDSOPTD systems with ratio L/a is 0.5. The different responses are obtained by varying the tuning parameter α from 0.1 to 1, and the observation of closed-loop performances are listed in Table 2.

The ratio L/a is considered as 1 for further analyzing the performance of PID controller, and it is represented by (CASE III),

$$\frac{C(s)}{R(s)} = \frac{1 e^{-6.24s}}{(6.24s + 1)^2} \quad (23)$$

It is controlled by using PID controller, and the controller settings are obtained by using direct synthesis method. The resultant CDSOPTD system responses are shown in Fig. 3 (CASE III).

The performances of direct synthesis method-based PID controller tuning is given in Tables 1 and 2. The PID controller settings of CDSOPTD system is tuned with the lesser value of parameter α (0.1, 0.2, 0.3) and controlled variables are unstable and oscillating behavior. The controller settings of PID controller is obtained with high value of α (>0.4). The performances of a closed system are enhanced, and it is shown in Figs. 1b, 2b and 3b. These responses are stable with reduced amount of peak overshoot.

5 Conclusion

The PID controller parameters of CDSOPTD systems are tuned using different tuning parameter values in a direct synthesis method. From the simulation responses, the tuning parameter α ranging from 0.1 to 0.3 (lower value)-based PID controller settings are delivered poor closed-loop performance. For the value of tuning parameter α ranging from 0.7 to 0.9 (higher value) is used to tune the setting of PID

controller, it provided better closed-loop performances in terms of stable responses with less peak over shoot and lower settling time. For intermediate tuning, parameter values are provided; the closed-loop responses are stable and oscillations are occurred in a controlled variable. In future work, the closed-loop performance of CDSOPTD system with direct synthesis-based PID controller is improved with optimizing the tuning parameter α .

References

1. Astrom K, Hagglund T (2005) Advanced PID control, research. Instrument Society of America, Triangle Park, NC
2. Hote YV, Jain S (2018) PID controller design for load frequency control: past present and future challenges. *IFAC Pap Online* 51(4):604–609
3. Luyben WL (2001) Effect of derivative algorithm and tuning selection on the PID control of dead-time processes. *Ind Eng Chem Res* 40(16):3605–3611
4. Darwish NM (2018) PID controller design in the frequency domain for time-delay systems using direct method. *T I Meas Control* 40(3):940–950
5. Zhao C, Guo L (2017) PID controller design for second order nonlinear uncertain systems. *Sci China Inf Sci* 60(2):022201
6. Stephanopoulos G (1990) Chemical process control: an introduction to theory and practise. Prentice Hall, New Delhi
7. Lui T, Gao F (2011) Industrial process identification and control design. Springer, London
8. Yu CC (2006) Auto tuning of PID controllers, 2nd edn. Springer, London
9. Basilio JC, Matos SR (2002) Design of PI and PID controllers with transient performance specification. *IEEE Trans Educ* 45(4):364–370
10. Ranganayakulu R, Uday Bhaskar Babu G, Seshagiri Rao A (2017) Fractional filter IMC-PID controller design for second order plus time delay processes. *Cogent Eng* 4(1):1366888
11. Ram VD, Chidambaram M (2015) On-line controller tuning for critically damped SOPTD systems. *Chem Eng Commun* 202(1):48–58
12. Santosh S, Chidambaram M (2015) Tuning of proportional integral derivative controllers for critically damped second-order plus time delay systems. *Indian Chem Eng* 57(1):32–51
13. Chen D, Seborg DE (2002) PI/PID controller design based on direct synthesis and disturbance rejection. *Ind Eng Chem Res* 41(19):4807–4822
14. Suganya G, Govinda Kumar E (2018) Control of third order processes using optimized synthesis PI controller. *Int J Eng Technol* 7(24):200–204
15. Narasimha Reddy S, Chidambaram M (2020) Model identification of critically damped second order plus time delay systems. *Indian Chem Eng* 62(1):67–77
16. Yerolla R, Bestha CS (2021) PI/PID controller design for critically damped SOPTD system and experimental validation. In: 2021 5th IEEE international conference on intelligent computing and control systems (ICICCS), pp 531–535

Exploiting Full-Duplex Relaying in Vehicular Cooperative NOMA for Residual Self-interference in Amplify and Forward and Decode and Forward with Incremental Relaying



Sravani Potula, Sreenivasa Rao Ijjada, and Karunakar Reddy Santhamgari

Abstract In this paper, the downlink cooperative non-orthogonal multiple access (NOMA) is considered in vehicular networks with full-duplex relay employing amplify and forward (AF) or decode and forward (DF) relaying protocol. We demonstrate performance of the proposed system using linear energy harvesting (EH) with power splitting protocol and incremental relaying to reduce the influence of residual self-interference at relay vehicle. The simulation results depict the effects of power division ratio on outage probability (OP) at vehicles (V_1 and V_2) when relay vehicle operates in AF or DF. The vehicle close to the roadside units (RSU) achieves better outage behavior even at low SNR region, and it is observed that outage behavior of far end user is better with DF rather than AF.

Keywords Vehicular networks · Non-orthogonal multiple access · Energy harvesting · Full duplexing · Amplify and forward · Decode and forward · Incremental relaying

1 Introduction

Recent advancements in IoT and latest technologies in wireless communications have supplanted intelligent transport system (ITS) by providing high data rates, better reliability and less delay. Due to the dynamic nature of the vehicles and stringent time constraints in vehicular communication, it has been an interesting topic to researchers, and many issues are still unresolved.

S. Potula (✉) · S. R. Ijjada · K. R. Santhamgari

Research Scholar, Department of Electronics and Communication Engineering, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India

S. R. Ijjada

e-mail: sijjada@gitam.edu

Contrasting to orthogonal multiple access (OMA), non-orthogonal multiple access (NOMA) serves multiuser by assigning different power with similar time and frequency [1]. Superposition coding is used at transmitter to superimpose multi-signals and successive interference cancellation [SIC] is used to decode the information successfully at receiver [2]. To further increase the fairness, link reliability and service coverage in vehicular communication, many researchers attempted to integrate cooperative communication with NOMA [3]. The impact of relay selection in cooperative NOMA was studied in [4]. Joint NOMA assisted with relaying was investigated in [5] to provide service to multiple cell edge users. In [6], the cooperative NOMA performance was studied with selection of relay and diversity order. By deploying users randomly, the authors in [7] investigated outage behavior and ergodic rate of NOMA. In [8], downlink NOMA in single cell scenario with imperfect SIC is evaluated. In [9], authors derived closed form expression for far end user OP with dedicated relay operating in half-duplex (HD) mode. Authors of [10, 11] depicted that cooperative NOMA with dedicated AF relaying outperforms cooperative NOMA. In [12], cooperative NOMA performance was assessed by OP and sum rate over Nakagami- m fading channel with dedicated DF. With AF and DF relaying and available channel state information (CSI), the order of the decoding of cell edge users was investigated in [13]. In full-duplex (FD) cooperative NOMA under imperfect SIC, the imbalance was analyzed in-phase and quadrature phase [14]. The feasibility of integrating FD in NOMA system model for uplink and downlink was investigated in [15]. They depicted that FD-NOMA offers better performance gain than HD-NOMA and OMA.

In [16], authors derived expressions for outage behavior and ergodic capacity in FD-NOMA in availability of perfect and imperfect SIC. Energy harvesting was investigated in network nodes from surrounding radio frequency by authors in [17–19]. It was observed that EH techniques allow network nodes to extend their operating times. In cooperative communication, HD relay system decreases the spectrum efficiency, whereas diversity gain is degraded in FD relay system with the self-interference. Motivated by the aforementioned works, incremental relaying (IR) is deemed with AF or DF in cooperative FD-NOMA vehicular networks. Relay vehicle uses EH to support far end user in decoding the intended information.

The rest of the paper is organized as follows. Section 2 depicts the system model under consideration with different channel models. The first half of Sect. 3 presents incremental relaying in FD-NOMA for AF and DF protocol, and remaining part of the section is used to present mathematical expression of outage probability of V_1 and V_2 using AF or DF relaying. In Sect. 4, parameters considered for the system under implementation and simulation results are presented for analysis. In Sect. 5, we conclude.

2 System Model

Figure 1 presents the system model of the downlink cooperative vehicular network with NOMA. The roadside unit (RSU) is denoted as S , intended vehicles are denoted as V_1 and V_2 , and relay vehicle is denoted as V_R . The RSU can select the best single relay from the set of decodable relays with the strongest instantaneous channel strength after updated CSI [20].

The relay vehicle is assumed to be operating in FD mode with AF or DF relaying protocol. V_R uses linear energy harvesting to collect the energy from source S and use the energy to forward the information to V_1 and V_2 . V_2 is assumed to be located at the end of the vicinity of RSU and has no direct contact with the source S . Self-interference is caused due to the use of FD in relay vehicle. All the channels in the system are assumed to endure frequency flat fading, independent and non-identical distributed nakagami- m fading with complex channel coefficients.

Channel response between any two communicating devices X and Y varies with time and can be denoted as $h_{xy}(t)$ with an exponential distributive variance σ_{xy}^2

$$\sigma_{xy}^2 = (\lambda/4\pi d_0)^2 (d_0/d_{xy})^2 \quad (1)$$

where

- α Pathloss exponent
- λ Carrier wavelength
- d_0 Reference distance
- d_{xy} Distance between X and Y .

The channel gain between X and Y is represented as $g_{xy} = |h_{xy}(t)|^2$ and the exponentially distributed with an average value $\bar{g}_{xy} = \sigma_{xy}^2$ whose probability density function (PDF) and the cumulative density function (CDF) are given as

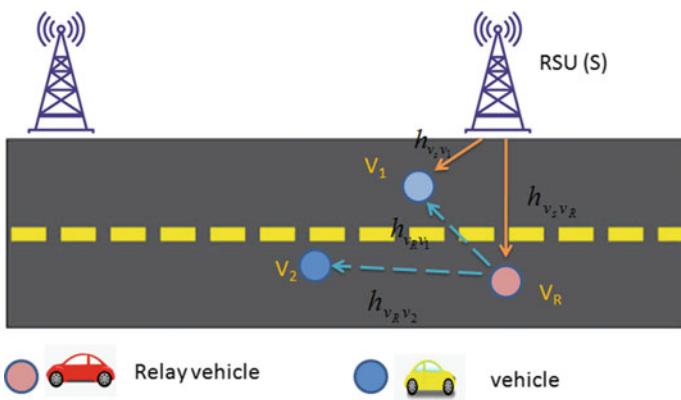


Fig. 1 System model of vehicular networks

$$f_{|h_{xy}|^2}(z) = \frac{1}{\lambda_{xy}} \exp\left(-\frac{z}{\lambda_{xy}}\right) \quad (2)$$

$$f_{|h_{xy}|^2}(z) = 1 - \exp\left(-\frac{z}{\lambda_{xy}}\right) \quad (3)$$

The self-interference (SI) in FD V_R is caused to the reflection paths. The direct and reflection paths are statistically independent, and the gain of the SI channel can be described as a complex Gaussian variable with $|h_{v_{RR}}|$ having Rayleigh distribution.

For all vehicles under communication, additive white Gaussian noise (AWGN) n_k ($k \in R, V_1, V_2$) with same noise distribution $\text{CN}(0, \sigma_{xy}^2)$ is considered. The transmitting power from the source (S) and relay vehicle (V_R) is presumed to be similar, i.e., $P_s = P_R = P$. The SNR of the received signal is given as $\gamma_{xy} = \frac{P|h_{xy}|^2}{\sigma_{xy}^2}$, where SI is exponential distributed with noise variance $\left(\frac{P}{\sigma_{xy}^2}\right)$.

Based on the NOMA principal, roadside unit (S) sends combined signal to V_1 and V_2 by using superposition coding techniques in the n th time slot.

$$x_n[n] = \sqrt{P_s a_1} x_1[n] + \sqrt{P_s a_2} x_2[n] \quad (4)$$

where the intended signals of and power allocation coefficients of vehicles V_1 and V_2 are represented as x_1, x_2 and a_1, a_2 , respectively. They should gratify $a_1 < a_2$ and $a_1 + a_2 = 1$. For vehicle V_i , V_i , $i \in \{1, 2\}$, $E\{|x_i|^2\} = 1$, $E\{\cdot\}$ denote expectation operation.

The received signal at the relay vehicle (V_R) is given as

$$y_R[n] = h_{sv_R}[n]x_s[n] + \sqrt{P_R}h_{v_{RR}}x_2[n - \tau] + n_R[n] \quad (5)$$

$\tau \geq 1$ is an integer representing the time delay due to signal processing in FD mode. For the τ time slots, the V_1 is effected with interference from V_R and the signal transmitted from V_R to V_2 . $h_{v_{RR}}$ is the loop interference channel at V_R , and as the relay cannot perform complete reduction with SIC, the residual self-interference (RSI) could be derived as $h_{v_{RR}}[n] = K_{\text{SIC}}h_{v_{RR}}[n] \neq 0$, where K_{SIC} is attenuation coefficient.

In the energy harvesting at the relay vehicle, some part of the total power received is used for decoding. The received signal at V_R , due to energy harvesting with power division factor α is given as:

$$y_R(\text{EH}) = \sqrt{1 - \alpha}y_R[n], \quad \alpha \in (0, 1) \quad (6)$$

The harvested energy at V_R is given as

$$E_R = \xi(1 - \alpha)\left(P_s|h_{sv_R}|^2 + P_R|h_{v_{RR}}|^2 + \sigma_R^2\right) \quad (7)$$

where ξ is the efficiency of the energy conservation.

The SINR to decode x_2 at V_R by treating x_1 as interference is

$$\gamma_{v_R, x_2} = \frac{\alpha P_s a_2 |h_{sv_R}|^2}{\alpha (P_s a_1 |h_{sv_R}|^2 + P_R |h_{v_R R}|^2 + \sigma_{v_R}^2) + \delta_{v_R}^2} \quad (8)$$

During the signal decoding, the variance of the AWGN is denoted as $\delta_{v_R}^2$. The received signal at V_2 can be articulated as

$$y_{v_2}[n] = \sqrt{P_R} h_{v_R R} x_2[n - \tau] + n_{v_2}[n] \quad (9)$$

where $n_{v_2}[n] \sim \text{CN}(0, \sigma_{v_2}^2)$ is the AWGN at V_2 .

The received signal at V_1 is expressed as

$$y_{v_1}[n] = h_{sv_1} x_s[n] + \sqrt{P_R} h_{v_R v_1} x_2[n - \tau] + n_{v_1}[n] \quad (10)$$

$n_{v_1}[n] \sim \text{CN}(0, \sigma_{v_1}^2)$ is the AWGN at V_1 .

Similarly, with the decoding principal of NOMA, V_1 interprets x_2 by assuming x_1 as interference. The expression for SINR to decode x_2 at V_1 by treating x_1 as interference is given as

$$\gamma_{v_1, x_2} = \frac{P_s a_2 |h_{sv_1}|^2}{P_s a_1 |h_{sv_1}|^2 + P_R |h_{v_R v_1}|^2 + \sigma_{v_1}^2} \quad (11)$$

As V_1 knows $x_2[n - \tau]$ and in spite of the previous decoding it cannot be eliminated. Therefore, $h_{v_R v_1}$ channel is modeled as inter-user interference. After successfully decoding x_2 , V_1 removes x_2 and performs decoding x_1 in the next step. SNR is determined as

$$\gamma_{x_1} = \frac{P_s a_1 |h_{sv_1}|^2}{P_R |h_{v_R v_1}|^2 + \sigma_{v_1}^2} \quad (12)$$

3 Incremental Relaying

The loop interference due to FD in relay vehicle affects the diversity gain of the system. By sending single bit feedback as overhead, the spectral efficiency can be improved with implementation of incremental relaying (IR) protocol. The relay vehicle in IR protocol is involved only when channel gain between source and the

destination vehicle is below threshold. The source broadcasts a pilot signal to vehicles and relay. The vehicle at the far end sends negative acknowledgement with 1 bit to source and relay vehicle if the intended information is not decodable due to channel gain. The achievable rate of V_1 and V_2 with IR is expressed as

$$\begin{aligned}
 R_1 &= \log_2(1 + a_1 \gamma_{sv_1}) \\
 R_2 &= \log_2(1 + \gamma_{v_R v_2}) \\
 &= \log_2\left(1 + \frac{P a_2 |h_{v_R v_2}[n]|^2}{P a_1 |h_{v_R v_2}[n]|^2 + \frac{|h_{v_R v_2}[n]|^2}{|h_{sv_R}[n-\tau]|^2} \sigma^2} + \sigma^2\right) \\
 &= \log_2\left(1 + \frac{a_2 \gamma_{v_R v_2}}{a_1 \gamma_{v_R v_2} + \frac{\gamma_{v_R v_2}}{\gamma_{sv_R}} + 1}\right)
 \end{aligned} \tag{13}$$

3.1 IR with Amplify and Forward

Once the negative acknowledgement is received to the relay vehicle V_R , in the next time slot, AF signal will be forwarded from S to V_1 and V_2 . The received signals at V_1 and V_2 after AF are given as

$$y_{v_1}(\text{AF}) = G h_{v_R v_1} y_R[n-1] + n_{v_1} \tag{14}$$

$$y_{v_2}(\text{AF}) = G h_{v_R v_2} y_R[n-1] + n_{v_2} \tag{15}$$

where $G = \sqrt{P_R/P_s |h_{v_R,s}|^2 + \sigma^2}$ represents the amplification factor of AF relay. The expression for received SINR at V_1 to interpret x_2 with SIC scheme is given as

$$\gamma_{v_1, v_2}(\text{AF}) = \frac{a_2 \rho_2 |h_{v_R v_2}|^2 |h_{v_R v_1}|^2}{a_1 \rho_s |h_{v_R v_2}|^2 |h_{v_R v_1}|^2 + |h_{v_R v_1}|^2 + |h_{v_R s}|^2 + (1/\rho_s)} \tag{16}$$

V_1 interprets its intended information with following SINR after successfully decoding and subtracting the information of V_2

$$\gamma_{v_1}(\text{AF}) = \frac{a_1 \rho_s |h_{v_R v_s}|^2 |h_{v_R v_1}|^2}{|h_{v_R v_1}|^2 + |h_{v_R s}|^2 + (1/\rho_s)} \tag{17}$$

Similarly, the expression for received SINR at V_2 to interpret its own information can be given as

$$\gamma_{v_2}(\text{AF}) = \frac{a_2 \rho_s |h_{v_R v_s}|^2 |h_{v_R v_2}|^2}{a_1 \rho_s |h_{v_R s}|^2 |h_{v_R v_2}|^2 + |h_{v_R v_2}|^2 + |h_{v_R s}|^2 + (1/\rho_s)} \quad (18)$$

3.2 IR with Decode and Forward

The relay is assumed to be capable of decoding two vehicle information

$$y_{v_1}(\text{DF}) = \sqrt{a_1 P_R} h_{v_R v_1} x_1 + \sqrt{a_2 P_R} h_{v_R v_1} x_2 + n_1 \quad (19)$$

$$y_{v_2}(\text{DF}) = \sqrt{a_2 P_R} h_{v_R v_2} x_1 + \sqrt{a_2 P_R} h_{v_R v_2} x_2 + n_2 \quad (20)$$

By means of SIC, the received SINR at V_1 to interpret x_2 is expressed by

$$\gamma_{v_1 v_2}(\text{DF}) = \frac{a_2 \rho_s |h_{v_R v_1}|^2}{a_1 \rho_s |h_{v_R v_1}|^2 + 1} \quad (21)$$

SINR required at V_1 to interpret its own information is expressed as

$$\gamma_{v_1}(\text{DF}) = a_1 \rho_s |h_{v_R v_1}|^2 \quad (22)$$

Similarly, the SINR required at V_2 to interpret own information is expressed by

$$\gamma_{v_2}(\text{DF}) = \frac{a_2 \rho_s |h_{v_R v_2}|^2}{a_1 \rho_s |h_{v_R v_2}|^2 + 1} \quad (23)$$

3.3 Outrage Probability with AF

OP of any channel can be expressed as the probability of instantaneous SINR of that channel is less than the threshold value. If t_1 and t_2 are threshold SINR to achieve target rate at r_1 and r_2 :

Then $t_1 = 2^{r_1} - 1$, and $t_2 = 2^{r_2} - 1$.

Outage probability of V_1 and V_2 with AF relaying is expressed as

$$P_{v_1}(\text{AF}) = \Pr(\gamma_{v_2}(\text{AF}) \geq t_2) [1 - \Pr(\gamma_{v_1 v_2}(\text{AF}) \geq t_2, \gamma_{v_1} \geq t_1)] \quad (24)$$

$$P_{v_2}(\text{AF}) = \Pr(\gamma_{v_2}(\text{AF}) < t_2, \gamma_2(\text{AF}) < t'_2) \quad (25)$$

where $t'_1 = 2^{2R_1} - 1$, $t'_2 = 2^{2R_2} - 1$ are the decoding thresholds with R_1 and R_2 target rates of V_1 and V_2 .

3.4 Outage Probability with DF

Outage probability of V_1 and V_2 with DF is expressed as

$$P_{v_2}(\text{DF}) = \Pr(\gamma_{v_2}(\text{DF}) < t_2, \gamma_{v_2}(\text{DF}) < t'_2) \quad (26)$$

$$P_{v_1}(\text{DF}) = \Pr(\gamma_{v_2}(\text{DF}) \geq t_2 [1 - \Pr(\gamma_{v_1 v_2} \geq t_2, \gamma_{v_1}(\text{DF}) \geq t_1)]) + \Pr(\gamma_{v_2} < t_2) \quad (27)$$

The closed form expression for the outage probabilities is referred from [21].

4 Numerical Results

The results present outage behavior for average SNR with respect to the power allocation factors while relay vehicle operating in FD-AF or FD-DF. The roadside unit (S) needs to know the channel coefficients $|h_{sv_R}|$ and $|h_{sv_1}|$ to allocate power factor for x_1 and x_2 information. Since the RSU and vehicle positions are considered to be fixed for evaluation, the power allocation coefficients are also considered as fixed, i.e., $a_1 = 0.3$ and $a_2 = 0.7$ or $a_1 = 0.2$ and $a_2 = 0.8$. Additionally, the target rates of V_1 and V_2 are $R_1 = 1.5$ bits/channel and $R_2 = 0.5$ bits/channel, respectively. If the d_{xy} be the distance between any two points X and Y , then $\lambda_{xy} = d_{xy}^{-\beta}$ for free space path loss transmission. The system parameters are set as $\beta = 3$, $d_{v_R s} = d_{v_R v_1} = d_{v_R v_2} = 63$ m, $d_0 = 40$ m, $d_{v_1 s} = 50$ m, $d_{v_2 s} = 85$ m, $\alpha = 0.6$ and $\xi = 0.8$.

Figure 2 depicts that, if a_2/a_1 is large, the outage behavior for the proposed system is better. However, when a_1 is considered too low, the decoding ability will be reduced at nearby vehicle V_1 . It is observed that at low SNR regime, the OP of the x_1 is lower than x_2 . Whereas at higher SNR values the outage of the x_2 decreases hastily than x_1 due to the distance.

Figure 3 depicts the OPs of x_1 and x_2 for different power allocation coefficients a_2/a_1 and power division ration α . Higher the α value reduces the outage probability of x_1 , and when α increases over a specific value, the outage probability of x_2 increases, because of reduction in energy harvesting at relay vehicle (VR).

Figure 4 presents the OP of the two vehicles with different target rates. If the achievable target rate R_1 increases, it is observed that OP of V_1 decreases, because the threshold for decoding increases with R_1 . Whereas V_1 can successfully decode the x_2 because it is not effected by R_2 adjustments. Similarly, V_2 outage probability decreases with increase in R_2 and does not depend on the R_1 adjustments. It is further

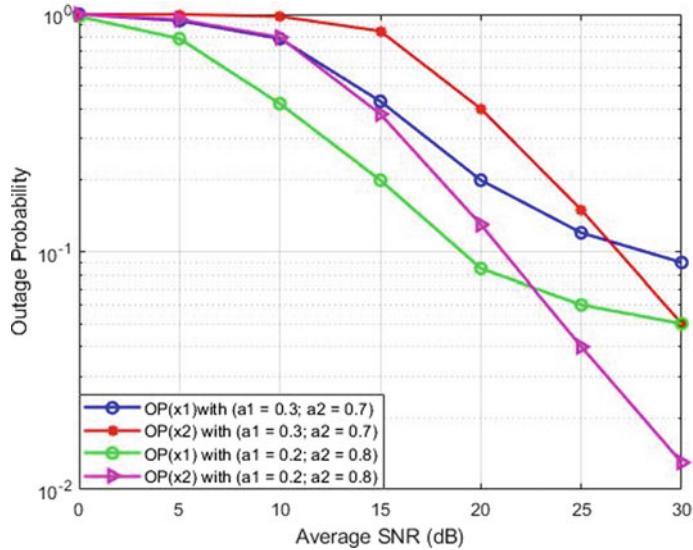


Fig. 2 Outage probability at V_1 and V_2 for average SNR for power allocation coefficients

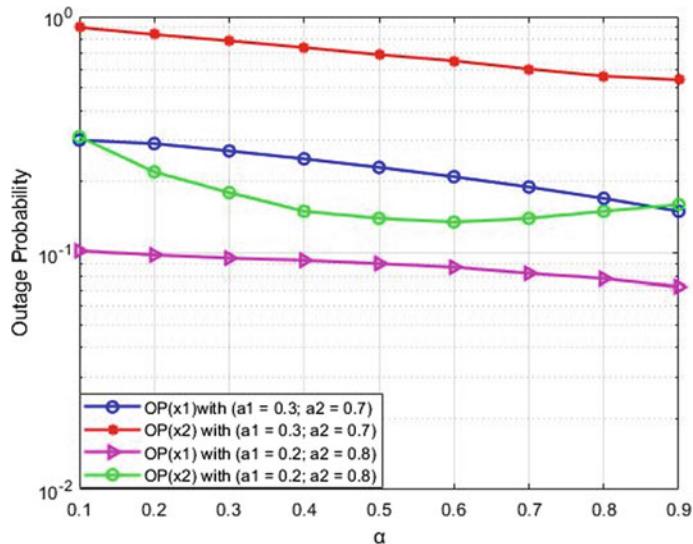


Fig. 3 Outage probability of V_1 and V_2 for different power allocation coefficients and α

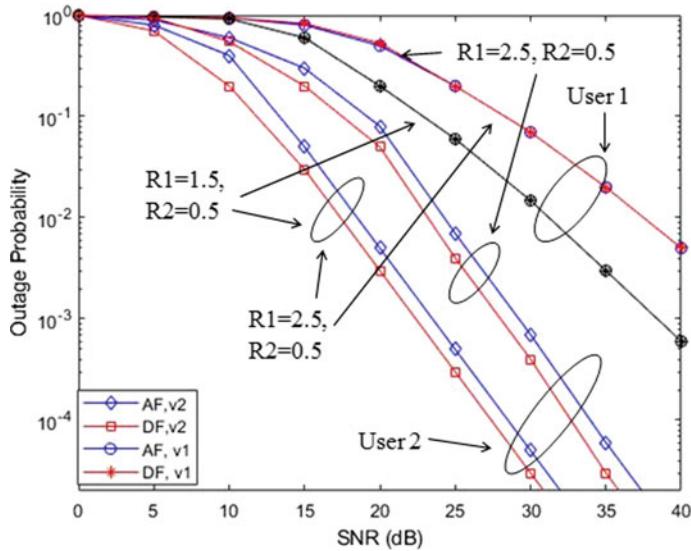


Fig. 4 Outage probability analysis for different target values considering power allocation coefficients as $a_1 = 0.2$, $a_2 = 0.8$ and $\alpha = 3$

observed that outage behavior of V_2 is better in DF rather than AF relaying protocol, because in AF, the x_1 is also amplified, is treated as noise signal and affects the decoding of original x_2 information at V_2 .

Figure 5 shows the plotting of outage probabilities of V_1 and V_2 for different power allocation coefficients. With the decrease in the a_1 , outage behavior of V_1 is noticed to be reducing, whereas V_2 outage behavior is improving. With increase in transmit power of the vehicle, the decodable performance increases. In NOMA, it is obvious that the far end vehicle has to be allocated more transmission power for better performance.

5 Conclusion

In this paper, we evaluated downlink full-duplex cooperative NOMA in vehicular networks, where the roadside unit has no direct contact with the far end vehicle. The relay vehicle is assumed to be operating in either AF or DF protocol. To reduce the effect of RSI in the FD relay, IR is implemented. Mathematical expressions are derived for appropriate power division ratio as function of average SNR. Simulation results show that as power division ratio increases the OP decreases and probability of decoding the information decreases. It is further perceived that, in incremental relaying (IR) DF has better performance than AF in FD cooperative NOMA system.

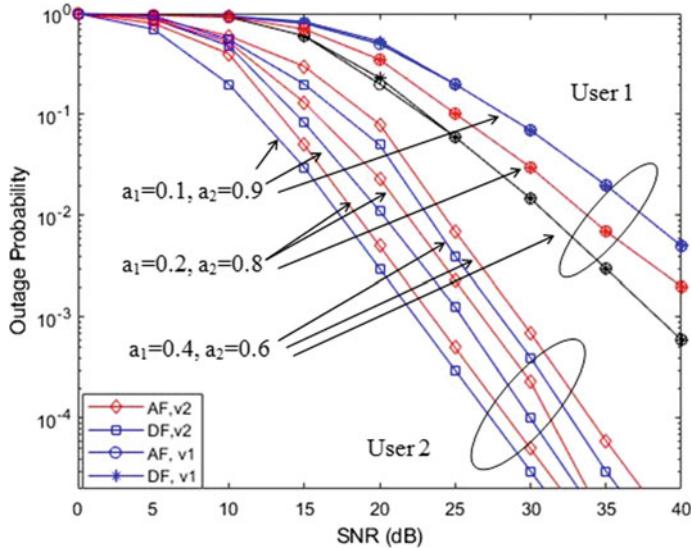


Fig. 5 Outage probability of V_1 and V_2 for different power allocation coefficients when achievable target rates are $R_1 = 1.5$ bits/channel, $R_2 = 0.5$ bits/channel and $\alpha = 3$

References

1. Yue X, Qin Z et al (2018) A unified framework for non-orthogonal multiple access. *IEEE Trans Comm* 66(11):5346–5359
2. Liu Y, Qin Z et al (2017) Non-orthogonal multiple access for 5G and beyond. *Proc IEEE* 105(12):2347–2381
3. Sravani P, Sharma A et al (2020) Outage analysis of half-duplex decode and forward relaying with cooperative NOMA in vehicular networks. *Mater Today Proc* (in press). <https://doi.org/10.1016/j.matpr.2021.01.218>
4. Ding Z, Dai H et al (2016) Relay selection for cooperative NOMA. *IEEE Wirel Comm Lett* 5(4):416–419
5. Liu X et al (2016) Outage probability and capacity analysis of the collaborative NOMA assisted relaying system in 5G. In: *IEEE-CIC international conference on communications (ICCC)*. China
6. Li X, Liu M et al (2019) Full duplex cooperative NOMA relaying system with I/Q imbalance and imperfect SIC. *IEEE Wirel Comm Lett* 9(1):17–20
7. Ding Z, Yang Z et al (2014) On the performance of non-orthogonal multiple access in 5G system with randomly deployed users. *IEEE Sig Proc Lett* 21(12):1501–1505
8. Yang Z, Ding Z et al, On the performance of non-orthogonal multiple access systems with part channel information. *IEEE Trans Comm* 64(2):654–667
9. Kim JB, Lee IH (2015) Non-orthogonal multiple access in coordinated direct and relay transmission. *IEEE Comm Lett* 19(11):2037–2040
10. Liang X, Wu Y et al (2017) Outage performance for cooperative NOMA transmission with an AF relay. *IEEE Comm Lett* 21(11):2428–2431
11. Yue X, Liu Y et al (2017) Performance analysis of NOMA with fixed gain relaying over Nakagamim fading channels. *IEEE Access* 5:5445–5454
12. Chu TMC et al (2018) Performance of non-orthogonal multiple access system with full duplex relaying. *IEEE Comm Lett* 22(11):2048–2087

13. Wan D, Wen M (2018) Cooperative NOMA system with partial channel state information over Nakagami-m fading channels. *IEEE Trans Comm* 66(3):947–958
14. Wang X, Jia M et al (2019) Exploiting full-duplex two-way relay cooperative non-orthogonal multiple access. *IEEE Trans Comm* 76(4):2716–2729
15. Ding Z, Fan P et al (2018) On the coexistence between full duplex and NOMA. *IEEE Wirel Comm Lett* 7:692–695
16. Ahiadormey RK, Anokye P et al (2020) Two-way relaying with imperfect successive interference cancellation in power line communication. *IEEE Open J Comm Soc* 1:1872–1885
17. Nguyen VD, Duong TQ et al (2017) Spectral and energy efficiencies in Full-Duplex wireless information and power Transfer. *IEEE Trans. Comm* 65:2220–2233
18. Hoang TM, Nguyen BG et al (2020) Outage analysis of RF energy harvesting cooperative communication system over Nakagami-m fading channels with Integer and non-integer. *IEEE Trans Veh Tech* 69:2785–2801
19. Zlatanov N, Schober et al (2017) Asymptotically optimal power allocation for energy harvesting communication networks. *IEEE Trans Veh Tech* 66:7286–7301
20. Zhang T. et al (2013) Opportunistic DF-AF selection relaying with optimal relay selection in Nakagami-m fading environment. In: IEEE international conference on communication, China
21. Van Toan H et al (2020) Outage probability and ergodic capacity of a two-user NOMA relaying system with an energy harvesting full duplex relay and its interference at the near user. *Sensor* (2020). <https://doi.org/10.3390/s20226472>

Lower Order Harmonic Torque Reduction in v/f Controlled Induction Motor Drive



Shivam Yadav, Anshul Kumar Mishra, Jayant Mani Tripathi, and Aseem Chandel

Abstract The high amplitude of lower order harmonic torque can lead to several losses, damage of drive, shaft failure, etc. Therefore, it is the foremost concern in modern high speed, high-power and motor drives. Presently, for its reduction, various classical methods (SHE, STPWM) are popularly being used. This paper proposes frequency domain (FD)-based approach in IM drive for reduction of harmonic torque of lower order. In this approach, the optimum switching angles are proposed and evaluated, for minimization or elimination of harmonic torque (lower order) for a particular case, i.e. twelve order harmonic torque. The approach being independent of load conditions and machine parameters can be easily extended for higher order harmonic torque and pulse number. The relative performance of the method is extensively being compared with selective harmonics elimination (SHE) and sine-triangles (ST) PWM through studies, and findings are supported as well as verified by the results of simulation in 3.73 KW IM drive. The frequency domain (FD) approach outperforms SHEPWM and STPWM in terms of harmonic torque magnitude, pulsating torque and torque THD both at no-load and loading conditions.

Keywords IM drive · VSI optimal PWM · Optimal PWM · FD-optimal · Torque ripples minimization

1 Introduction

Electrical drives are getting popularity mainly due to scarcity of fossil fuels and emphasis on clean environment [1]. High-power level IM drives are used in numerous applications like electrical vehicles, combustion engines, various sectors of energy

S. Yadav (✉) · A. K. Mishra · J. M. Tripathi · A. Chandel
Electrical Department, Rajkiya Engineering College, Mainpuri, India

A. K. Mishra
e-mail: anshulmishra@ieee.org

A. Chandel
e-mail: aseemchandel@ieee.org

and oil, etc. [2, 3]. In high-power drives system, switching efficiency is utmost important, to achieve this switching frequency is kept lowered to avoid switching losses. At high-power levels, in IM drive due to high energy loss in switching transition of semiconductor devices, the pulse number (P) is considered to be low [4]. Pulse number being the number of pulses in the output voltage of DC within one time period of the source voltage of AC.

In VSI operated high-power IM drive, switching loss is the foremost concern leading to low switching frequency of PWM inverters. PWM inverters are mainly used to control the output voltage (fundamental) due and to reduce harmonic voltage with their harmful effects. Pulsating torque is produced in modern AC drives owing to non-sinusoidal nature of PWM inverters [4–7]. For low P (pulse number), IM drive has voltage harmonics (line-line) of order 5th, 7th, 11th, 13th, 17th ... [6]. These voltage harmonics lead to harmonics in current and flux, their interaction results in harmonic torque of lower order 6th, 12th, 18th, 24th ... [3, 6, 8]. Higher amplitude of lower order harmonics in torque causes torsional oscillations resulting in several losses, damages to drive and can even lead to shaft failure [3, 7, 9].

Research community in the recent years has shown continuous interest in reducing harmonic torque of lower order in high-power drives for this, various methods, techniques and control algorithms are being proposed. Modulation approach like PWM technique and its variants including sine-triangle pulse width modulation (STPWM) is well-known traditional approach and is most widely used. In PWM, width of the pulses is designed to get desired fundamental voltage and harmonics shifted towards higher frequency bands. It is very effective technique only if switching losses are not concerned (switching only up to few KHz acceptable). Nowadays, another technique SHE selective harmonic elimination is gaining popularity due to low switching frequency ratio with acceptable performance. In this approach, switching angles are chosen to get the desired fundamental output and particularly chosen harmonics are suppressed or eliminated. This technique eliminates selected harmonics with some reduced losses.

This paper proposes frequency domain (FD)-based scheme for evaluation and minimization of harmonic torque which is independent of both machines and load parameters. Further, the approach can be easily extended both for torque harmonics (higher order) and pulse number (N). The minimization of FD method has been extensively compared with selective harmonics elimination (SHE) PWM, sine-triangles (ST) PWM and is verified by simulation results.

2 Harmonic Torque Expression

In VSI operated PWM-based IM drive having pulse number (P) = $(2N + 1)$, here, N being the switching angle per quarter per cycle so $(N - 1)$ voltage harmonics are eliminated [6], i.e. for $N = 3$, the 5th and 7th both harmonics voltages are eliminated. Pairs of harmonic voltages lead to harmonic torque. For example, harmonic voltage pair (11th, 13th) leads to harmonic torque (12th). Similarly, harmonic voltages pair

(17th, 19th) leads to harmonic torque (18th). Torque harmonics of specific order $6n$ (n is positive integer) can be eliminated by classical method if harmonic current amplitude of order $(6n - 1)$ and $(6n + 1)$ are equated [3, 9]. For two switching angle per quarter ($N = 2$), the harmonic torque varies as 6th, 12th, 18th Similarly, for three switching angle per quarter ($N = 3$), the harmonic torque varies as 12th, 18th

Considering a particular case of VSI operated PWM-based induction motor drive having three switching angle per quarter ($N = 3$). The strategy for PWM waveform for three switching angle per quarter (α_1, α_2 and α_3) is illustrated in Fig. 1.

Therefore, $N = 3$ results in switching seven times the fundamental frequency with pole voltage shown in Fig. 2, the corresponding harmonics torque (lowest order) is 12th for this, FD (frequency domain)-based approach independent of machine and load parameter is proposed.

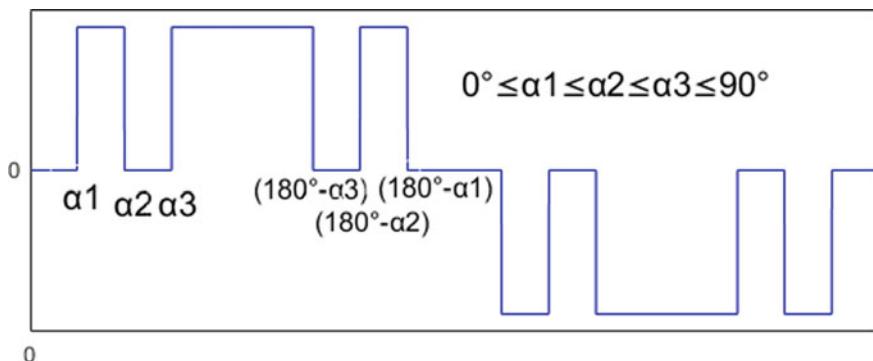


Fig. 1 SHEPWM strategy (quarter wave symmetrical 3°)

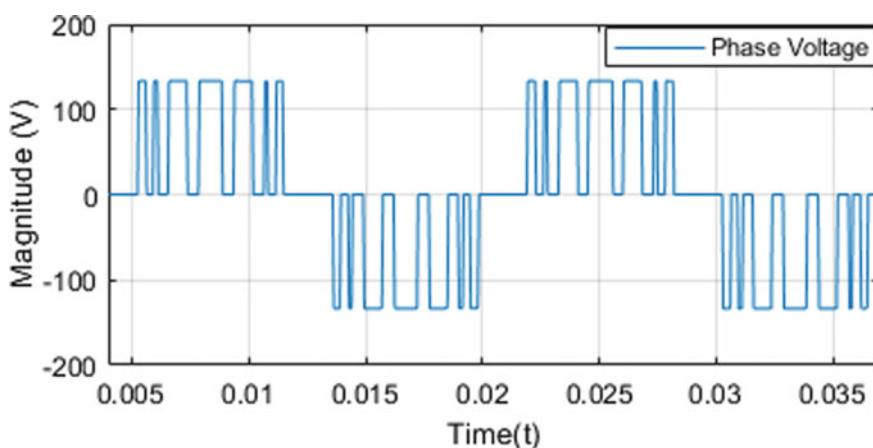


Fig. 2 VSI operated phase voltage waveform with P (pulse number) = 7 or N = 3

The fundamental amplitude and odd harmonic voltages given by (1)

$$E_n = \left(\frac{2E_{dc}}{n\Pi} \right) [1 - 2 \cos(n\alpha_1) + 2 \cos(n\alpha_2) - 2 \cos(n\alpha_3)] \quad (1)$$

E_{dc} is the voltage (DC bus), α_1 , α_2 and α_3 are switching angles, and n is the order of harmonics.

Harmonic torques results owing to interaction of flux (fundamental) with current (harmonic) and that of current (fundamental) with flux (harmonic) [7, 10]. For example, the 12th harmonic torque is resultant of 11th and 13th order fluxes and harmonic current interaction with fundamental current and flux. Similarly, 18th harmonic torque produced by 17th and 19th order fluxes and harmonic current. Torque expression for ζ_{12} and ζ_{18} is expressed below [10, 11].

$$\begin{aligned} \zeta_{12} \sin(12\omega_s t + \delta_{12}) &= K_2 [\phi_{1m}(I_{r11} - I_{r13}) \sin(12\omega_s t) \\ &\quad + I_{r1}(\phi_{11m} + \phi_{13m}) \cos(12\omega_s t)] \end{aligned} \quad (2a)$$

$$\begin{aligned} \zeta_{18} \sin(18\omega_s t + \delta_{18}) &= K_1 [\phi_{1m}(I_{r17} - I_{r19}) \sin(18\omega_s t) \\ &\quad + I_{r1}(\phi_{17m} + \phi_{19m}) \cos(18\omega_s t)] \end{aligned} \quad (2b)$$

Here, ϕ_{1m} fundamental harmonic air gap flux of order (11th,13th,17th,19th) are $\phi_1, \phi_{11}, \phi_{13}, \phi_{17}, \phi_{19}$. Similarly, components of corresponding rotor currents are $I_{r1}, I_{r11}, I_{r13}, I_{r17}, I_{r19}$. $K1$ and $K2$ are machine parameters constant. ω_s is the frequency in rad/s. δ_{12} and δ_{18} are phase angle for 12th and 18th harmonic torque.

In harmonic torque expression (Eqs. (2a) and (2b)), the second term is neglected [7, 10, 11]. The expression leads to approximation.

Equations

$$\zeta_{12} \approx K_1 [\phi_1(I_{r11} - I_{r13})] \quad (3a)$$

$$\zeta_{18} \approx K_2 [\phi_1(I_{r17} - I_{r19})] \quad (3b)$$

3 Classical Solution of Harmonic Torque

A method is approximated to eliminate harmonic torque ζ_{12} and ζ_{18} in VSI controlled IM drive having three switching angle per quarter.

3.1 Approximate Solution Evaluation

For the expression in Eqs. (3a) and (3b), the harmonics torque (ζ_{12}) and (ζ_{18}) eliminated if following conditions are satisfied

$$I_{r11} = I_{r13} \quad (4a)$$

$$I_{r17} = I_{r19} \quad (4b)$$

Above condition stated in Eqs. (4a) and (4b) are satisfied if corresponding voltage harmonics holds relation in (5a) and (5b).

$$(E_5/11) = (E_7/13) \quad (5a)$$

$$(E_{11}/17) = (E_{13}/19) \quad (5b)$$

where E_{11} , E_{13} , E_{17} and E_{19} are harmonic voltage of order (5th, 7th, 11th, 13th).

Equations (5a) and (5b) can be re-written in terms of alpha (α_1 , α_2 , α_3) as specified below and shown in Fig. 3.

$$\begin{aligned} & \left(\frac{1}{121} \right) (1 - 2 \cos(11\alpha_1) + 2 \cos(11\alpha_2) - 2 \cos(11\alpha_3)) \\ &= \left(\frac{1}{169} \right) (1 - 2 \cos(13\alpha_1) + 2 \cos(13\alpha_2) - 2 \cos(13\alpha_3)) \\ & \left(\frac{1}{289} \right) (1 - 2 \cos(17\alpha_1) + 2 \cos(17\alpha_2) - 2 \cos(17\alpha_3)) \end{aligned}$$

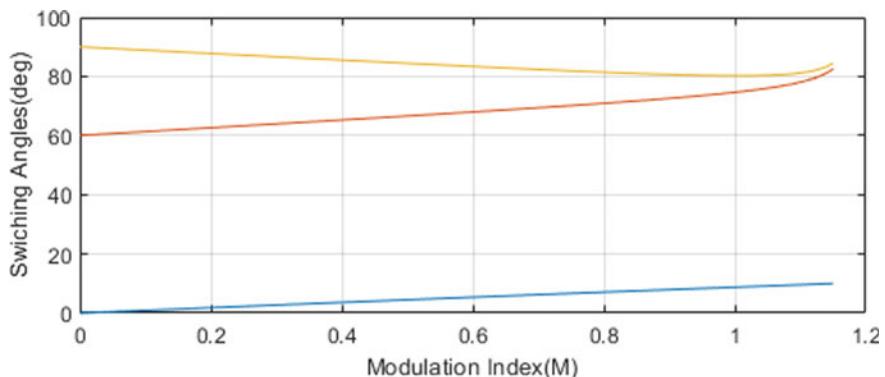


Fig. 3 Three switching angles per quarter corresponding to modulation index (M)

$$= \left(\frac{1}{361} \right) (1 - 2 \cos(19\alpha_1) + 2 \cos(19\alpha_2) - 2 \cos(19\alpha_3))$$

The solution of Eqs. (5a) and (5b) can have multiple solution and no solution for different range of modulation index (M). M is the normalized value ($\frac{2E_{dc}}{n}$) of fundamental voltage ($E1$) in square wave operation.

4 Harmonic Torque Minimization

In FD-based approach, harmonic torque (ζ_{12}) is minimized in this section with three switching angle per quarter.

4.1 Frequency Domain-Based Minimization

Harmonic torque (ζ) magnitude minimization needs minimizing Eq. (2a) which in turn reduces to minimization of second term of Eq. (2a) as the first term is eliminated, i.e. minimization of ($\Psi_5 + \Psi_7$). This is realized by minimizing a machine independent parameter U defined as follows.

$$U_n = \left\| \left(\frac{V_{n-1}}{n-1} + \frac{V_{n+1}}{n+1} \right) \right\| \quad (6)$$

Minimization of ζ_{12}

In frequency domain (FD), ζ_{12} minimization may be specified as by minimizing U (Eq. 7) with constraints specified in Eqs. (8)–(10).

Minimize

$$U_{12} = \left\| \left(\frac{E_{11}}{11} + \frac{E_{13}}{13} \right) \right\| \quad (7)$$

$$(E_{11}/11) = (E_{13}/13) \quad (8)$$

$$M = 1 - 2 \cos(\alpha_1) + 2 \cos(\alpha_2) - 2 \cos(\alpha_3) \quad (9)$$

$$0^\circ \leq \alpha_1 \leq \alpha_2 \leq \alpha_3 \leq 90^\circ \quad (10)$$

With the constraints, the solution space limits to finite set of points. The optimal solution for minimization of ζ_{12} is obtained by choosing lowest value of U_{12} . As the stated minimization procedure, Eqs. (7)–(10) have only three switching angles

per quarter of pole voltage waveform. Hence, stated optimization is independent of load conditions and machine parameters. As it is simple procedure, it can easily be extended for higher order torque harmonics discussed in next section.

Minimization of ζ_{18}

The extension of above scheme for ζ_{18} minimization is presented in this section to validate the proposed procedure. Similarly, U_{18} minimization with constraints is specified below.

Minimize

$$U_{18} = \left\| \left(\frac{E_{17}}{17} + \frac{E_{19}}{19} \right) \right\| \quad (11)$$

$$(E_{17}/17) = (E_{19}/19) \quad (12)$$

$$M = 1 - 2 \cos(\alpha_1) + 2 \cos(\alpha_2) - 2 \cos(\alpha_3) \quad (13)$$

$$0^\circ \leq \alpha_1 \leq \alpha_2 \leq \alpha_3 \leq 90^\circ \quad (14)$$

Optimal solution is lowest value of U_{18} at given M as it is very much effective when four switching angle per quarter is considered.

Hence, this procedure can be extended for both higher torque harmonics and for higher pulse number (P) as well. Pulsating torque minimization through frequency domain (FD) is quick and simple procedure used for optimization.

5 Simulation Results (and Discussion)

Synchronous STPWMs, SHEPWMs and FD-optimal PWM are simulated in VSI operated 3.73 KW IM drive as open loop with constant V/Hz controlled, switching seven times the fundamental frequency ($P = 7$) with parameters shown in Table 1.

The results of simulation carried out at different modulation index (M) = 0.8 and (M) = 0.6, frequency (f) = 60 Hz under both no-load and loading conditions (52.5% Load and 62.4% Load) are arranged in below sections.

Line to Line Voltage and Current

Under no-load conditions V_{ry} (voltage line-line) and I_r (line current) at $M = 0.8$ and $f = 60$ Hz simulated waveforms are shown in Fig. 4 and Voltage harmonic spectrum is shown in Fig. 5. The proposed FD-optimal PWM at no-load leads to lower voltage THD than SHEPWMs and STPWMs.

Table 1 Parameters of induction motor

Specifications of induction motor (IM)	
Stator resistances (R_s)	0.09961 Ω
Rotor resistances (R_r)	0.05837 Ω
Stator inductance (L_s)	0.000867 H
Rotor inductance (L_r)	0.000867 H
Voltage (line-line)	460 V
Mutual inductance (L_m)	0.03039
Inertia (J)	0.4 (kg m ²)
Friction factor (F)	0.202187 (N m s)
Frequency (f)	60
Poles (p)	2

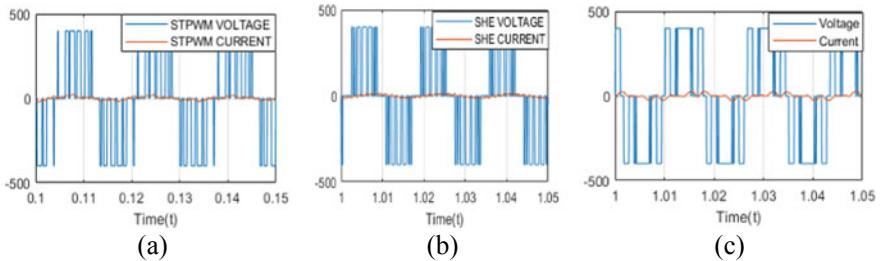


Fig. 4 Line voltage and line current waveforms ($P = 7$) at $M = 0.8$, $f = 60$ Hz, under no loading conditions **a** sine-triangles PWMs, **b** SHEPWMs, **c** FD-based

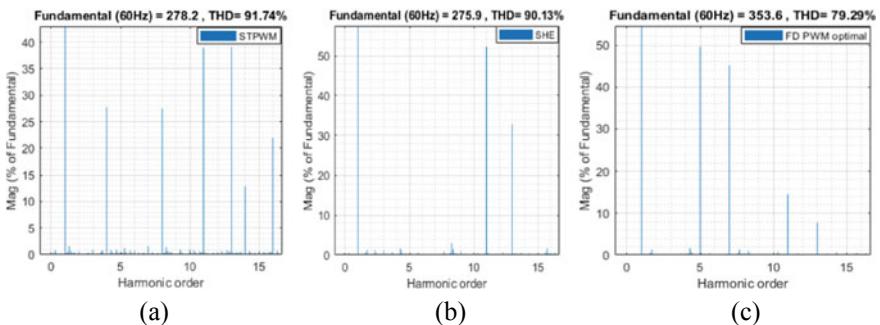


Fig. 5 Voltage THD ($P = 7$) at $M = 0.8$, $f = 60$ Hz, under no loading conditions **a** sine-triangles PWMs, **b** SHEPWMs, **c** FD-based

Harmonic Spectrum and Torque Ripple

Torque ripple and harmonic spectrum under no loading conditions at $M = 0.6$ and $f = 60$ Hz are simulated and shown in Figs. 6 and 7. The proposed FD-optimal PWM at No-load leads to lower torque THD and torque ripple than SHE and STPWMs.

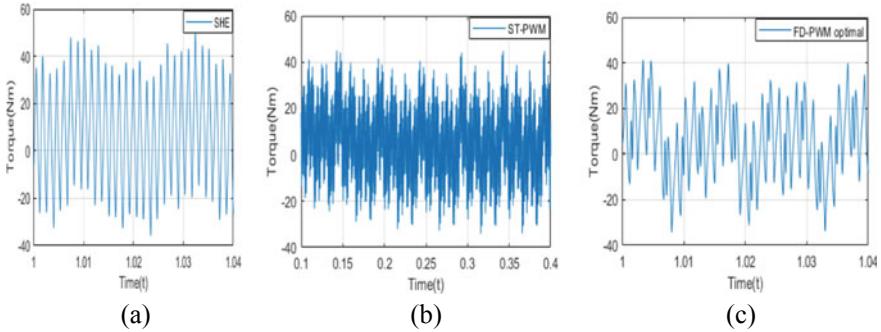


Fig. 6 Torque ripple ($P = 7$) at $M = 0.6, f = 60$ Hz, under no loading conditions **a** sine-triangles PWMS, **b** SHEPWMS, **c** FD-based

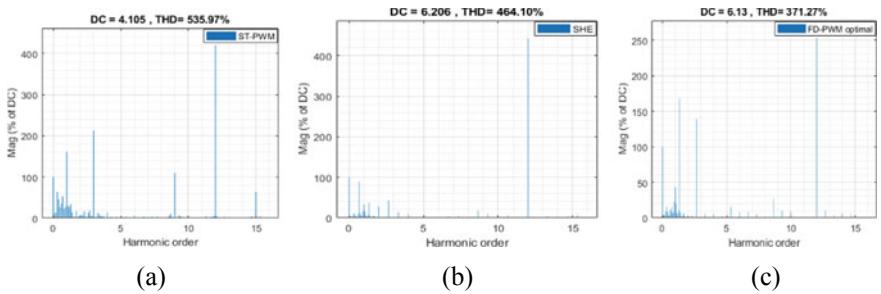


Fig. 7 Torque THD ($P = 7$) at $M = 0.6, f = 60$ Hz, under no loading conditions **a** sine-triangles PWMS, **b** SHEPWMS, **c** FD-based

Steady state peak to peak torque ripple (no-load) at $M = 0.6, f = 60$ Hz, $P = 7$ is calculated for STPWMs, SHE and FD-optimal PWM and revealed in Table 2.

Evaluation of Harmonic Torque

The harmonic torque (ζ_{12}) magnitude and THD for the proposed optimal (FD) PWMS, SHEPWMS and STPWM are compared through simulations at different modulation index and are arranged in Tables 3 and 4. It reveals that FD PWM optimal has lower harmonic torque (ζ_{12}) than STPWM and SHE.

Table 2 Peak to peak torque ripples at $M = 0.6, f = 60$ Hz

Peak to peak torque ripples at $M = 0.6, f = 60$ Hz	
STPWM	83.9878
SHE	78.9898
FD PWM optimal	72.4193

Table 3 Magnitude of harmonic torque (ζ_{12}) of STPWMs, SHEPWMs and FD PWM optimal at different modulation index (M) in range 0.2–0.9 in steps of 0.1

Modulation index (M)	Magnitude of harmonic torque (ζ_{12}) in pu		
	SHEPWM	FD PWM optimal	STPWM
0.2	66.67	34.72	94.87
0.3	153.7	24.95	151.8
0.4	251.7	112	244
0.5	352.2	197	356.2
0.6	443.8	252.2	419.8
0.7	501.4	260.5	455.7
0.8	518.9	75.84	465.4
0.9	479	237.5	415.2

Table 4 THD variation of STPWMs, SHEPWMs and FD PWMs optimal at different modulation index (M) in range 0.3–0.9 in steps of 0.1

Modulation index (M)	Total harmonic distortion (THD) in %		
	SHEPWM	FD PWM optimal	STPWM
0.2	110.60	133.58	131.04
0.3	172.35	168.16	186.03
0.4	279.47	179.58	283
0.5	386.72	276.3	424.04
0.6	465.90	371.27	535.97
0.7	525	350	613.30
0.8	561.09	508.01	728.61
0.9	544.81	430.3	862.4

Load Performance Evaluation

Simulations of IM drive under diverse loading conditions (62.4 and 52.57%) are performed at $M = 0.5, f = 60$ Hz, $P = 7$ and arranged below. The torque THD and torque ripple for 52.57% loading corresponds to Figs. 8 and 10 and for 62.4% loading corresponds to Figs. 9 and 11. The loading conditions clearly demonstrate that FD-optimal PWM has lower THD and torque ripple compared to other two PWM techniques (STPWM and SHE).

6 Conclusion

This paper deals with VSI operated high-power IM drive switching at seven times fundamental frequency. The optimum FD-based approach is proposed for twelfth order harmonic torque minimization using load and machine independent parameter

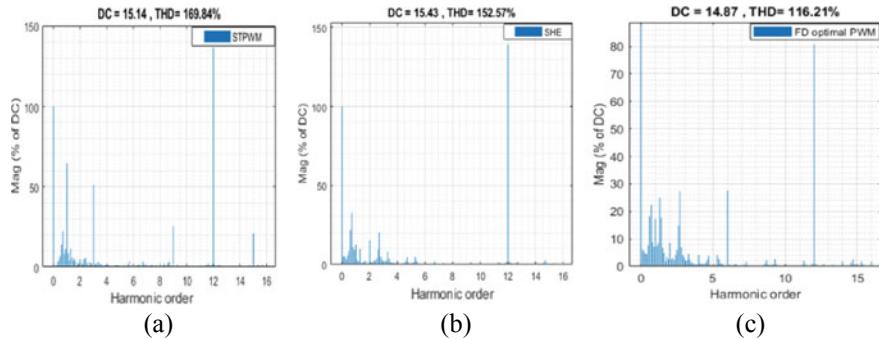


Fig. 8 Torque THD ($P = 7$) at $M = 0.5, f = 60$ Hz, under 52.57% loading condition **a** sine-triangles PWMS, **b** SHEPWMS, **c** FD-based

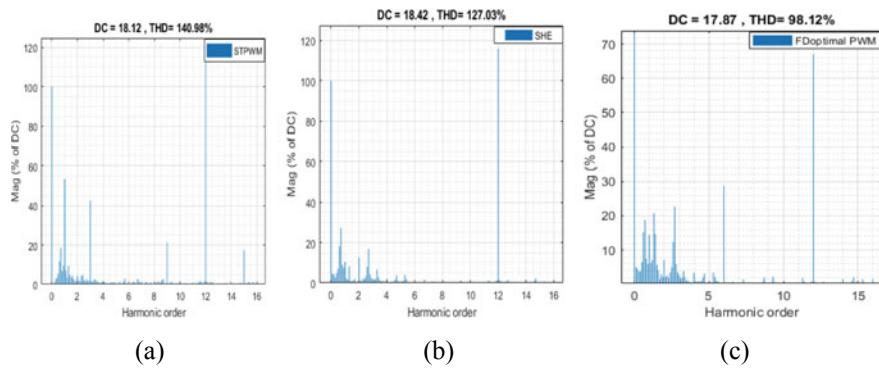


Fig. 9 Torque THD ($P = 7$) at $M = 0.5, f = 60$ Hz, under 62.4% loading conditions **a** sine-triangles PWMS, **b** SHEPWMS, **c** FD-based

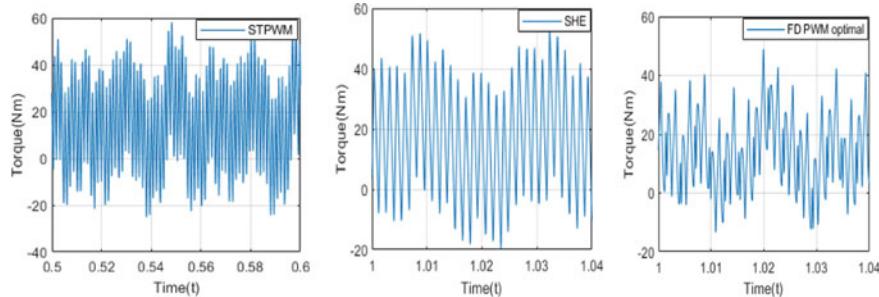


Fig. 10 Torque ripples ($P = 7$) at $M = 0.5, f = 60$ Hz, under 52.57% loading condition **a** sine-triangles PWMS, **b** SHEPWMS, **c** FD-based

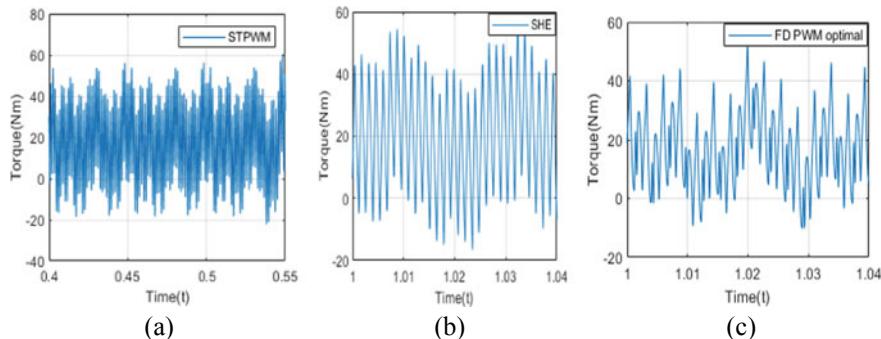


Fig. 11 Torque ripples ($P = 7$) at $M = 0.5$, $f = 60$ Hz, under 62.4% loading condition **a** sine-triangles PWMs, **b** SHEPWMs, **c** FD-based

having three switching angle per quarter. However, the torque harmonics can be lowered if the proposed condition, i.e. sum of harmonic voltages is minimized with constraints as equal amplitude of harmonic current. The proposed procedure can be easily extended both for higher pulse number (P) and higher order harmonic torque.

In this paper, VSI operated IM drive is simulated circuit and its output waveforms are analyzed and compared with selective harmonic elimination (SHE) and sine-triangle pulse width modulation (ST-PWM) in detail. The magnitude of harmonic torque, torque THD and steady state torque ripple is compared with different PWM techniques (FD-optimal, SHE, STPWM) both at no-load and at different loading conditions (52.57% loading, 62.4% loading). The comparative performance is verified, and findings are confirmed by simulation results. The proposed method, i.e. FD-based approach outperforms the other two techniques (SHE and STPWM) in terms of harmonics torque magnitude, pulsating torque and torque THD both at no-load and loading conditions.

References

1. Yongxing W, Xuhui W, Xinhua G, Feng Z, Wei C (2011) Vector control of induction motor based on selective harmonic elimination PWM in medium voltage high power propulsion system. In: 2011 international conference on electric information and control engineering, 15–17 Apr 2011, pp 6351–6354
2. Ben-Brahim L, Yoshino T (2014) High-power drive systems for industrial applications: practical examples, pp 695–726
3. Tripathi A, Narayanan G (2019) Analytical evaluation and reduction of torque harmonics in induction motor drives operated at low pulse numbers. *IEEE Trans Industr Electron* 66(2):967–976
4. Tripathi A, Narayanan G (2016) Evaluation and minimization of low-order harmonic torque in low-switching-frequency inverter-fed induction motor drives. *IEEE Trans Ind Appl* 52(2):1477–1488

5. Mishra S, Palazzolo AB, Han X, Li Y, Kulhanek C (2020) Torsional vibrations in open loop volts hertz variable frequency drive induction motor driven mechanical systems. In: 2020 IEEE Texas power and energy conference (TPEC), 6–7 Feb 2020, pp 1–6
6. Tripathi A, Narayanan G (2014) Evaluation and minimization of low-order harmonic torque in low-switching-frequency inverter fed induction motor drives. In: 2014 IEEE international conference on power electronics, drives and energy systems (PEDES), 16–19 Dec 2014, pp 1–6
7. Grieve DW, McShane IE (1989) Torque pulsations on inverter fed induction motors. In: 1989 fourth international conference on electrical machines and drives conference publication no. ??, 13–15 Sept 1989, pp 328–333
8. Patel HS, Hoft RG (1973) Generalized techniques of harmonic elimination and voltage control in Thyristor inverters: part I—harmonic elimination. *IEEE Trans Ind Appl IA-9(3)*:310–317
9. Zach FC, Martinez R, Keplinger S, Seiser A (1985) Dynamically optimal switching patterns for PWM inverter drives (for minimization of the torque and speed ripples). *IEEE Trans Ind Appl IA-21(4)*, 975–986
10. Nystrom A, Hylander J, Thorborg K (1988) Harmonic currents and torque pulsations with pulse width modulation methods in AC motor drives. In: Third international conference on power electronics and variable-speed drives, 13–15 July 1988, pp 378–381
11. Taniguchi K, Inoue M, Takeda Y, Morimoto S (1994) A PWM strategy for reducing torque-ripple in inverter-fed induction motor. *IEEE Trans Ind Appl 30(1)*:71–77

Attention-Based Comparison of Automatic Image Caption Generation Encoders



Cheboyina Sindhu NagaDurga and T. Anuradha

Abstract Generating captions to images has still been a challenging task. Image captioning is a combination of both computer vision and natural language processing (NLP) which has many applications in social networking and is advantageous to people who are impaired visually. There are different encoders (CNN) for feature extraction from the input image and decoders (RNN) for the language model and attention mechanisms which concentrate on relevant data to improve the model's performance. In this paper, for the comparison of encoders, VGG19 and ResNet152 are used and LSTM as a decoder to generate captions. Along with the decoder, visual attention mechanism is used which allows the human or a system to concentrate on the essential parts from the input data. Visual attention mechanism is also widely used in video analytics. The proposed work uses the MSCOCO dataset for both architectures. The generated captions are then compared with the actual captions using the BLEU score. From the proposed models, the generated captions are 80 per cent accurate.

Keywords Deep learning · Image captioning · ResNet152 · VGG19 · Visual attention · BLEU

1 Introduction

Image captioning is still an interesting task in the field of artificial intelligence even many methods have achieved better results. When a person is shown an image and requested to identify the objects and describe the image, it would be easy for them. But, for the machine, it might be difficult where it should classify the image, identify the object, need to pre-process the image, extract the features from the image, and then generate the caption. Moreover, the advancement in deep learning methods made it easy to generate captions. Image captioning has wide applications in social media, indexing, helping people who are not having clear eyesight, and virtual assistants.

C. Sindhu NagaDurga (✉) · T. Anuradha
Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

Every image has its relative content based on the identified objects that are in it [1]. There are different models which are used as encoders and decoders. Most of the convolution neural networks are used as encoders and recurrent neural networks as decoders [2, 3]. Many datasets are used for image captioning where each image in the dataset contains five descriptions.

2 Related Work

Most of the works have proved that image caption is an interesting task and not an easy one where every work strives to implement a new architecture. Farhadi et al. [1] proposed a learning method between the captions and images using a graphic model. Vinyals et al. [2] and Fang et al. [3] explained that CNN is used for extracting the features and RNN for language modelling. Performance of State-of-art [4] image captioning. Vinyals et al. [5] introduced novel approach using convolution and recurrent neural decoder (LSTM). Karpathy and Fei-Fei [6] proposed a model where detectors are used to extract the features from the image. Xu et al. [7] proposed a novel approach with visual attention to generating image caption where [8] used this approach as the base for proposed work to get better results. Demirel et al. [9] proposed a method that uses object detection to identify unseen objects in the image. You et al. [10] proposed a semantic attention model which uses a top-down approach. Radford et al. [11] identified the gap between supervised and unsupervised learning. Hossain et al. [12] compared different encoders and the captions that are generated with different existing models. Chen et al. [13], Lu et al. [14], and Gao et al. [15] use encoder–decoder architecture by dividing the image into grid format considering the CNN used, and the output is represented in grid visual.

3 Proposed Method

In the proposed model, VGG19 and ResNet152 are used as encoders and LSTM as decoder with visual attention. The process flow of the proposed method for generating image captions is as shown in Fig. 1. To analyse the generated captions, an evaluation metric is used which compares the predicted caption with the original caption.

3.1 Algorithm

- Take an image from the MSCOCO dataset as input initially to build the model.
- Use CNN encoder to extract features from the input image based on its weights.

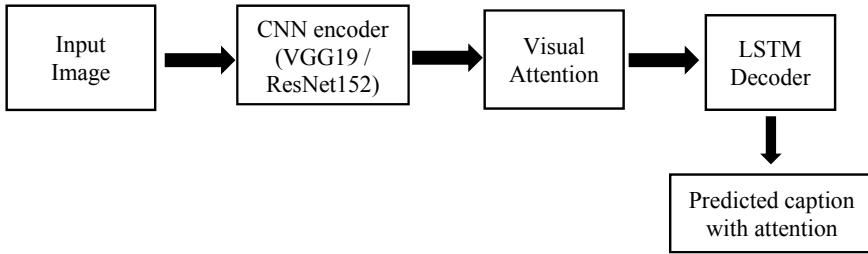


Fig. 1 Architecture diagram for proposed method

- Build ResNet152 encoder with 152 layers having a fixed vector size of 224×224 with a kernel size of 3×3 .
- Build VGG19 encoder with 16 layers having a fixed vector size of 224×224 with a kernel size of 3×3 .
- Give the output of the last convolution layer as input to the decoder.
- Generate the caption's using LSTM is used as the decoder.
- Use LSTM with visual attention to extract only the required features instead of all features.

3.2 Decoder

Long short-term memory (LSTM) [16] is used as the decoder which is one of the recurrent neural network designed to store information for a long time. LSTM possesses three components; one is to update, to remember, and to forget. LSTM is used as a language model for generating the correct caption. Equations (1)–(5) represent the mathematical functionality of LSTM.

$$f_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_t) \quad (1)$$

$$C_t = f_t \cdot C_{t-1} + i_t * C_t \quad (2)$$

$$i_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_t) \quad (3)$$

$$\overline{C}_t = \tanh(W_t \cdot [h_{t-1}, x_t] + b_t) \quad (4)$$

$$h_t = o_t * \tanh(C_t) \quad (5)$$

3.3 Encoders

VGG19: [17] is one of the pre-trained convolution neural networks which contains 19 layers. It contains a fixed input size of 224×224 with a 3×3 kernel size. It contains three fully connected layers where the first two layers contain 4096 and the third layer contains 1000 channels. The built model of VGG19 is shown in Fig. 2 which is used in the proposed work.

ResNet152: ResNet [18] is a residual network that learns by the residual function representation. ResNet152 is one of the deep residual networks which has 152 layers. This model is trained with the weights of the ImageNet. The model is as shown in Fig. 3.

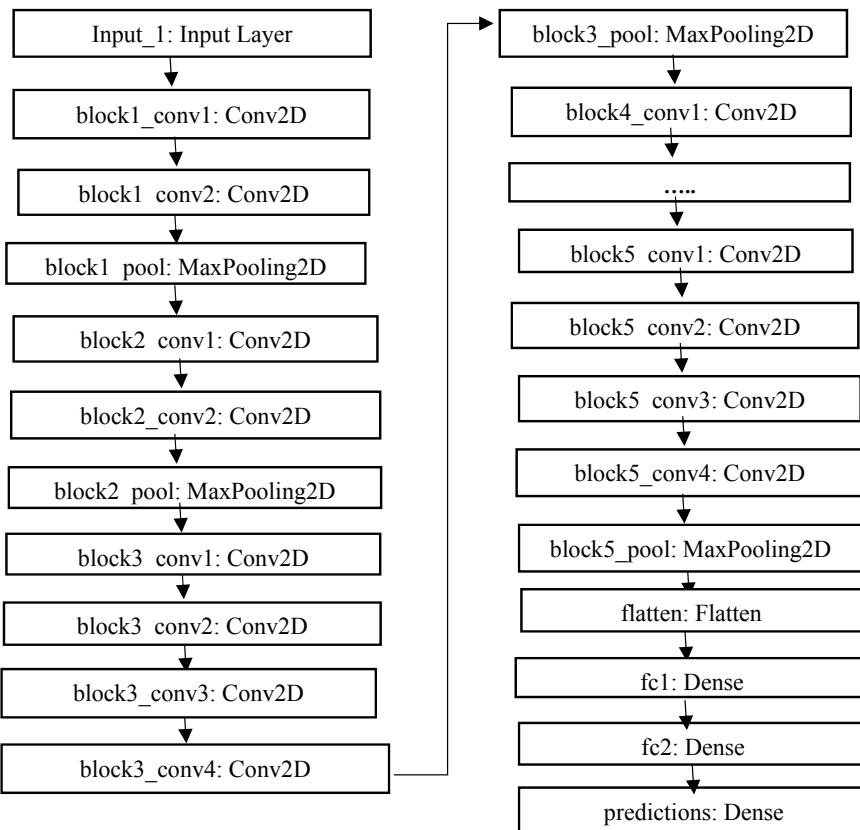


Fig. 2 Layers in VGG19 architecture

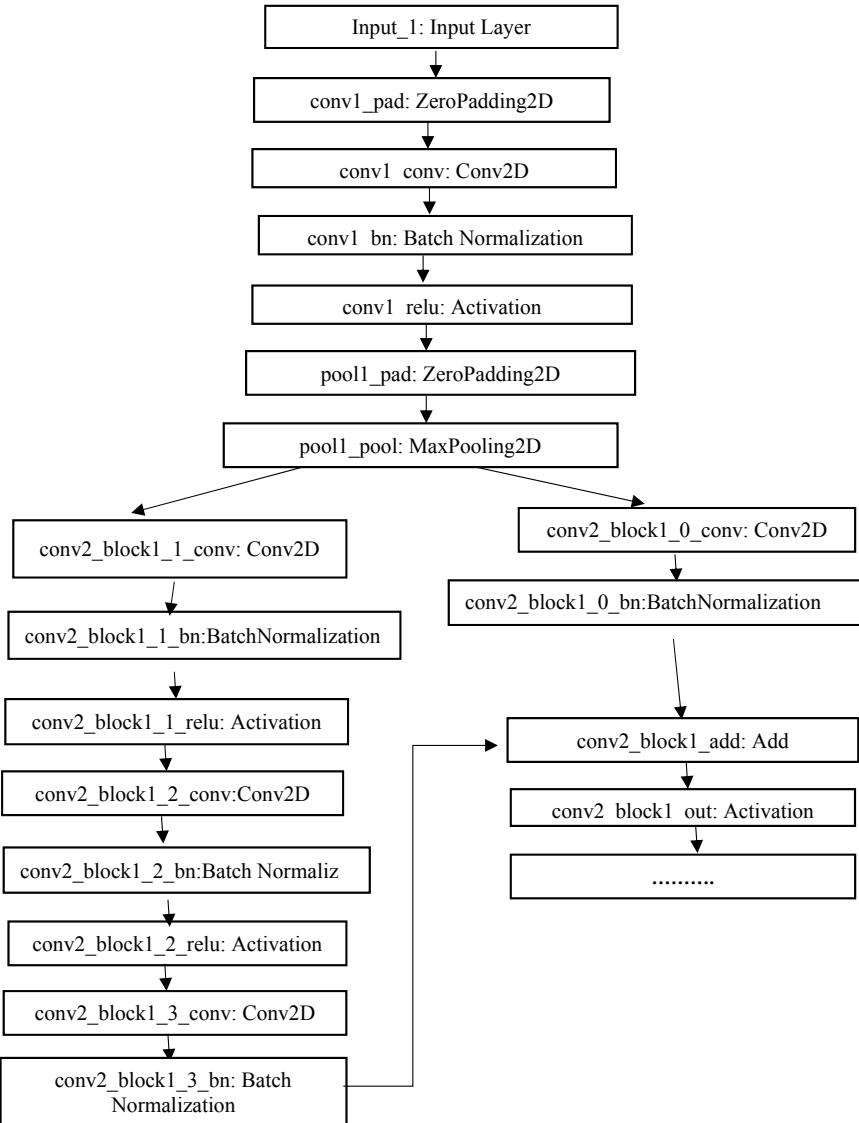


Fig. 3 Layers in ResNet152 architecture

3.4 Visual Attention

Attention mechanisms replicate human behaviour. An input image will be divided into n modules, and they generate n hidden states. Local and global are two types of mechanisms, whereas in this paper, the local mechanism (Bahdanau's attention [19]) is used.

3.5 *Optimization*

Adam optimizer [20] is one of the stochastic gradient descent algorithms which computes the individual learning rates of different parameters. The proposed method uses regularization and a dropout of 0.5.

4 Experimental Work

4.1 *Dataset*

MSCOCO [21] dataset is implemented by the Microsoft team, where the images are collected naturally. It contains 328 k images with five descriptions for each image in the dataset. The size of the image in the dataset is $480 \times 640 \times 3$. In this paper, around 30,000 images are used for training.

4.2 *Pre-processing*

For image caption generation, we need to pre-process both images and captions. When the image is given as input, the image will be resized based on the encoder weights. For VGG19 and ResNet152, the vector size is $224 \times 24 \times 3$. The features are extracted based on these weights. The caption size should range between 8 and 49.

4.3 *Implementation*

The encoder-decoder architecture is used for implementing the method where VGG19 and ResNet152 are used as an encoder which is responsible for extracting the features from the image after resizing the image into their fixed vector size. The output of the last convolution layer is given as input to the attention, that mechanism will try to take the required features, and then, the decoder LSTM will try to remember the previous output and generates the caption word by word in grid-based format by having the tokens <start> and <end> which says the model when to end the caption. The generated caption will then be analysed with the original caption by evaluation metrics.

4.4 Bilingual Evaluation Understudy (BLEU)

BLEU metrics is one of the evaluation metrics which is to evaluate the generated captions from the original caption. The score value ranges from 0 to 1. While evaluating, it considers different parameters and human judgement. BLEU score is calculated with the comparison of individual and generally translated sentences to find the high correlation, and then, the generated scores are averaged using Eq. (6).

$$p = m/w_t \quad (6)$$

where m is count of words in the original sentence, and w_t is the count of words in the reference text.

5 Results

To generate the captions for the image, automatically two different models are used, and the features are extracted based on them. To avoid overfitting, a large dataset MSCOCO is used. To reduce the loss value, Adam optimizer is used with regularization and dropout. As the number of epochs is increased, there is a decrease in loss value. The epochs range from 20 to 30. The result of loss function for both encoders (VGG19 and ResNet152) is as shown in the below figures. When compared to loss, ResNet152 is better than VGG19.

Fig. 4 Loss plot (VGG19) for 30 epochs shows a decrease in loss value from 0.9 to 0.3 with the increase in epochs

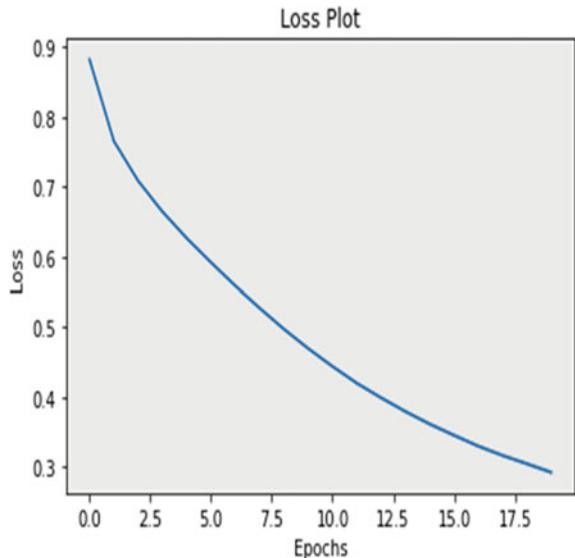


Fig. 5 Loss plot
(ResNet152) for 30 epochs
shows a decrease in loss
value from 1.2 to 0.2 with
the increase in epochs

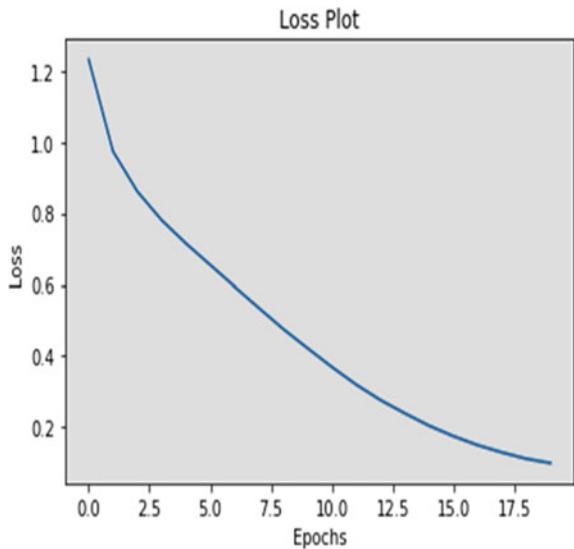


Figure 4 represents the loss plot for the VGG19 encoder from the loss that is generated for each epoch. The loss value for the first epoch is initially at 0.9, and later with the increase in the number of epochs, the loss value gradually decreased to 0.3.

Figure 5 represents the loss plot for ResNet152 encoder, from the loss that occurred at each epoch. The loss value for the ResNet152 encoder initially started at 1.2, and with the increase in epoch rate, the loss value decreased to 0.1 for 30 epochs. But the difference is VGG19, and loss value is decreasing slow when compared to ResNet152.

5.1 Output Images for VGG19 Model

Figure 6 shows the output generated using VGG19 encoder with visual attention mechanism for the input image. It shows both real and predicted captions without BLEU score. Figure 7 shows predicted caption along with the BLEU score.

5.2 Output Images for ResNet152

Figure 8 shows the output image generated, and Fig. 9 represents the predicted caption along with BLEU score for the ResNet152 encoder. The output images represent both real and predicted captions with a visual attention mechanism.

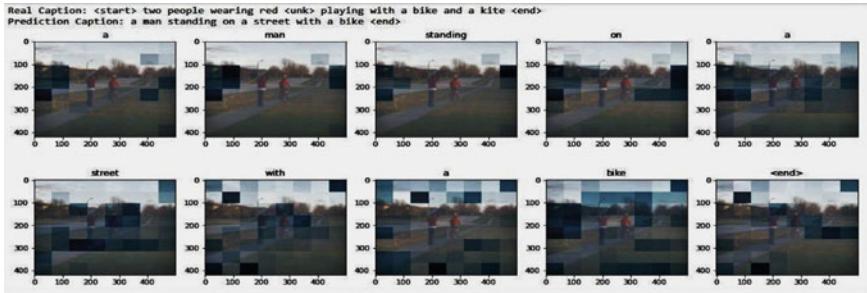


Fig. 6 Predicted caption without BLEU score—VGG19

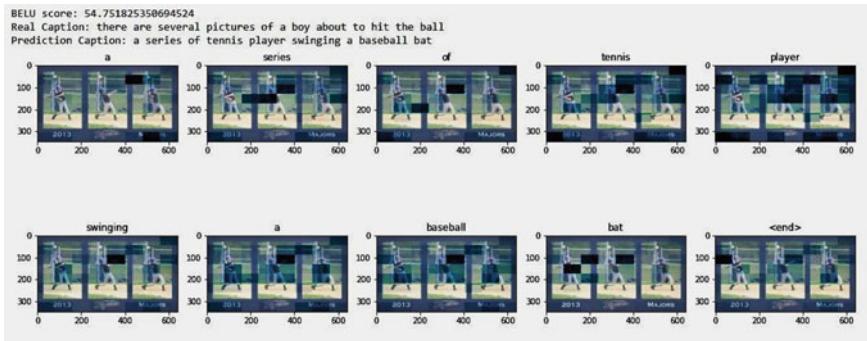


Fig. 7 Predicted caption with BLEU score—VGG19

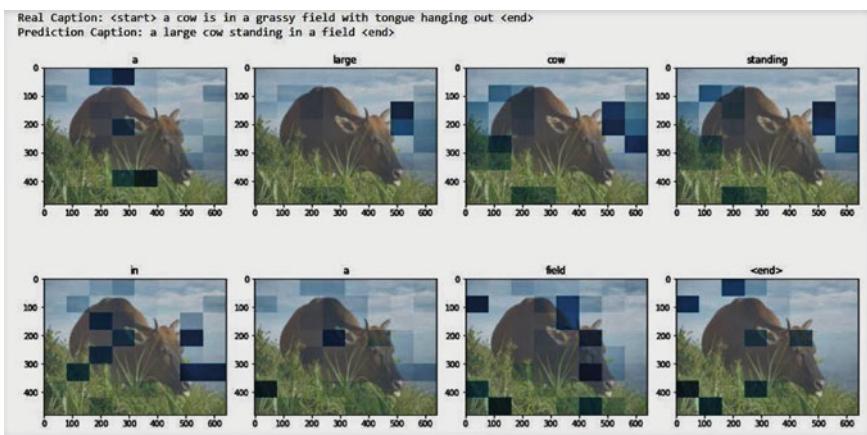


Fig. 8 Predicted caption without Bleu score—ResNet152



Fig. 9 Predicted caption with BLEU score—ResNet152

6 Conclusion

In this paper, the implementation is done using TensorFlow and Keras. Both encoders are used to extract the features from the image dataset MSCOCO based on their vector sizes and LSTM for language models. Among the two models worked, the encoder ResNet152 is doing better than VGG19, where few images are giving the least score by the evaluation. In the future, we work on extracting better features for the images which are having low pixel values or blurred to acquire better performance and to generate accurate captions for every image in the dataset.

References

1. Farhadi A, Hejrati M, Sadeghi MA, Young P, Rashtchian C, Hockenmaier J, Forsyth D (2010) Every picture tells a story: generating sentences from images. In: Computer vision—ECCV. Springer, pp 15–29
2. Vinyals O, Toshev A, Bengio S, Erhan D (2015) Show and tell: a neural image caption generator. [arXiv:1411.4555](https://arxiv.org/abs/1411.4555)
3. Fang F, Wang H, Tang P (2018) Image captioning with word level attention. In: 25th IEEE international conference on image processing (ICIP), pp 1278–1282
4. Anderson P, He X, Buehler C, Teney D, Johnson M, Gould S, Zhang L (2018) Bottom-up and top-down attention for image captioning and visual question answering. In: Computer vision and pattern recognition
5. Vinyals O, Toshev A, Bengio S, Erhan D (2017) Show and tell: lessons learned from the 2015 MSCOCO image captioning challenge. *IEEE Trans Pattern Anal Mach Intell* 39(4):652–663

6. Karpathy A, Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: Computer vision and pattern recognition. [arXiv:1412.2306](https://arxiv.org/abs/1412.2306)
7. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Proceedings of the 32nd international conference on machine learning. PMLR 37:2048–2057
8. Lu J, Xiong C, Parikh D, Socher R (2017) Knowing when to look: adaptive attention via a visual sentinel for image captioning. In: Computer vision and pattern recognition. [arXiv:1612.01887](https://arxiv.org/abs/1612.01887)
9. Demirel B, Cinbis RG, Ikitzler-Cinbis N (2020) Image captioning with unseen objects. arxiv.org
10. You Q, Jin H, Wang Z, Fang C, Luo J (2016) Image captioning with semantic attention. In: Computer vision and pattern recognition. [arXiv:1603.03925](https://arxiv.org/abs/1603.03925)
11. Radford A, Metzand L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arxiv.org
12. Hossain MDZ, Sohel F, Shiratuddin MF, Laga H (2018) A comprehensive survey of deep learning for image captioning. In: Computer vision and pattern recognition. [arXiv:1810.04020](https://arxiv.org/abs/1810.04020)
13. Chen L, Zhang H, Xiao J, Nie L, Shao J, Liu W, Chua T (2017) SCA CNN: spatial and channel-wise attention in convolutional networks for image captioning. In: Computer vision and pattern recognition. [arXiv:1611.05594](https://arxiv.org/abs/1611.05594)
14. Lu J, Xiong C, Parikh D, Socher R (2017) Knowing when to look: adaptive attention via a visual sentinel for image captioning. In: Computer vision and pattern recognition. [arXiv:1612.01887](https://arxiv.org/abs/1612.01887)
15. Gao L, Guo Z, Zhang H, Xu X, Shen HT (2017) Video captioning with attention-based LSTM and semantic consistency. IEEE transactions on multimedia. <https://doi.org/10.1109/TMM.2017.2729019>
16. Gong Y, Wang L, Hodosh M, Hockenmaier J, Lazebnik S (2014) Improving image sentence embeddings using large weakly annotated photo collections. In: European conference on computer vision. Springer
17. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large- scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
18. Chu Y, Yue X, Yu L, Sergei M, Wang Z (2020) Automatic image captioning based on ResNet50 and LSTM with soft attention Hindwai
19. Bahdanau D, Cho KH, Bengio Y (2014) Neural machine translation by jointly learning to align and translate ICLR. arxiv.org
20. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arxiv.org
21. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollar P, Zitnick CL (2014) Microsoft COCO: common objects in context ECCV. Springer
22. Papineni K, Roukos S, Ward T, Zhu WJ (2002) BLEU: a method for automatic evaluation of machine translation. ACL '02: Proceedings of the 40th annual meeting on association for computational linguistics, July 2002, pp 311–318. <https://doi.org/10.3115/1073083.1073135>

Implementation of Image Denoising with Reversible Multipliers Using Xilinx System Generator



Bandlapalli Sudharani and Gunapati Sreenivasulu

Abstract To trade-off precision in performance metrics like power, area, and delay, approximate computations have been proposed as a potential alternative for error-tolerant applications. Researchers have proposed many approximate designs for enhancing one of these metrics. In this work, a new approximate adder was designed with reversible logic gates which can have better performance than existing approximate adders. The addition of partial products is achieved by utilizing newly proposed approximate half adder and full adder circuits with reversible logic circuits, which can lessen the time delay and area of the conventional multiplier design. In this paper, a novel 8×8 -bit multiplication was designed, and it is implemented in image processing as an application. The entire concept of this proposed work is taken as application in image processing where we use it in image denoising. A novel technique of image denoising has been unveiled in this proposed work, which is called modified adaptive median filter. A qualitative noise removal image can be obtained by combining the method of post-processing of median filter with small-sized masks of iterative processing. Novel denoising processes can be carried out with the aid of the proposed multiplier design.

Keywords Approximate computing · Reversible multiplier · Image denoising · Xilinx system generator

1 Introduction

For modern electronic designs, a faster design with a smaller area and reduced power consumption is necessary. The design of the multiplier in microelectronics is a fundamental unit and is commonly used in many circuits for which the multiplication process should be correctly optimized [1]. In general, multipliers have extended latency, large area and consume vast power. In VLSI design, the major thing is to design a power-efficient multiplier. For reducing the area and power consumption,

B. Sudharani (✉) · G. Sreenivasulu

Department of Electronics and Communication Engineering, S.V.U. College of Engineering, S.V. University, Tirupati, Andhra Pradesh, India

approximate computing method was considered. By using approximation, there are more chances to have error rate. Applications such as image processing [2], wireless communications, and mining are tolerant of certain errors. Errors are not considered in those applications.

Some defects do not make huge variations in applications such as image, dedicated to human sensing due to visual limitations of human beings. In several DSP systems, inputs are noisy which are considered from real time. So, these systems are not accurate. There are many additional applications that involve the use of statistical computation, such as clustering and recognition techniques which are used in data processing [3]. Small errors in the calculation would not impose significant output loss due to statistical nature of these applications. Therefore in several applications that can withstand the loss of exactness, imprecise computing is applicable. The areas of digital signal processing, robotics, and machine learning are addressed by our applications.

2 Earlier Work

The three different steps for multiplier design are: The partial product generation, reductions in partial products, and adders or compressors are used to generate the final products [4]. The third step, which is designed with approximate adders or compressors, is selected for this research work.

In the first stage, PP is reduced by using AND, OR logic compression, and the further stages are calculated with four different full adder structures which are shown from Figs. 1, 2, 3, 4, and 5.

An approximate full adder design 1 is designed with 3 reversible gates RG1, RG2, and RG3 with two inputs as '0' and A, B, C_i , two outputs sum and carry, and two garbage outputs which is shown in Fig. 1.

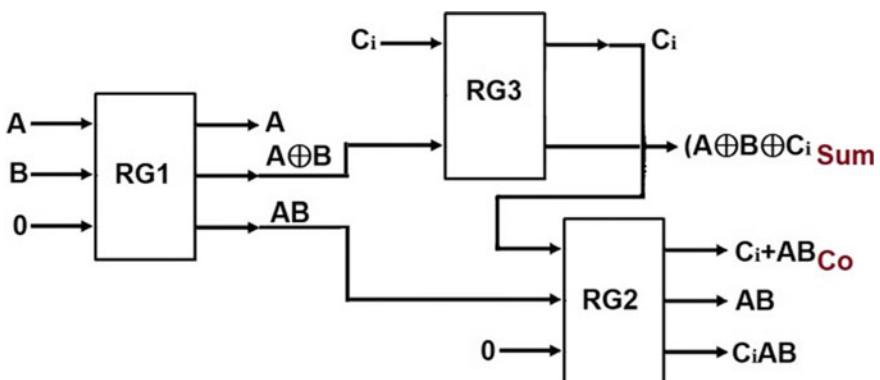
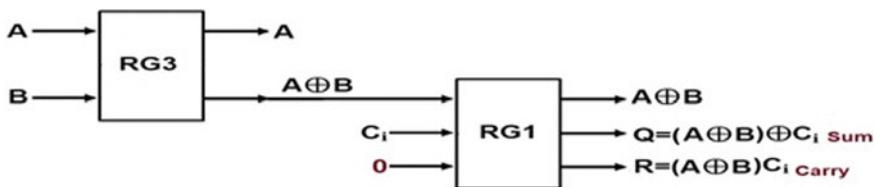
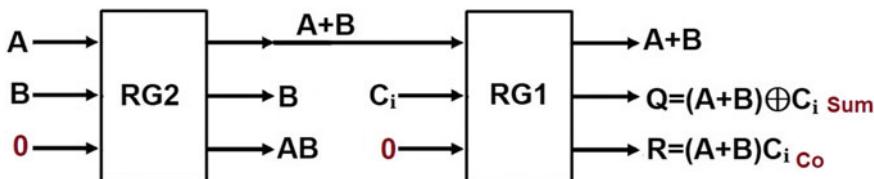
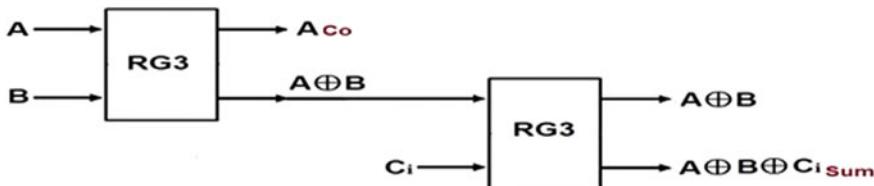
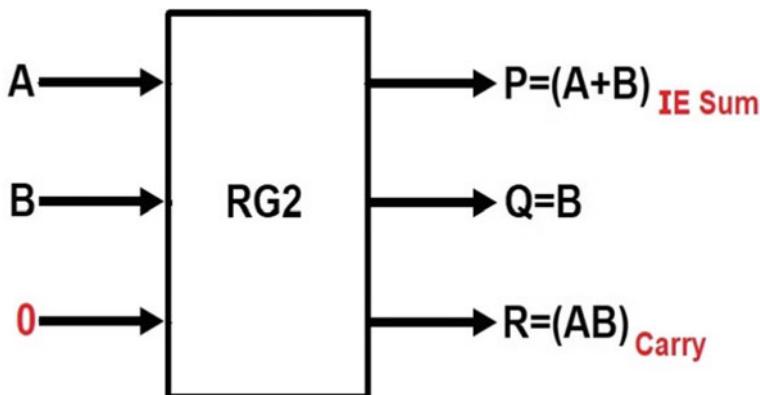


Fig. 1 Approximate adder design 1

**Fig. 2** Approximate adder design 2**Fig. 3** Approximate adder design 3**Fig. 4** Approximate adder design 4**Fig. 5** Approximate half adder design

An approximate full adder design 2 is designed with 2 reversible gates RG1 and RG3 with inputs as ‘0’, A, B, C_i , two outputs sum and carry, and two garbage outputs which is depicted in Fig. 2.

An approximate full adder design 3 is designed with 2 reversible gates RG2 and RG1 with two inputs as ‘0’ and A, B, C_i , two outputs sum and carry, and three garbage outputs which is depicted in Fig. 3.

An approximate full adder design 4 is designed with reversible gates using two RG3 with inputs as A, B, C_i , two outputs sum and carry, and one garbage output which is depicted in Fig. 4.

An approximate half adder design is implemented with reversible gates using RG2 with inputs as $A, B, '0'$, two outputs sum and carry, and one garbage output which is depicted in Fig. 5.

3 Methodology

Denoising is a technique for minimizing the inherent noise from an acquired image. For reducing the noise from an image, there are lot of techniques. The method of denoising is selected according to the kind of image and the model of noise [5]. There are two primary ways for image denoising which are available:

1. Spatial domain filtering and
2. Transform domain filtering.

3.1 *Spatial Domain Filtering*

Spatial filters are the traditional method for noise reduction from image data [6]. The preferred approach is spatial filtering when additive noise is merely present. It is divided into two classes: linear and nonlinear.

3.1.1 Linear Filters

This is the best method when additive noise occurs exclusively. In terms of mean square error, for Gaussian noise, the simplest linear filter is the mean filter. Sharp edges are twisted, lines are obliterated, and all fine details are lost in the image. It comprises a mean filter and a Wiener filter [7].

3.1.2 Nonlinear Filters

This is the method to utilize when multiplicative and functional-based noise is present. Noise can be reduced without defining it solely by utilizing nonlinear filters.

The value of an output pixel is determined in this instance by the pixels median of the neighborhood [8]. Low pass filters of pixel groups use spatial filters, which indicate that noise occurs in larger regions of spectrum. In most cases, spatial filters reduce noise to an adequate level, but by blurring images and making picture edges invisible [9].

Nonlinear filters are used to remove noise from the image, although no clear effort to define noise has been undertaken. Such a filter generally reduces noise to a sufficient extent and causes the edges to disappear in the image because of the risk of blurring images [10]. Analysis on denoising (research) is going on, and there have been some new proposed approaches to deal with those problems.

Median Filter

The median filter is considered to be a nonlinear filter. The median filtering process is performed first by determining the median value of the entire window and then replacing each entry with the median value of the pixel. The median is simple to calculate, if a window has an odd number of entries: t is the middle value just after all the elements in the window have been numerically arranged [11]. However, if the number of entries is even, there may be many medians.

In image processing and time series processing to enhance smoothness, this median filter is utilized. The benefit of median filtering is that it is less susceptible to extreme values than the mean called outliers. As a consequence, it is capable of deleting these outliers without compromising the image's sharpness.

4 Implementation of the Proposed Method

The flowchart for the proposed method is shown in Fig. 6.

Step 1

In grayscale images, impulse noise is represented by random values (RV) of pixels (values ranging from 0 to 255) in the corrupted image, or fixed values (FV), also known as ‘salt and pepper’ noise, which is created by a random partial distribution of (value 0) black pixels and (value 255) white pixels into the image. The image $O(i, j)$ -corrupted salt and pepper noise is described by $NRV(i, j)$ which is represented as follows.

$$NRV(i, j) = \begin{cases} n(i, j) \in [0, 255] \\ 0 \end{cases}$$

Step 2

The $3 * 3$ window pixel representation of the median filter is shown in Fig. 10. In this figure, the median of the pixel is calculated as arranging the pixels in either

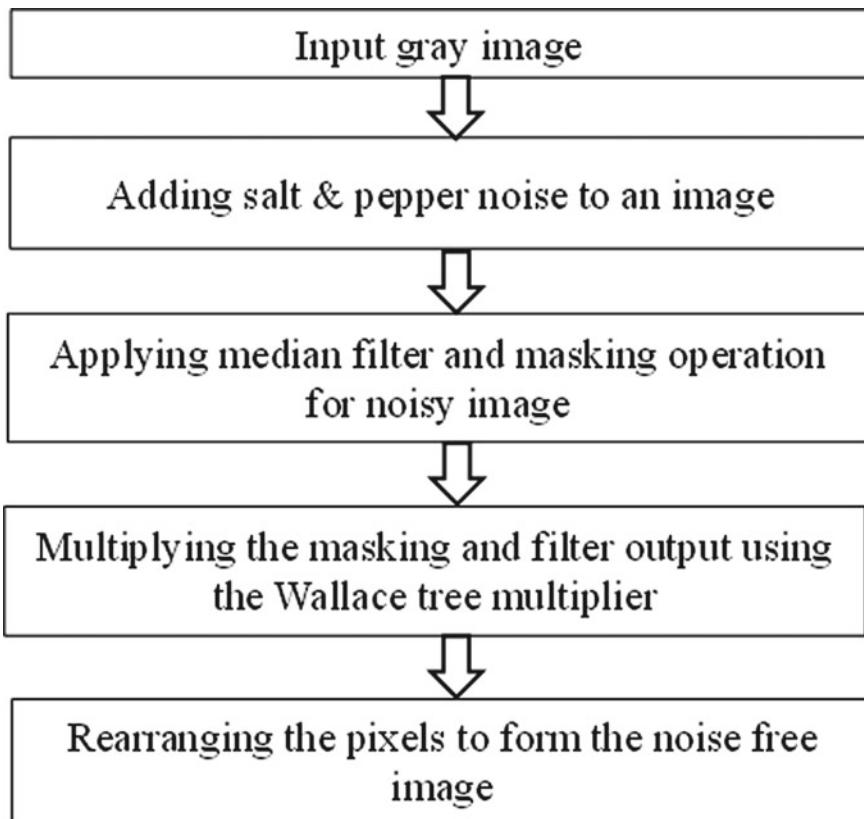


Fig. 6 Flowchart for proposed method

ascending or descending order. The median value is replaced by the center pixels of the 3×3 window which is shown in Fig. 7.

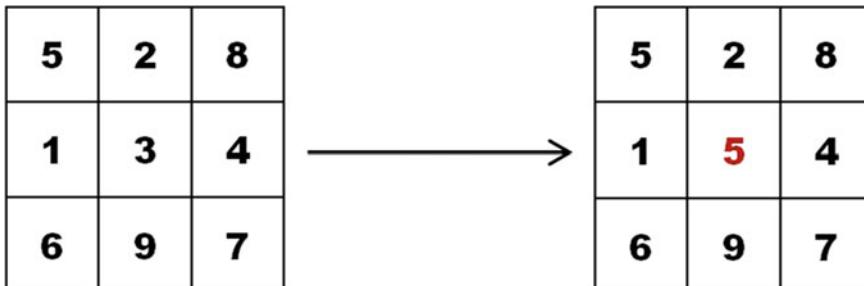


Fig. 7 Median filter for 3×3 matrix

The image denoising technique was implemented with the proposed multiplier structure. Let us consider $\{y_{i,j}^{(0)}, y_{i,j}^{(1)}, y_{i,j}^{(n)} \dots\}$ as the pixels in the grayscale image and $\{g_{i,j}^{(0)}, g_{i,j}^{(1)}, g_{i,j}^{(n)} \dots\}$ as the pixels of the distorted image, and those are identified in accordance with the existence of distortion in the pixel (n) of $y_{i,j}^{(n)}$ as follows:

$$g_{i,j}^{(n)} = \begin{cases} 0, & \text{if } y_{i,j}^n \text{ is not distorted} \\ 1, & \text{if } y_{i,j}^n \text{ is distorted} \end{cases} \quad (2)$$

where $y_{i,j}^{(0)}$ represents the starting pixel of a noisy image, $y_{i,j}^{(n)}$ represents the nth iteration of $\{g_{i,j}^{(n)}\}$, the value of $g_{i,j}^{(n)} = 0$ represents the noise less pixels of the respective (i, j) coordinates, and value of $g_{i,j}^{(n)} = 1$ represents the distorted pixels of the (i, j) coordinates to be constant. If $y_{i,j}^{(0)} = 0$ or $y_{i,j}^{(0)} = 255$, then the initial values of $g_{i,j}^{(0)}$ set to 1, and if $0 < y_{i,j}^{(0)} < 255$, then the initial values of $g_{i,j}^{(0)}$ are set to zero.

Step 3

Let us consider the distorted image as N and applying the median filtering and masking of an image as M_f and M_k , respectively. The distortionless pixels are generated as follows.

$$D_L = M_f * M_k$$

The novel Wallace tree multiplier is considered from the black box through the gateway In, the inputs of the black box are median filter image and mask pixels, and the outputs of the black box are passed through the gateway Out which is depicted in Fig. 8.

The median filter applied to an image and mask operation can be done for that image which is shown in Fig. 9.

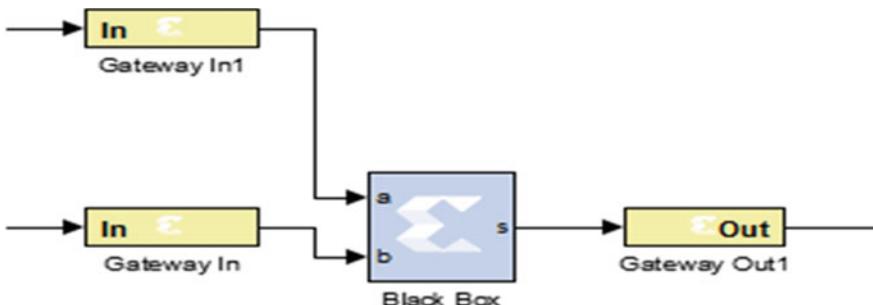


Fig. 8 Black box of the proposed multiplier architecture

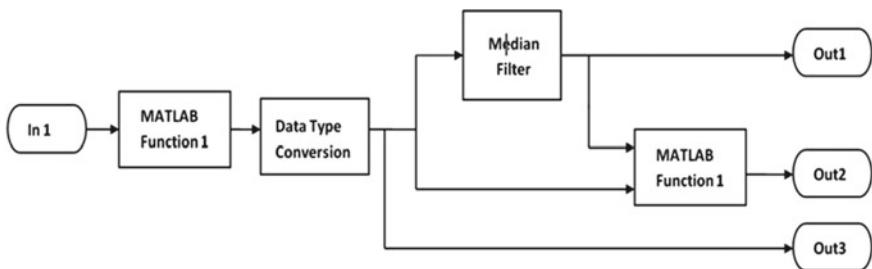


Fig. 9 Modified median filter architecture

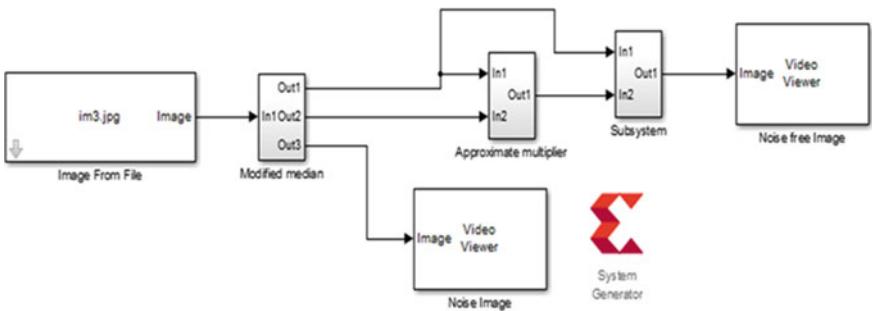


Fig. 10 Modified median filter implementation in XSG

The mask pixels and the median filter pixels were multiplied, and the modified median filter is shown in Fig. 10.

Step 4

Each pixel of the D_L can be rearranged from the noise-free image which is given as follows

$$f(x) = \{N = D_L\} \quad \text{when } D_L < 0$$

where $f(x)$ is noise-free image.

5 Results and Discussion

The image with size $256 * 256$ was considered and applied the salt and pepper noise to an image with the default commands by using software package of Xilinx system generator (Xilinx ISE 14.7 configured with MATLAB 2013b). The input noisy image can be obtained by adding noise to the image which is shown in Fig. 11.



Fig. 11 Noisy image

The mask pixels and the median filter pixels were multiplied, and rearranging the pixels with the level of the rejection, the final output image is shown in Fig. 12.

6 Conclusion and Future Scope

In this paper, a new reversible approximate multiplier was proposed based on Wallace tree scheme, so that the multiplier's efficiency is improved by means of area and delay. The entire novel multiplier was used in the image denoising technique. By using median filter and the masking technique, the performance of the image will be high.



Fig. 12 Noise-free image

Acknowledgements We would like to thank the management and Principal of Sri Venkateswara College of Engineering, Tirupati for their support in carrying this research work.

References

1. Momeni A, Han J, Montuschi P, Lombardi F (2015) Design and analysis of approximate compressors for multiplication. *IEEE Trans Comput* 64(4):984–994
2. Liu C, Han J, Lombardi F (2014) A low-power, high-performance approximate multiplier with configurable partial error recovery. *Proc Des Autom Test Eur Date* 1:1–4
3. Kishore TR, Rao KD (2008) A fast and reliable median filter for extremely corrupted images. In: IEEE Region 10 annual international conference, proceedings/TENCON [Internet]. Cited 22 July 2021. Available from: <https://www.infona.pl/resource/bwmeta1.element.ieee-art-00004766558>
4. Sivaranjani B, Krishnaveni R, Priya PS, Sathishkumar M, Anand IV (2020) Design of low power multiplier unit using Wallace tree algorithm. *Int J Eng Res* 9(02):118–122 (2020)
5. Simonetti OP, Ahmad R (2017) Low-field cardiac magnetic resonance imaging. *Aha J Int* 1–7. Available from: <http://ahajournals.org>
6. Ng PE, Ma KK (2006) A switching median filter with boundary discriminative noise detection for extremely corrupted images. *IEEE Trans Image Process* 15(6):1506–1516
7. Raut NP (2013) FPGA implementation for image processing algorithms using Xilinx system generator. *IOSR J VLSI Sig Process* 2(4):26–36
8. Vijaykumar VR, Vanathi PT, Kanagasabapathy P, Ebenezer D (2008) High density impulse noise removal using robust estimation based filter. *Int JAEENG Int J Comput Sci* 35(3). *IJCS_35_3_02*

9. Fan L, Zhang F, Fan H, Zhang C, Brief review of image denoising techniques. Available from: <https://doi.org/10.1186/s42492-019-0016-7>
10. Guragain DP, Ghimire P, Budhathoki K (2018) Implementation of FPGA based image processing algorithm using Xilinx system generator, 1457–1462
11. Rao AS, Devsingh M (2019) Image denoising using linear and non-linear filters. 6(3):1–6

Development of an Automatic Fruit Classification Using Hybrid Deep Learning Model for Super Markets



B. Sridhar, K. Sitharam Sai Kiran, N. Likhitha, K. P. V. A. Vardhan, and K. V. Nikhil

Abstract Nowadays, deep learning techniques are penetrates in all areas of applications to focus especially super markets which provide huge services to the people daily requirements. Automation of super markets is essential to support the needs of the people. In order to separate the vegetables and fruits based on various classifications such as type, color, and quality, a supervised hybrid deep learning techniques are superior than others. Proposed work is to classify and grade the fruits with consideration of various quality parameters based on image classification methods. Convolution neural networks are used to classify the image objects with multiplier layer combinations, however, to train the huge data base of 31 types of fruit image. Single layer CNN is not suitable for these type of learning huge database. So hybrid deep learning model is required to make the system automation. Here, a combinational of dense CNN and autoencoder are applied to achieve the optimized results. Proposed model exhibits high performance results with accuracy 99.35% with minimum loss 0.0245. In order to test the proposed work, we have applied a hybrid set of fruit data to a proposed hybrid network. The experimental results show the improved the accuracy and quality of the fruits classification with reduced cost of maintenance in supply chain of the super markets located in metropolitan cities.

Keywords Image classifications · Super markets · Fruit classification · Hybrid CNN models · Autoencoders · Convolution neural networks

1 Introduction

The color of the individual pixels is depend upon the neural networks which are being ideal for the segmenting data. For example, if you want to choose a vegetation zone from a satellite image, you can train the network using manually selected pixels from sample photos representing vegetation, and then use the trained NN to classify each pixel in the unknown image [1–4].

B. Sridhar (✉) · K. Sitharam Sai Kiran · N. Likhitha · K. P. V. A. Vardhan · K. V. Nikhil
Department of Electronics and Communication Engineering, Lendi Institute of Engineering and Technology, Vizianagaram, India

Few neurons are in the buried layer, as well as in the output neurons and in the input neurons, this are the only three input neurons (for R, G, and B color components).

Huge generic image datasets, like ImageNet and Pascal VOC, are being involved by the convolutional neural networks. By the way, each and every dataset for bio pictures has its own way of design and style of output and degree of complexity in a unique way. By this, strong performance on one of the dataset is not always applicable good performance on another dataset that means it changes according to the datasets [2].

The very first and beginning design model is based on the ceresin only, which consists of 10 layer model which is included with the 4 convolutional layers, where each one of it is followed by a maximum pooling layer, also with the 2 fully connected layers. The deep convolutional neural network AlexNet which is being constructed by them with in a convolutional version. The model is being tested by the 3 separate segmentation benchmark datasets. On the above 3 segmentation benchmark dataset, the variations of the base FCN have the ability to outperformed state-of-the-art approaches, demonstrating that fully convolutional networks are available alternative to traditional classification convolutional networks. The Ronneberger which is the one used as a deep fully convolutional network which consists of 23 convolutional layers on the 3 bio-image datasets: The 3 datasets are being developed by the above layers [3].

2 Back Ground

To be aware of losing added value in the markets, CNN-based approaches are very much useful and accurate in detection and classification of fruit quality. To this end, many methods and efforts are being made to improve the quality of detection of damage, disease, and fruit ripening. The quality control obtained here is carried out before, during, and after harvesting the fruits. Here, we use the recently used CNN analysis in the above area and perform the analysis. A modified AlexNet model with some layer structure. Classification includes three well-known algorithms: support vector machines (SVMs), particle swarm optimization (PSO), and inverse neural networks (BP), and the dataset has the laser-induced back scatter images, the process called acquisition is being carried out in this process. A similar kind of a apple samples of about 500 are present in dataset. The algorithms, like BP, SVM, and PSO algorithm, have the recognition rate less than the proposed CNN model for apple detection which has 92.50% [5, 6].

Using this total sample of about 341 healthy and unhealthy limes, we took a 4320 × 3240 pixel RGB image. The image is preprocessed and also needs small changes, such as background removal and resizing. The CNN models used are compared with some methods such as, fuzzy methods, KNN [7], SVM, artificial neural networks, and decision trees. The extraction method is done here because which can extracts features with local binary pattern (LBP) and graph directed chromatic gradient (HOG). As a result, an average accuracy of 100% is found in the proposed CNN model [8].

An light separation lab (LSL) is being used dataset of 270 images to capture illumination-separated of grapevine berries, for the phenol typing the distribution of epicuticular waxes here. For the image analysis method, a CNN model is being used by them here. Here, accuracy is show up to 97.3% over the six grapevine cultivars validation. To identify the papaya disease, a CNN model is used and it is presented by Munasingha. A collection of diseased images was being collected by them using a surveillance with normal conditions of the papaya farms. Here, some of the images were taken by the Internet as it available in it publicly. 5 main papaya diseases are being classified by the network itself. The classification accuracy of ~92% is being achieved for new images. The diseased region of the fruit is being detected to facilitate effective classification. Segmentation of images with CNN achieved 93% accuracy KNN and SVM-based classifiers. The nutritional deficiencies in plants of tomato are being predicted by the autoencoder models and Inception-ResNet V2. Within the fruiting and leafing phases, 571 images were captured. They have reached accuracy with autoencoders and ResNet V2 for predictive performance of 3 models upto 91% [9].

Sustika evaluated GoogLeNet, MobileNet, VGGNet, AlexNet, and Xception architectures against the two-level CNN architecture used as a standard. They also use the strawberry rating system for quality controls to assess on two datasets, two classes of strawberries, and the other four classes. VGGNet achieves the best accuracy for the datasets of 96.49 and 89.12% [10].

The internal damage of blueberries using data of hyperpectral transmittance is being detected by the Wang applied residual networks. Here, 4 machine learning algorithms were also applied in comparative linear regression (LR), random forest (RF), sequential minimum optimization (SMO), and multilayer cognition (MLP) experiments [11].

To observe the performance of classifier, the precision-recall and ROC curves are being plotted here. With the traditional ML methods, CNN models have shown better results for classification The ResNet/ResNeXt achieves accuracy of 0.8952 and 0.8844, and the classifiers SMO/RF/LR/Bagging/MLP are being obtained with the 0.8082 and 0.8784, respectively. The classification for the bananas ripening stages, Zhang prepared and found a CNN architecture designed for it. The image data contain 17,312 images of bananas of different stages of ripening, captured in standard RGB format with a resolution of 3200×2400 pixels and saved as PNG. The final accuracy was found of CNN is 95.6%, which is better than the other methods [5].

The introduction of a framework which is used to CNN-based combines a stacked sparse autoencoder (SSAE), which is well known for CNN-SSAE. In the classification of the SSAE, here, CNN network is considered the majorly for the self-learning of local image features.

On average, 85.6 and 78.3% accuracy were confirmed using the spectroscopic function according to the test results. They proposed method with the combination of spectral and spatial characteristics improved accuracy up to 91.1 and 88.6% at time of execution [9].

3 Methodology

Firstly, we need to give the datasets to the neural networks as their inputs. The image segmentation and the convolutional neural networks are being studied and observed carefully and also their applications which are present in many fields [6].

Here, we design and analyze the proposed network for cell detection.

To get the exact detecting section of the malignant cell, the image must be segmented and also various hidden layers are being implemented by the convolution neural networks [3].

Here, we observe the different fruits by using different algorithms in PYTHON and with the formation of different layers with more classifications within the layers. In this, we have 2 methods for the fruit testing process, one is training and the other is testing.

3.1 Autoencoder

Autoencoders are automatic neural networks with identical inputs and outputs. It can be used to rewrite the output by compressing the data into low dimensional code. Code, also called a latent space representation, is a compressed “summary” or “compress” of an input. The encoder, the token, and the decoder are three parts of an autoencoder. The input is applied to encoder which compresses data and generates the code at output the data reconstructed with decoder [5].

To create an automatic encoder, you need an encoding method, a decoding method, and a loss function to check the difference between the target and the output.

Autoencoders are basically dimensionality reduction algorithms. Some important characteristics of autoencoders are

- Data-specific: Autoencoders are compress the data accordingly without losing the main information of the data as they are trained.
- Lossy: The eutoencoder output will not be same as input it has some loss and it goes through lossy compression.
- Unsupervised: The autoencoder train itself when we give the data. As the autoencoders does not need explicit labels while training, these are considered as unsupervised learning technique. The autoencoders are self-supervised and create their own labels [8].

3.2 Architecture

The encoder and decoder are completely connected to a feedforward neural network, mainly an ANN. The code is a single layer of an ANN with the dimensions. The scope of the nodes within the code layer (code size) is a set of hyperparameters

before the autoencoder is trained a special offer for automatic encoders. First, the encoder is a fully connected ANN takes the input and decoders with the same ANN architecture provide the most efficient output for using the code. After process, both input and output seems to be same.

The structure of the decoder is a duplicate image of the encoder. This is not always a requirement, but it generally is. The only condition is that the dimensions of the inlet and outlet are the same. Everything in between can be played at the same time.

Parameters to consider before training an autoencoder:

- Code size: Consider the number of nodes in the middle layer. The smaller the volume, the higher the compression ratio.
- Number of layers: The number of layers depends on the code.
- Nodes per layer: The autoencoder architecture is a stacked on top of each other of autoencoders.

These layers look like a “sandwich”. The number of nodes per layer decreases with encoding and increases again in the decoding process. Decoders and encoders have a similar structure.

Loss function: It is also called binary cross-entropy or mean square error (MSE). If the input values are in the range 0, 1. Otherwise, use the root mean square error range is given as [0 1].

Autoencoders are trained in the same way as ANNs by back propagation.

3.3 Convolutional Neural Networks

A convolutional neural network (CNN) is a system which is used in deep learning to takes an input images, and it will assign related learnable weights and biases the different features/entities in image, and it will differentiate between them. When we contrast to other categorizing methods, the data for pre-processing require by CNN is considerably less [5, 6].

While basic approaches.

Requires self-constructing filters, CNN can learn these filters/attributes with enough training.

The architecture of a CNN is corresponding to that of the integrate pattern of neurons in the brain of humans and it was inspired by the organization of the visual cortex. Every neuron respond to stimuli in a restricted area of the visual region called as receptive field. Different fields like this receptive fields overlap to cover the total visual area.

Image is collection of matrix of pixel values. So, we flatten the image and feed it to multilevel preceptors for categorizing purpose. For basic binary images, this method may show an average precise value while predicting the classes but comes to complex images accuracy depends on pixels.

4 Structure of Convolution Autoencoder Attention-Based DenseNet

4.1 Classification and Parameter Learning

The construction of AND is shown in below figure, first, we will give training dataset (input images) to unsupervised pre-training, i.e., convolutional autoencoder as shown in Fig. 1. In this, unsupervised pre-training images does not have any labels those images will be trained. This is Stage-1. After this stage, we calculate optimal W and optimal B then we will move on to the stage-2, i.e., supervised training in this attention-based DenseNet is present training dataset also given to this stage. Supervised images means image having their respective images. Then, we will move on to the Stage-3; in this stage, classifiers are present to make predictions by classifying the images using different classifiers. The different classifiers present in deep learning. Those are LAZY LEARNERS and EAGER LEARNERS.

LAZY LEARNERS: These simply store the given training data and wait for the testing data to appear. The classification is done based related data which are stored in training data.

Ex.: K -nearest neighbor.

EAGER LEARNERS: These construct a classification model before receiving data for classification based on the given training.

Ex.: Decision tree, artificial neural networks.

We use a function called softmax to make the prediction at final. Accordingly, we train the model by minimizing the loss between the predicted results and the attributes.

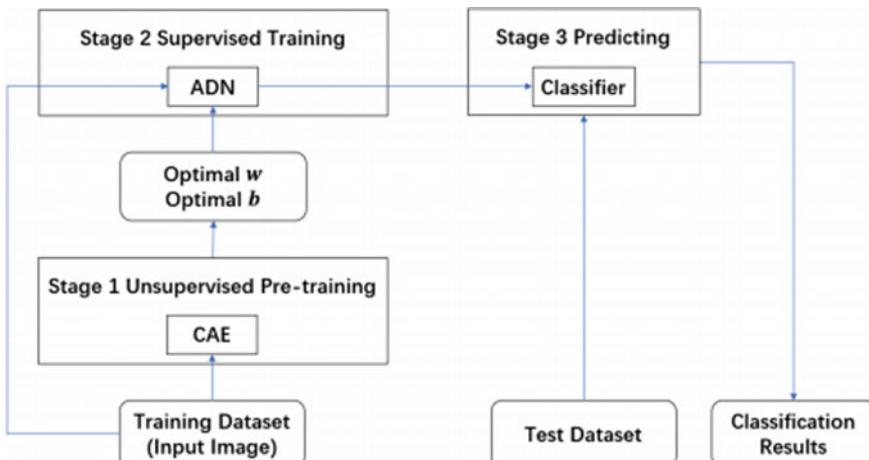


Fig. 1 Learning process of proposed work

$$\theta = \arg \min_{\theta} \left\{ - \sum_{i=1}^M g_i \log(p_i) \right\},$$

where

- M Number of images in training dataset
- g_i Image ground truth
- p_i After softmax neural network output
- θ Frame work parameters.

5 Results

5.1 Datasets

To test the efficacy of the model, we used a fruit dataset with many images. We collected the fruit dataset containing 131 classes, displayed the fruit with a black background in the image, removed the background, and sampled the image at 224×224 pixels with a batch size of 32. The fruit 360 dataset included 94,830 fruit images spread across 131 classes. 62,380 images are for training and 32,000 images are for testing. Below figure shows some samples of fruit 360 both training and test. We calculate accuracy and loss for both training and test by epochs. These epochs are like batches of the images, each epoch contains 1694 images, and total 6 epochs are there by performing each epoch, the accuracy will increases and loss will decrease. This we can observe in the result output image. The executed results of the proposed work as shown in Fig. 2, for training, for testing Fig. 3. The process of leaning of various set of fruits as shown Fig. 4. The final evaluated results are shown in the epochs are shown in Fig. 5, the accuracy and loss are sequentially evaluated and final values of the proposed algorithm have displayed.

Training Data Samples

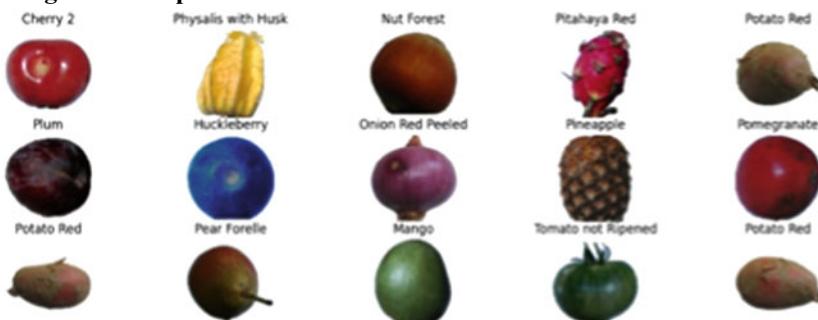


Fig. 2 Training results of the proposed method

Test Data Samples

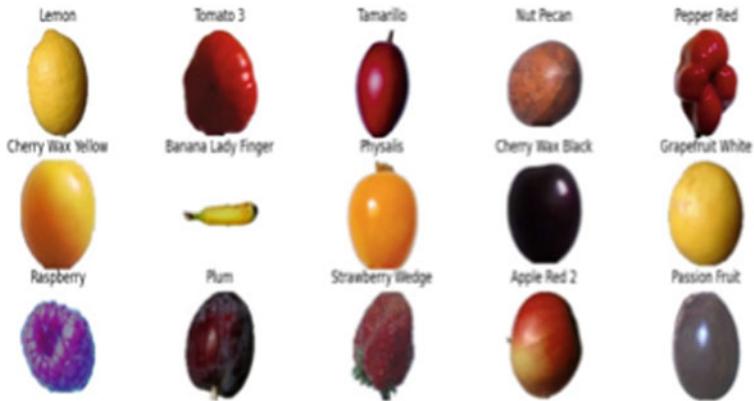


Fig. 3 Testing results of the proposed method

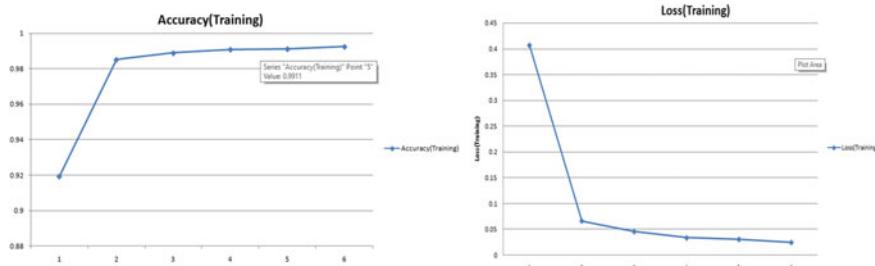
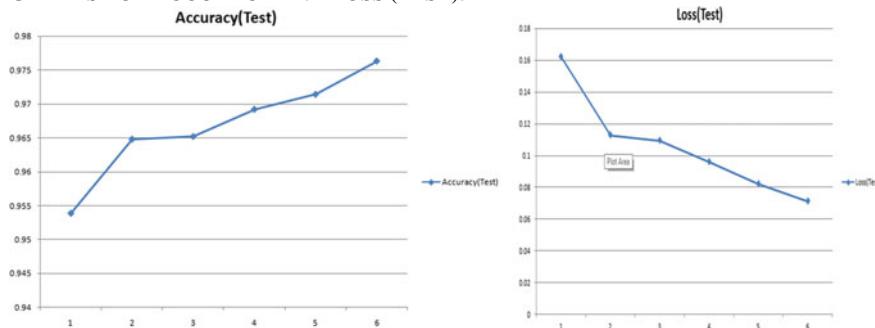
```
Found 54190 images belonging to 131 classes.  
Found 13502 images belonging to 131 classes.  
Found 22688 images belonging to 131 classes.
```

Fig. 4 Learning process of the ADE_CNN

Number of Classes in the Fruit DataSet

Fig. 5 Evaluation result of the method

Results of training and testing accuracy and loss factors as explained below.

GRAPHS FOR ACCURACY AND LOSS (TRAINING):**Fig. 6** Graph for training accuracy and loss function for versus number of iterations**GRAPHS FOR ACCURACY AND LOSS (TEST):****Fig. 7** Graph for testing accuracy and loss function for versus number of iterations

6 Graphs for Accuracy and Loss (Training)

Figure 6 depicts the average loss and accuracy curves during the model training process for tenfold cross-validation. In both curves, after the 6th-epoch, the loss values keep stable and close to 0, and the accuracy results are above 80% for training and vary for validation obtaining 99.35% at 6th-epoch. Similarly, the testing results of the proposed method also given a similar approached results as shown in Fig. 7.

7 Conclusion

Proposed work is to classify and grade the fruits with consideration of various quality parameters based on image classification methods. Convolution neural networks are used to classify the image objects with multiplier layer combinations, however, to train the huge data base of 31 types of fruit image. Single layer CNN is not suitable for this type of learning huge database. So hybrid deep learning model is required to make

the system automation. Here, a combinational of dense CNN and autoencoder are applied to achieve the optimized results. Proposed model exhibits high performance results with accuracy 99.35% with minimum loss 0.0245. It can be summarized that hybrid models such as convolutional autoencoder with prior training which can enhance the performance of multilayer convolutional neural network algorithms in fruit classification problems. Also, compared to traditional algorithms, deep learning is a way to simulate human visual perception through neural networks by extracting the local properties of the lower layer; each layer receives information from the upper layer.

References

1. Matsuda Y, Hoashi H, Yanai K (2012) Recognition of multiple-food images by detecting candidate regions. In: Proceedings of the 2012 IEEE international conference on multimedia and expo (ICME). Melbourne, Australia, 9–13 July 2012
2. Wu A, Zhu J, Ren T (2020) Detection of apple defect using laser-induced light backscattering imaging and convolutional neural network. *Comput Electr Eng* 81:106454
3. Jahanbakhshi A, Momeny M, Mahmoudi M, Zhang YD (2020) Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. *Sci Hortic* 263:109133
4. Barré P, Herzog K, Höfle R, Hullin MB, Töpfer R, Steinhage V (2019) Automated phenotyping of epicuticular waxes of grapevine berries using light separation and convolutional neural networks. *Comput Electron Agric* 156:263–274
5. Munasingha LV, Gunasinghe HN, Dhanapala WWGDS (2019) Identification of papaya fruit diseases using deep learning approach. In: Proceedings of the 4th international conference on advances in computing and technology (ICACT2019), Kelaniya, Sri Lanka, 29–30 July 2019
6. Ranjit KN, Raghunandan KS, Naveen C, Chethan HK, Sunil C (2019) Deep features based approach for fruit disease detection and classification. *Int J Comput Sci Eng* 7:810–817
7. Tran TT, Choi JW, Le TTH, Kim JW (2019) A comparative study of deep CNN in forecasting and classifying the macronutrient deficiencies on development of tomato plant. *Appl Sci* 9:1601
8. Sustika R, Subekti A, Pardede HF, Suryawati E, Mahendra O, Yuwana S (2018) Evaluation of deep convolutional neural network architectures for strawberry quality inspection. *Int J Eng Technol* 7:75–80
9. Wang Z, Hu M, Zhai G (2018) Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. *Sensors* 18:1126
10. Zhang Y, Lian J, Fan M, Zheng Y (2018) Deep indicator for fine-grained classification of banana's ripening stages. *EURASIP J Image Vid Proc* 46
11. Cen H, He Y, Lu R (2016) Hyperspectral imaging-based surface and internal defects detection of cucumber via stacked sparse auto-encoder and convolutional neural network. 2016 ASABE annual international meeting. American Society of Agricultural and Biological Engineers, St. Joseph, MI, USA, p 1

Design a Low-Cost Air Pollution Monitoring IoT System



B. Sridhar, R. Mounika, P. Nagendra Babu, Y. Pavan Kumar,
and S. Y. Ravi Raja

Abstract Air pollution is a big problem these days. Air pollution, which has become a major problem all over the world, is the most serious, dangerous, and most dangerous pollution among other pollutants such as water pollution, soil pollution, water pollution, soil pollution, and water pollution. Quality detection and control are essential. Air pollution, including noise pollution, light pollution, and heat pollution, is a major cause of diseases such as asthma, cancer, bronchitis, birth defects, and diseases such as the immune system. Therefore, we propose an air quality monitoring system that can find and check air quality in real time through the Internet of Things. It uses air sensors to detect the proximity of destructive gases in its environment and sends this information to a microcontroller. The proposed system implements a combination of Android applications, server, and gas sensor (CO_2 and DHT11) to detect ambient air quality and display the actual air condition. Sensors synthesize this information and send it to the web. This allows you to detect air pollution in different areas and take steps to prevent it. There is also a temperature sensor that estimates the room temperature. By solving the shortcomings of the existing air quality sensor, it is possible to perform several monitoring operations simultaneously. The hardest thing is that this system provides real-time data and displays the air quality based on the standard air quality. The system is used to display the air quality to the user and informs the user of how polluted or safe the surrounding air is according to the specified criteria.

Keywords Internet of Things · Arduino · Node MCU · Cloud computing · Air sensors · Sensor node technology

1 Introduction

Air pollution monitoring systems are primarily a measure of the severity of air pollution, and higher numbers indicate that air quality is more dangerous to human

B. Sridhar (✉) · R. Mounika · P. Nagendra Babu · Y. Pavan Kumar · S. Y. Ravi Raja
Department of Electronics and Communication Engineering, Lendi Institute of Engineering and Technology, Vizianagaram, India

health. The unexpected weather that has occurred in our country in recent months has exposed the environment to natural disasters, resulting in a particularly foggy environment. Fog definitely affects the specific area where the fire is occurring [1–5]. On the other hand, industrial waste such as non-supplied gas can affect the atmosphere, especially near industrial parks. These days, people who want to have an air pollution monitoring system have to pay a lot of money to install it because it is expensive. The large size makes it uncomfortable to wear. These systems are often used by government agencies and large companies with industrial plants to monitor air quality. The paper has been suggested to be portable, requires a cheaper version of the air pollution monitoring system, has a lower cost of materials, and is readily available from the Arduino and Node MCU [6–10].

The main objective of this work is the disinfection of unpolluted cities and people in our country. Air pollution estimates in smart cities these days are becoming very erratic, and there is a certain period of time to detect air pollution without knowing whether the air pollution has filled up or not. This leads to air pollution in some areas, causing unrelated diseases. Therefore, the proposed work starts monitoring all kinds of pollution to keep cities clean. When air pollution is not controlled, send a message to higher authorities at a specific place. This reduces pollution and increases human health. This is how our project helps the community maintain a good environment [11–16].

There are several steps to be achieved at the end of this work [17, 18]. They are:

- (a) Design of low-cost portable air pollutant monitoring systems that use gas sensors
- (b) Integration of Sensor, Arduino, and Node MCU to form a complete air pollution monitoring system.
- (c) Transmission and reception of air pollution data via short message service (SMS) using the Node MCU.

Various areas for achieving the purpose of the process work are described in steps [19, 20]

- (i) Detection of ambient air pollutants using appropriate sensors and signal conditioning circuits.
- (ii) Create a suitable codec to read the sensor data to the Arduino.
- (iii) The data received from the Arduino creates an appropriate command for the Node MCU and sends the data to Short Message Service (SMS).
- (iv) Connect the cloud server and monitor the data online.

The paper is divided into six main parts, and each one can be conceptually divided into two parts.

The first introduction relates to the capabilities monitoring unit, and the second relates to the vision unit. In Sect. 2, we analyze past applications of air pollution systems and the articles in the literature that inspire us to develop this project. In Sect. 3, we describe the theoretical and practical technical background applied to the development of this project. Section 4 introduces the hardware interface problem, starts with the common features, and then gets into the details that will show you

how to build the unit step by step. Section 5 introduces hardware functioning and results. Finally, Sect. 6 presents results and conclusion remarks with results obtained from real-world scenarios of air pollution systems and some potential use cases, also final session to explain some suggestions for future work.

2 Background

2.1 IoT-Based Air Quality Solution

Close monitoring is a particularly difficult part of air quality management. In order to monitor the entire cities, many monitoring stations are required to continuously monitor the air quality, which makes it difficult for experts to accurately measure the level of urban air pollution and its impact on urban life. The advent of the Internet of Things (IoT) could change these challenges. Advances in low-power wide-area networks (LPWA) have increased the availability of small, always-on handheld sensors.

With low data rates and long distances, LPWA sensors can be connected to bicycles, outdoor furniture, or people to measure and report air quality more regularly and accurately. The power of big data, such as weather and transport, can improve the background of the data and can provide insight into the causes and changes in air pollution levels. While they are reluctant to replace traditional monitoring networks, governments around the world are investing in policies and partnerships to find IoT-based air quality solutions. For example, Chongqing is currently working with China Mobile to improve the weather with built-in sensors. Quality control, on the other hand, Telefónica and Orange works with cities in Portugal, Spain, France, and Brazil [18].

Kim and Paulos [19] have developed and implemented a system called air that can measure, visualize, and transmit data on indoor air quality. Air quality from the DC 1100 is used to measure indoor air pollution, an AVD-based Arduino is built into an air quality monitor, and the iPod Touch processes, displays, and transmits data to the Arduino wirelessly. Data is transmitted every 15 s while the Arduino encodes the data as a series of beeps. It is like the iPod Touch reading your modem through the microphone port. We can use Wi-Fi to exchange real-time data from a central server.

Air pollution is a vital issue for a country. An air polluted country falls prey to many different environmental issues and health hazards. Recently in India, there are various debatable and questionable conditions which arises when it comes to air pollution, cause, from different foreign surveys it is found that, the air quality in India is very hazardous, and it is one of the worst's in the world ranking [12]. In 2002, continuous air quality monitoring was commenced through the establishment continuous air quality monitoring station (CAMS) by Department of Environment (DoE) under the worlds bank financed Air Quality Management Project (AQMP) [15]. But there is no organization or any government projects that work in detail on air pollution and air

quality. There are some international websites where they display the live air quality and unfortunately, they do not have enough data about India's air quality in their database. For India, the AQI measured by US embassy is followed widely. But the problem remains about the accuracy because the measurement device is installed on the premise of US embassy, and it can only detect quality of the surrounding air. This value cannot represent the AQI for all the cities, even the other part of Delhi City [21–25].

- The main goal of our work is to detect the gases responsible for air pollution and measure the air quality and view the pollution level so that we can evaluate which gases are responsible for pollution in which area the most and in which rate the air is being affected.

3 Experimental Studies

This section describes the project methodology applied to the project. Guidelines are used to complete the project from start to finish. The procedure includes several activities implemented under the project: successful completion [26].

3.1 System Design

This session will discuss in detail the progress made in the development of a real-time air quality reporting system. Thus, this chapter deals with the design and construction of the original system as shown in Fig. 1. The system can be divided into two types: Those are description of hardware and software implementation [23].

3.2 Hardware Development Part

The low-cost system consists of the Arduino and Node MCU sensor node system. To connect Wi-Fi and fast monitoring of the sensors data Node MCU included in the work, which accesses the network easily and remote server. Arduino Uno interfaces with DHT11 sensor for monitoring temperature and humidity of the given environment and process the data based on threshold value with reference of the other sensors those are barometric pressure and air quality sensor, which are interfacing with Node MCU. The gas sensor are detecting the carbon dioxide and carbon monoxide gases with reference of pressure and DHT11 sensors [14].

The threshold values are adjusted in the system as per the quality standards of air. This can continuously monitor with IoT could server-ThinkSpeak. Users can access the data either by using Android mobile or by Internet. When the gases have reached above the threshold value, the system has sent immediate alarm to the users. So they

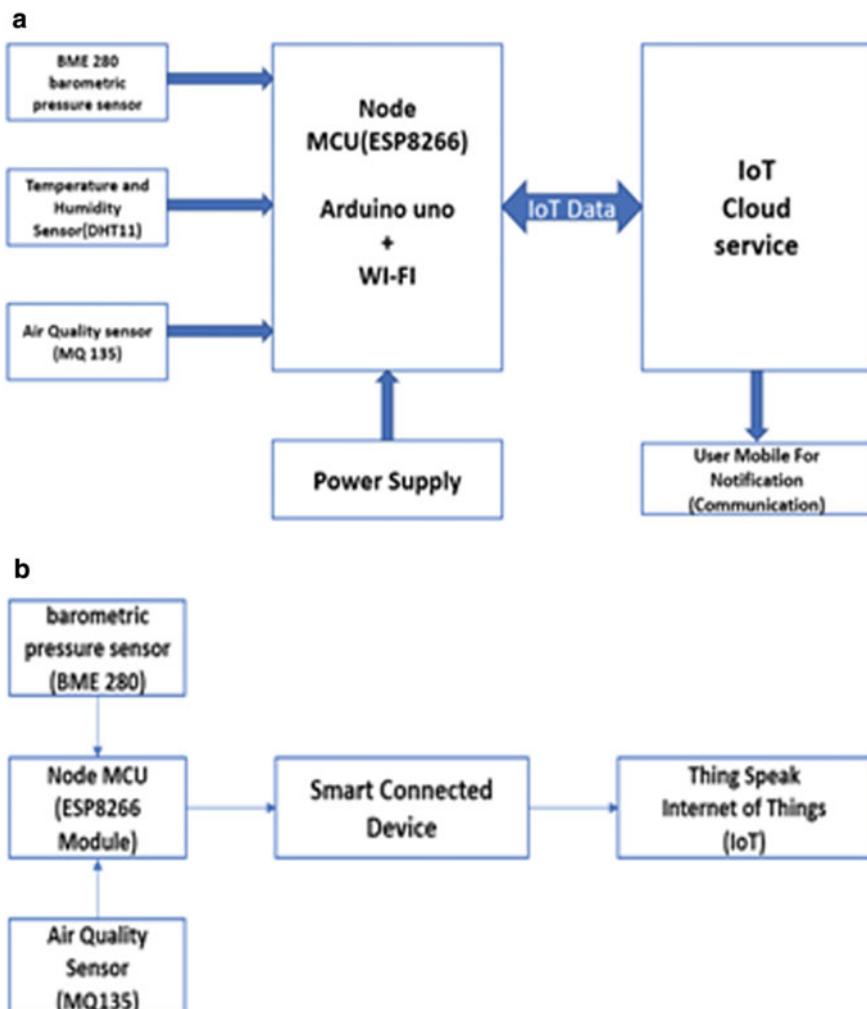


Fig. 1 **a** Block diagram of air quality monitoring and sensing. **b** Block diagram of sending the data to ThingSpeak using NodeMCU

immediately save their lives, and also, the proposed system has connected exhaust fans to exhaust the gas from that environment. The detailed hardware has shown in the block diagram Fig. 1.

3.3 Software Development Part

The process of the software development has given in flowchart Fig. 2. Flowchart for system programming, another flowchart is for displaying temperature, humidity and air quality on an LCD. Results of the obtained data to send temperature, humidity to ThingSpeak in order to display the data.

4 Results and Discussions

Connections

As shown in the picture, 3 DHT11 voltages are connected to ground +5 and 0 V, and all signal pins of Arduino Uno can be connected to 8 pins, and the rest of the components like MQ135 voltage level and ground to +5 V and analog PIN connectors. The analog pins of the Arduino Uno [22] are connected to A0 (Fig. 3).

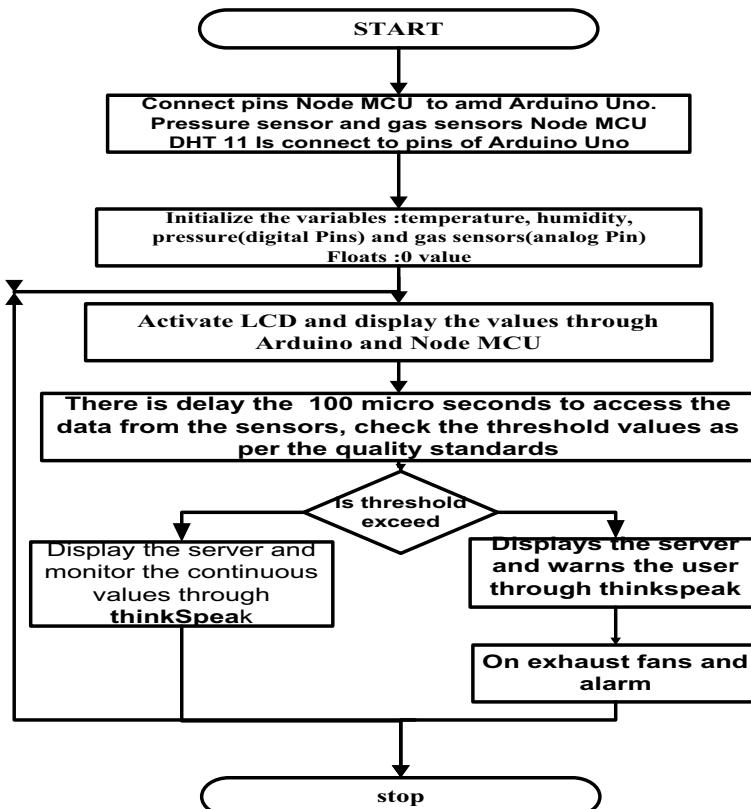


Fig. 2 Flowchart: programming for the system

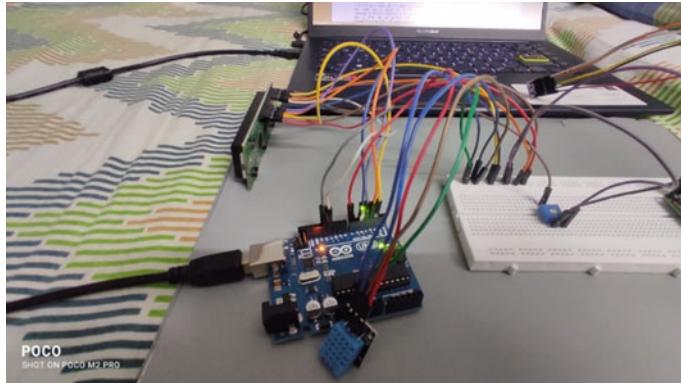


Fig. 3 Connections of Arduino with LCD

Fig. 4 Temperature and humidity display on LCD



- As shown in the picture, 4 RS pins to pin 12, pin 11, pin 4 to digital pin 5, D5 to digital pin, D6 to pin 2, R7 W pin to ground, VSS pin to ground, VCC to 5 pin V, resistor from 10 kΩ to +5 V and output to ground and dehumidifier VO PIN (Fig. 4).

The DHT11 info buttons connect to the digital pin of the MCU node, and the MQ135 connects to the analog pin as shown in Fig. 5.

4.1 Process Code Execution Process

Programs are written using Arduino software (IDE) as shown in the figure are called sketches. The editor has the ability to cut/paste and find/modify text. The scope of the message provides feedback on recording and exporting and also indicates errors.

The console displays text results through the Arduino software (IDE), including error messages and other information. The configuration board and serial port are displayed in the lower right corner of the window. The toolbar button allows us to test and load applications, sketch, open and save, and open serial monitors [24] (Fig. 6).

The complete proposed work execution of the program is shown in Fig. 7

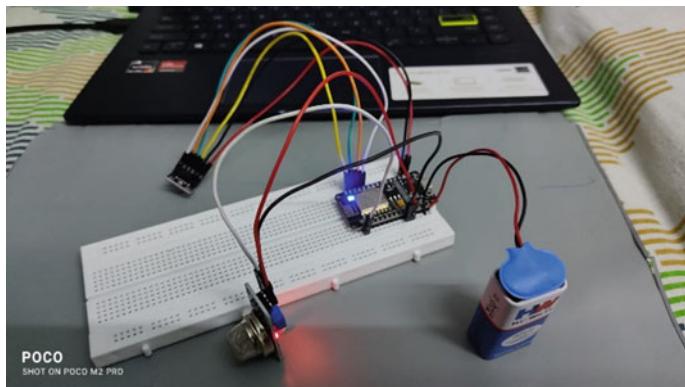


Fig. 5 Connection of DHT11 and MQ135 with NodeMCU

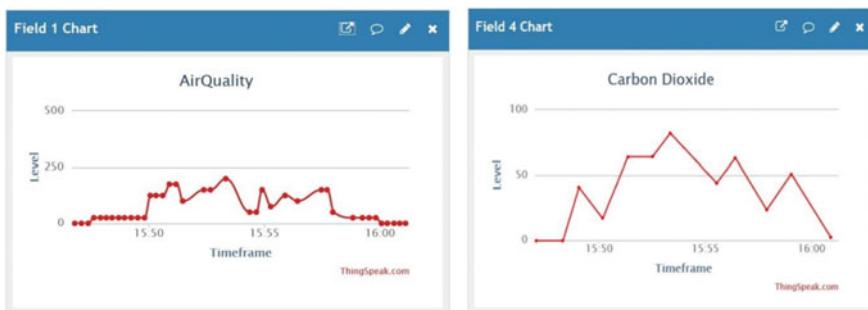


Fig. 6 ThingSpeak cloud server data display

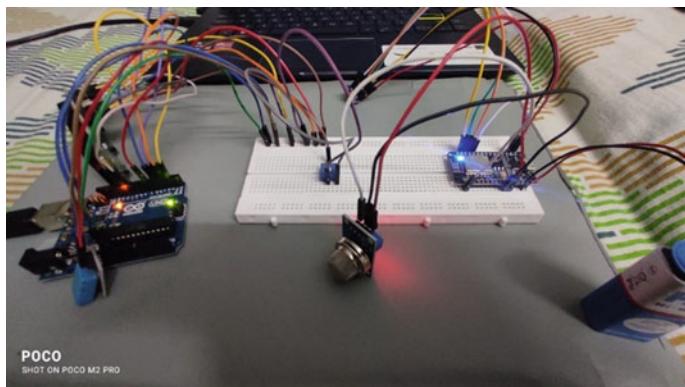


Fig. 7 Execution state of the proposed work

The proposed work is well-advanced, and the sensors are easily accessible, simple, compact, and convenient to use. The sensors have a longer lifespan and are less expensive. Air quality can be checked both indoors and outdoors and determines a wide range of physical parameters, including carbon dioxide. The proposed work will be implemented in several directions, depending on the type of environment. This is indoor air quality monitoring. Monitoring of industrial parameters and roadside pollution monitoring is available to the common man [24].

5 Device Impact

In this chapter, we have discussed the different impacts of the device in different aspects. Besides this we have also discussed cost analysis and economic impact of this device.

5.1 *Health and Safety Impact*

Our prototype model is designed in a way, so that it can be helpful to people with safety issues. In recent days, in our society gas leakage is a common incident nowadays. Many people lose their valuable lives because of this. For this reason, we have designed a prototype model of air pollution monitoring system where our system will automatically detect harmful gas presence. After detecting, our system will send the SMS by using ThinkSpeak. So we can see, for health and safety issues our system can bring a very powerful impact on society [11].

5.2 *Lower Energy Consumption*

Another major problem of our society is wastage of energy, and in recent days, this problem is becoming more acute. As the fastest developing country, many factories and mega malls are being established, and as a result, we are facing the energy problem. To reduce these kinds of problem, many agendas and project works are developing but still this problem has not been solved properly. So we all should come forward in order to solve excess energy consumption; otherwise in near future, we will face more energy issues. As it is designed for rooms and open areas, so it can be installed in home, office, schools, colleges, and many more places easily. It only consumes 5 V power supply. As a result, huge amount of energy and electricity can be saved [15].

5.3 Cost Analysis and Economic Impact

There are many systems related to air pollution monitoring system available in the market but most of them are highly priced and not affordable for every person. But our system is an integrated system which has many important features and its easily affordable for any person. Because it only costs under **2000 Rupees**, most importantly our system can be installed in many places, like room of homes, office rooms, schools, colleges, or university classroom. So it is really an additional plus point for our prototype model that it can perform multiple important operations at a time as well as it costs very less than the air pollution monitoring system available in market [16–20].

6 Conclusions

The proposed IoT-based air pollution system consists of a hardware that include Arduino and Node MCU with various types of sensors. The proposed system implements a combination of Android applications, server, and gas sensor (CO_2 and DHT11) to detect ambient air quality and display the actual air condition. Detects parameters causing contamination, each time the parameter level increases, the sensor detects the situation and sends a message to the authorized person via the Node MCU. Authorities can close down the power in polluting area via the Internet of Things. Sensors synthesize this information and send it to the web for real-time monitoring. This allows you to detect air pollution in different areas and take steps to prevent it. There is also a temperature sensor that estimates the room temperature. The proposed system is accurate and dynamically provides current situation at any place to check the air quality. The proposed system can be installed in every industry to monitor the various types of the emissions check the hazards for human. The system is also implemented in the home to check the air quality.

References

1. Alonso M, Finn EJ (1967) Fundamental university physics, vol 2. Addison-Wesley Reading
2. Continental Automated Buildings Association (2013) Impact of smart grid on connected homes landmark research study. In: Impact of smart grid on connected homes, pp 1–106
3. Cui Y, Kim M, Gu Y, Jin Jung J, Lee H (2014) Home appliance management system for monitoring digitized using cloud computing technology in ubiquitous sensor network environment. Int J Distrib Sens Netw 10(2):174097
4. Deepa K, Kumar PA, Krishna VS, Rao PNK, Mounika A, Medhini D (2017) A study of comparative analysis of different PWM techniques. In: 2017 international conference on smart technologies for smart nation (SmartTechCon), pp 1144–1149
5. Erfani S, Ahmadi M, Chen L (2017) The internet of things for smart homes: an example. In: 2017 8th annual industrial automation and electromechanical engineering conference (IEMECON), pp 153–157

6. Fernandes R, Matos JN, Varum T, Pinho P (2014) Wi-fi intruder detection. In: 2014 IEEE conference on wireless sensors (ICWiSE), pp 96–99
7. Gabriele T, Pantoli L, Stornelli V, Chiulli D, Muttillo M (2015) Smart power management system for home appliances and wellness based on wireless sensors network and mobile technology. In: 2015 XVIII AISEM annual conference, pp 1–4
8. Gu H, Diao Y, Liu W, Zhang, X (2011) The design of smart home platform based on cloud computing. In: Proceedings of 2011 international conference on electronic mechanical engineering and information technology, vol 8, pp 3919–3922
9. Han J, Jeon Y, Kim J (2015) Security considerations for secure and trustworthy smart home system in the IoT environment. In: 2015 international conference on information and communications technology convergence (ICTC), pp 1116–1118
10. Haque SMA, Kamruzzaman SM, Islam MA (2010) A system for smart home control of appliances based on timer and speech interaction. CoRR, abs/1009.4992
11. Hasan N, Khan AAM, Uddin N, Mitul AF (2013) Design and implementation of touchscreen and remote control based home automation system. In: 2013 2nd international conference on advances in electrical engineering (ICAEE), pp 347–352
12. Hassan MA, Abdullah AR, Bahari N, Jidin A (2013) Incorporating brushless DC motor in outdoor fan control of low voltage air-conditioning system. In: 2013 IEEE student conference on research and development, pp 550–555
13. Huaiyu X, Ruidan S, Xiaoyu H, Qing N (2009) Remote control system design based on web server for digital home. In: 2009 ninth international conference on hybrid intelligent systems, vol 2, pp 457–461
14. Huang L-C, Chang H-C, Chen C-C, Kuo C-C (2011) A zigbee-based monitoring and protection system for building electrical safety. Energy Build 43(6):1418–1426
15. Ikuta K, Kage H, Seki M, Hirai T (2011) Human intruder detection with leaky coaxial cables for wide area surveillance system. In: 2011 IEEE international conference on systems, man, and cybernetics, pp 103–108
16. Keshamoni K, Hemanth S (2017) Smart gas level monitoring, booking and gas leakage detector over IoT. In: 2017 IEEE 7th international advance computing conference (IACC), pp 330–332
17. Lei Z, Hu W, Li H, Yang Z (2013) Web-based remote networked control for smart homes. In: Proceedings of the 32nd Chinese control conference, pp 6567–6571
18. Lin H (2012) The development of control and energy usage information modules for smart homes. In: 2012 international conference on control, automation and information sciences (ICCAIS), pp 236–240
19. Liu L, Liu Y, Wang L, Zomaya A, Hu S (2015) Economical and balanced energy usage in the smart home infrastructure: a tutorial and new results. IEEE Trans Emerg Top Comput 3(4):556–570
20. Longe OM, Ouahada K, Rimer S, Zhu H, Ferreira HC (2015) Effective energy consumption scheduling in smart homes. In: AFRICON 2015, pp 1–5
21. Mao X, Li K, Zhang Z, Liang J (2017) Design and implementation of a new smart home control system based on internet of things. In: 2017 international smart cities conference (ISC2), pp 1–5
22. Nag D, Majumder D, Raquib CM, Pramanik S, Basu A, Rana TK, Rana B (2017) Green energy powered smart healthy home. In: 2017 8th annual industrial automation and electromechanical engineering conference (IMEECON), pp 47–51
23. Nguyen TV, Lee DG, Seol YH, Yu MH, Choi D (2007) Ubiquitous access to home appliance control system using infrared ray and power line communication. In: 2007 3rd IEEE/IFIP international conference in Central Asia on internet, pp 1–4
24. Panwar A, Singh A, Kumawat R, Jaidka S, Garg K (2017) Eyrie smart home automation using internet of things. In: 2017 computing conference, pp 1368–1370
25. Quwaider M (2017) Real-time intruder surveillance using low-cost remote wireless sensors. In: 2017 8th international conference on information and communication systems (ICICS), pp 194–199

26. Sahoo SK, Ramulu A, Batta S, Duggal S (2012) Performance analysis and simulation of three phase voltage source inverter using basic PWM techniques. In: IET Chennai 3rd international on sustainable energy and intelligent systems (SEISCON 2012), pp 1–7

Discrete Wavelet Transform-Based Fusion of Mammogram Images for Contrast Improvement



Anushka Dikshit, Vikrant Bhateja, and Ashruti Rai

Abstract Enhancement increases accuracy and efficiency of radiologist interpretation of mammogram images. However, most modalities exhibit inadequate contrast, weak tissue boundaries, etc. Thus, in this paper, contrast limited adaptive histogram equalization (CLAHE) is done to improve contrast of mammographic images in order to ease the further processing and get accurate results. In most of the cases, the acquired enhanced image lacks some important features of the original mammogram which gets removed as noise by the enhancement technique leading to contradictions in the interpretation of the mammographic image. To extract better features from the mammogram images, discrete wavelet transform-based image fusion technique is implemented which fuses the original and the enhanced mammographic image together, leading to restoration of features of both the mammographic images. The performance is evaluated using dedicated metrics such as enhancement measure (EME) and entropy (E).

Keywords CLAHE · Discrete wavelet transform · Image fusion · EME

1 Introduction

The second most fatal disease in women around the world after lung cancer is breast cancer [1]. In India, out of all women cancer patients in general, 14% are of breast cancer; based on the said statistic, 1 out of 28 women in India suffers by this disease during her lifetime. Most of the traditional techniques have failed to recognize the breast cancer at early stages [2]. The development of breast cancer occurs due to excessive growth of cells/tissue in a woman's breast at a rapid pace, forming a tumor. A tumor if cancerous in nature can be malignant which implies to rapid multiplication

A. Dikshit · V. Bhateja (✉) · A. Rai

Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM), Tiwari Ganj, Faizabad Road, Lucknow, Uttar Pradesh 226028, India

Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh 226031, India

of cells which spreads to other parts of the body if left untreated or a benign tumor in which cells multiply, but it will not affect surrounding tissues by spreading. Therefore, early detection of breast cancer can increase the chance of survival rate and increases the treatment option [2]. Accurate diagnosis and routine monitoring suggested for the cancer patients are obstructed by various factors such as the brightness, contrast, or noise present in the obtained images. Therefore, approach of computer-aided detection and diagnosis using mammogram images has been preferred for breast cancer which assists radiologists' interpretation by means of medical image analysis due to enhanced detection and evaluation of complex imaging features [1]. Therefore, various contrast enhancement techniques have been implemented such as HVS technique which incorporates the contrast manipulation algorithm to evaluate the degree of enhancement [3]. CLAHE has also been used for the purpose of improving the image contrast and is said to be suitable for medical images. Real-time image enhancement has also been carried out using CLAHE proving its suitability for VLSI and FPGA implementation [4]. The aim of enhancement is to improve the contrast of the image by removing the noise. Sometimes, important features of the original image get identified as noise and get removed in this process. Thus, in order to restore these features, image fusion has been a great technique and is being used extensively. Image fusion based on PSO and genetic algorithm has been able to provide suitable results especially for medical images [5, 6]. Also, a local entropy maximization-based image fusion that used incorporating discrete wavelet transform has given good results for mammogram images [2]. Combination of wavelet analysis and nonlinear enhancement functions has been used to find minute calcifications in the mammogram images effectively along with various applications in using extended spatial frequency [7, 8]. Fusion techniques involving PCA employed on contourlet and curvelet domains [9–12]. The paper proposes to utilize the discrete wavelet transform-based image fusion on the original mammogram and the CLAHE-enhanced mammogram using maximum and average fusion rules. The paper is organized into following sections: Sect. 2 describes the general overview of CLAHE and discrete wavelet transform along with proposed methodology; Sect. 3 summarizes the results, performance metrics and discusses the outcomes of this methodology. Section 4 summarizes the conclusion of the proposed work.

2 Proposed Design Methodology

2.1 Discrete Wavelet Transform

Wavelet transforms are a set of multi-resolution image decomposition tools which provide a variety of features of the image in different sub-bands of frequencies at multi-scale. The discrete wavelet transform works by converting the image into approximate coefficient (low-frequency component of the image) and detail coefficient (high-frequency component of the image); i.e., the transform basically converts

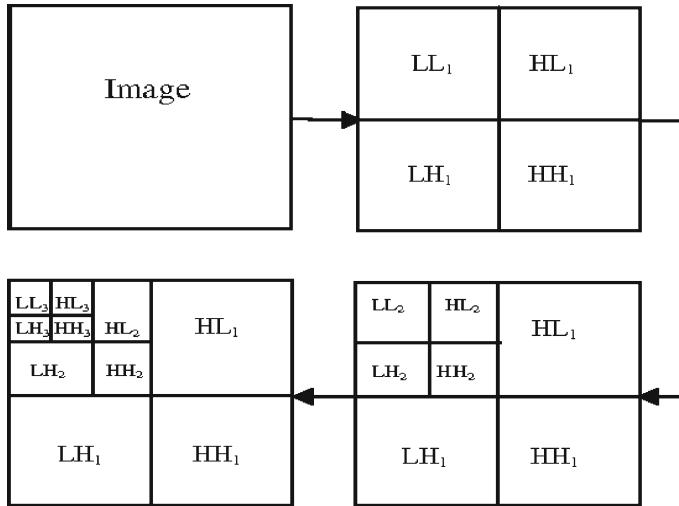


Fig. 1 Three-level decomposition in discrete wavelet transform [13]

an image from spatial domain to frequency domain as shown in Fig. 1 where the horizontal and vertical lines show the level of decomposition, and high- and low-frequency bands are represented as L and H [3].

2.2 Fusion Rules

The fusion of the mammogram is performed by fusion of the coefficients obtained by the combining approximate coefficients of the original and the CLAHE mammogram together and the detail coefficients of the original and CLAHE mammogram. The output fused image is obtained by applying inverse discrete wavelet transform on the fused mammogram. The fusion process in this paper has been performed using two techniques discussed in the subsections to follow.

2.2.1 Average Rule-Based Image Fusion

The average rule is the simplest approach in the fusion process in which the output pixel intensity from the resultant fused image is obtained by taking the average of the pixel intensity of the input images, i.e.,

$$F(i, j) = A(i, j) + B(i, j)/2 \quad (1)$$

where $F(i, j)$ is output image and $A(i, j)$ and $B(i, j)$ are the input images [14, 15].

2.2.2 Maximum Rule-based Image Fusion

The maximum rule is based on selection of a pixel having maximum intensity among the other set of pixels belonging to the considered input image, i.e.,

$$F(i, j) = \sum_{i=0}^m \sum_{j=0}^n \max A(i, j) B(i, j) \quad (2)$$

where $F(i, j)$ is output image and $A(i, j)$ and $B(i, j)$ are the input images [14, 15].

2.3 Proposed Fusion Approach

The working of the proposed approach is illustrated in the block diagram in Fig. 2.

The mammogram image is taken from the Digital Database of Screening Mammography, containing 2D mammogram images [16]. The input image (A) considered is enhanced using the CLAHE transform (B). The CLAHE transform operates by converting the mammogram images into small regions or tiles. The pixels covered by these tiles are clipped above the threshold and redistributed over

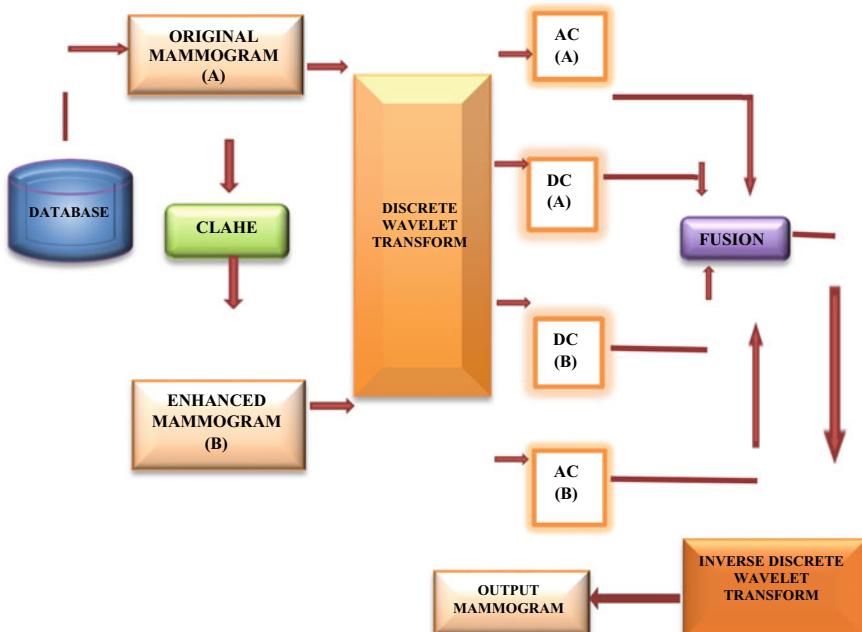


Fig. 2 Block diagram of proposed fusion algorithm

the histogram for histogram equalization to take place. In the final step, these tiles are recombined using binary interpolation to remove artificial boundaries. Discrete wavelet transform is now applied on the original and the CLAHE-enhanced image of the same mammogram to obtain the approximate coefficients (AC) and detail coefficient (DC). The fusion process will take place as a result of combining DC of the original and CLAHE-enhanced image and the AC of the original and CLAHE-enhanced mammogram, respectively, using the desired fusion rule. The output mammogram image will be obtained by applying inverse wavelet transform on the fused coefficients.

The proposed procedure of mammogram fusion is summarized in the algorithm below:

Algorithm 1: Procedural Steps for Proposed Fusion Approach

BEGIN

Step 1: *Input* test image A.

Step 2: *Convert* A from RGB to Grayscale to obtain image B.

Step 3: *Apply* CLAHE transform on image B.

Step 4: *Sub-band* decomposition of images A and B using DWT.

Step 5: *Fuse* the decomposed level by using Max. Fusion rule.

Step 6: *Fuse* the decomposed level by using Average Fusion rule.

Step 7: *Output* fused images from step 5 and 6 are obtained using Inverse DWT.

END

3 Results and Discussions

The enhancement of an image is basically the modification in the pixel values. But, due to the amplification of the noise also thus, we cannot rely solely on the contrast improvement to assess the enhanced mammogram. The evaluation is, therefore, done using dedicated metrics such as enhancement measure (EME) and absolute mean brightness error (AMBE) which is the difference between the brightness of the enhanced and the original image. For assessing the fusion process, the metrics used are entropy (E) which gives the information content of the image and standard deviation which measures the dispersion of image gray-level intensity [17]. The mammogram images used in this paper are acquired in the 2D format from the Digital Database of Screening Mammography (DDSM) [16]. The mammographic image is converted from RGB to grayscale format first in order to apply CLAHE on it. Then, for restoration of features, the input and the CLAHE-enhanced mammograms are fused together using the techniques mentioned in Sects. 2.2.1 and 2.2.2 separately. The results are evaluated along with the output images shown in Fig. 3.

The quality assessment of images has been conducted using entropy and EME described in Sect. 3, and their values are tabulated in Table 1.

Figure 3 shows two test images, i.e., C_0006_1. RIGHT_CC and C_0019_1. RIGHT_MLO, respectively, along with their CLAHE-enhanced and fused outputs.

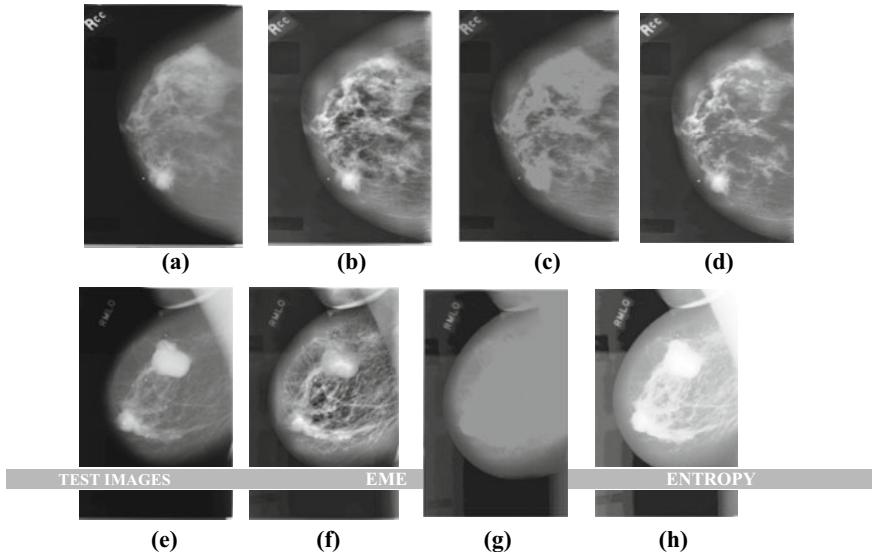


Fig. 3 **a** Original mammogram image C_0006_1. RIGHT_CC **b** CLAHE **c** Average fusion rule **d** Max. fusion rule **e** Original mammogram image C_0019_1. RIGHT_CC **f** CLAHE **g** Average fusion rule **h** Max. fusion rule

Table 1 Values of EME and entropy for proposed approach

Test images	EME				ENTROPY			
	Original	CLAHE	Max. rule	Average rule	Original	CLAHE	Max. rule	Average rule
C_0006_1. RIGHT_CC	2.1841	2.0602	1.7355	1.7291	6.2634	6.9107	6.8647	6.0868
C_0011_1. RIGHT_CC	1.5795	2.5677	2.9437	1.0749	5.3903	6.9907	6.0663	5.2916
C_0019_1. RIGHT_MLO	2.4067	2.9847	2.1403	2.1167	6.1792	6.3211	6.8800	5.4758
D_4084_1. RIGHT_MLO	2.9893	2.3623	4.7952	4.5968	6.5118	7.2882	6.2232	6.1641

The first image has an ill-defined dense mass of benign (non-cancerous) category of tumor. Upon CLAHE enhancement, the tumor is visible in the lower part of the mammogram. On applying the fusion rule, the visibility of the tumor is better visible on the max. fusion rule as compared to the average fusion rule. The second image consists of a circumscribed mass, i.e., a mass with well-defined structure of malignant category (cancerous) which upon CLAHE enhancement is visible as two blob-like structures. On using both the fusion rules, it shows that the max. fusion rule shows better performance as compared to the average fusion rule in terms of visuality of the

tumor. Table 1 shows incrementing values of entropy and EME on applying CLAHE enhancement and image fusion using maximum fusion rule.

4 Conclusions

In this paper, discrete wavelet transform-based fusion technique is proposed to fuse the mammogram image with an aim to restore the essential features of the mammogram image which get eradicated as noise during the enhancement procedure. The enhancement has been carried out using the advanced approach of histogram equalization, i.e., CLAHE. The paper also discusses about the maximum and average fusion rules and highlights the motive of using the maximum fusion rule as it only considers the pixels having high pixel intensity, thus reducing the creation of artificial boundaries during the fusion rules. The proposed approach is implemented on the sample images taken from DDSM database and assessed using dedicated metrics, i.e., EME and entropy which indicate that the maximum fusion rule gave better results as compared to average fusion rule. Thus, it can be concluded that the essential features of the original mammogram images can be preserved using the proposed approach which may further assist in efficient feature extraction and classification techniques improving the overall efficiency of the radiologist diagnosis and treatment.

References

1. Bajaj V, Pawar M, Meena VK, Kumar M, Sengur A, Guo Y (2017) Computer-aided diagnosis of breast cancer using bi-dimensional empirical mode decomposition. *Neural Comput Appl* 31:3307–3315
2. Pawar MM, Talbar SN (2018) Local entropy maximisation based image fusion for contrast enhancement of mammogram. *J King Saud Univ Comput Inf* 23:1190–1197
3. Trivedi M, Jaiswal A, Bhateja V (2008) A novel HVS based image contrast measurement index, vol 1(3). Springer, Berlin, pp 685–690
4. Reza AM (2003) Realization of the contrast limited adaptive histogram equalisation (CLAHE) for real time image enhancement. *J Image Process* 38(6):15–34
5. Kumar MP, Kumar PR (2016) Image fusion of mammography images using meta heuristic method particle swarm optimization (PSO). *Int J Appl Eng Res* 11:6254–6258
6. Kumar MP, Kumar PR (2019) Enhancing bio-medical mammography image fusion using optimized genetic algorithm. *J Med Imag Health Inf* 9(3):502–507
7. Bhateja V, Devi S (2009) Combination of wavelet analysis and non-linear enhancement function for computer aided detection of microcalcification. *Int J Appl Eng Res* 9(13):645–649
8. Zheng Y, Essock EA, Hansen BC, Haun AM (2005) A new metric based on extended spatial frequency and its application to DWT based image fusion algorithms. *Image Process Rem Sens* 8:117–192
9. Himanshi, Bhateja V, Krish A, Sahu A (2014) An improved medical image fusion approach using PCA and complex wavelets. *Int Conf Med Imag* 14(3):978–4799
10. Himanshi, Bhateja V, Krishn A, Sahu A (2015) Medical image fusion in curvelet domain employing PCA and maximum selection rule. *Adv Intell Syst Comput* 379(3):1–9

11. Bhateja V, Urooj S, Pandey A, Misra M, Lay-Ekuakille A (2013) Improvement of masses detection in digital mammograms employing non-linear filtering. In: Proceedings of IEEE international multi-conference on automation, computing and control, vol 32(119), pp 406–408
12. Bhateja V, Srivastava A, Moin A, Le D (2011) Multispectral medical image fusion in Contourlet domain for computer based diagnosis of Alzheimer disease. *Rev Sci Instrum* 5(19):341–346
13. Giri K, Peer MA, Nagabhushan P (2014) A robust color image watermarking scheme using discrete wavelet transform. *Int J Image Graph Sig Process* 7(1):47–52
14. Rani K, Sharma R (2005) Study of different image fusion algorithms. *Int J Emerg Technol Adv Eng* 8:177–192
15. Sruthy S, Parameshwaram L, Sasi AP (2013) Image fusion technique using DT-CWT. *Int J Appl Eng Res* 13:984–988
16. Digital Database for Screening Mammography (DDSM) images available at: <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>. Last Updated on 14 Sept 2017
17. Trivedi M, Jaiswal A, Bhateja V (2013) No reference image quality index for contrast and sharpness measurement. In: 3rd IEEE international advance computing conference (IACC), India, pp 1234–1239

Assessment of Land Degradation Dynamics Using Spectral Angle Mapper Method and Demographic Analytics



Amar Kumar Kathwas and Rakesh Saur

Abstract Landuse landcover (LULC) dynamics demonstrate and forms relationship between human-induced development and biodiversity conservation. With the ever-rising human population and developments, the occurrence of LULC transformations is significantly high, globally. The LULC dynamics induced due to human developments had drastic negative impacts on climate, hydrogeomorphology, flora and fauna and earth system processes. Therefore, an understanding of LULC drivers is requisite for formulation of environmental policies, aimed to build synergistic between human and ecology for environmental sustainability. Considering the above rationale, the present study is envisioned to investigate the impact of LULC transformation induced due to human development and changing demography. The task, delineation of LULC categories, was accomplished using Spectral Angle Mapper (SAM) technique, and the impact of demography on LULC dynamics was studied utilizing landscape fragmentation metrics. The findings of the study are astute for devising environmental policies and management strategies for sustainable development of the region.

Keywords Spectral angle mapper · Landuse landcover · Demography · Landscape fragmentation metrics

1 Introduction

Sustainability can be defined as the developments and actions induced by humankind for evading diminution of natural resources for maintaining ecological balance. The notion of environmental dynamics evolved from apprehension of natural resource status altered by human activities at a swift pace in recent decades. According to the United Nations, it took 150 years (1750–1900) for the world population to triple from 0.7 to 2.5 billion, whereas it took only 60 years (1950–2010) to triple again to

A. K. Kathwas (✉)

Haryana Space Applications Centre, CCS Haryana Agricultural University, Hisar, Haryana, India

R. Saur

Department of Remote Sensing, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India

more than 7.25 billion [1]. This upsurge in human population can be accredited to intensified resource consumption and, can be manifested through agriculture, food production and urbanization intensification. These developments eventually led to transformation of existing landuse and landcover (LULC) with negative impacts on the natural balance of climate, hydrology, biodiversity and various earth system resources and processes [1, 2]. Since, sustainability is the prime principle behind formulation of environmental policies and decision-making, LULC emerged globally as central research theme for assessing its role and impact on environment at hyperlocal to global scale.

Investigation of LULC for development of management strategies and future LULC dynamics concerning environmental sustainability for ameliorating living standard of humans requires insightful of LULC drivers. [3] outlined population growth, urbanization and pasturing as one of prime drivers of LULC transformation, globally. Moreover, natural hazards, viz. climate change, drought, soil erosion, rainfall variability, flood and fire are also identified as one of the major drivers [4–6]. Conception of these drivers is prime requisite for modeling LULC dynamics and development of management strategies and policy-making. Furthermore, various researchers integrated social and ecological data to investigate their impact on LULC dynamics. [7] assessed the relationship and impact of tribal population with LULC dynamics in tribal dominated Ranchi district of Jharkhand state India. The study revealed the acute dependency of various tribes on natural vegetation cover for their livelihood and survival which subsequently resulted in natural vegetation cover conservation and expansion. [8] studied the impact of urban expansion on LULC and environment. The study unveiled the severe negative impacts of urbanization on natural resources. [9] showed austere impacts of mining-induced LULC dynamics on human living standard, environment and biodiversity of the region. Furthermore, Patel and Kathwas 2011 exhibited the impact of LULC dynamics on the natural surface process, viz. soil erosion.

Monitoring and assessment of LULC change requires timely and accurate information. The emergence and advancement of remote sensing and Geographic Information System (GIS) proved out as powerful tools for assessing LULC dynamics over large areas. GIS provides a flexible platform for collecting, storing, displaying and analyzing digital datasets, whereas remote sensing provides satellite imageries with synoptic view of earth surface, an important data resource of GIS. Various geospatial techniques (image classification, differentiation, band ratioing, on-screen digitization, image regression, temporal analysis, principal component analysis) developed are present in literatures for monitoring and assessment of LULC. Studies showed that post-classification techniques provide significantly accurate information and nature of changes.

Here, in the present study, the primary aim is to assess the LULC dynamics and its impact relation concerning environmental sustainability. In order to achieve the desired objective, Spectral Angle Mapper (SAM) algorithm was implemented for classification of satellite imageries of respective time-periods. The study was carried out for Tinsukia district located in eastern state of India, Assam. The major LULC dynamics drivers considered in the present study involves demography along with

landscape fragmentation indices. The present research will help in determining the impact and magnitude of urban center expansion, population outburst and human-induced landscape fragmentation on LULC and environment. Moreover, it will also provide an insight for formulation of problem-specific management strategies for sustainable resource development of the region.

2 Study Area

The study area (Fig. 1) Tinsukia district is located in the eastern-most part of Assam state of India, positioned between $27^{\circ} 14' 03''$ and $27^{\circ} 48' 05''$ north latitudes and $95^{\circ} 13' 30''$ and $96^{\circ} 00' 00''$ east longitudes. The district encompasses an approximate area of 3790 sq. km. of Brahmaputra River basin flowing NE–SW direction. The study area experiences sub-tropical humid climate endowing with high-intensity rainfall all year around. Out of 2323 mm yearly rainfall, 65% is received during the monsoon months of June to September. The summer season last for 5 months from May to September with temperature ranging from 25° Celsius to 32° Celsius, whereas the winter season last for 2 months from December to February with temperature varying from 9° Celsius to 22° Celsius. Physiographically, the study area can be characterized into plains of Brahmaputra and hills in southern part. The distinguishable geomorphic landforms identified in the study area are flood plain, structural hills and old

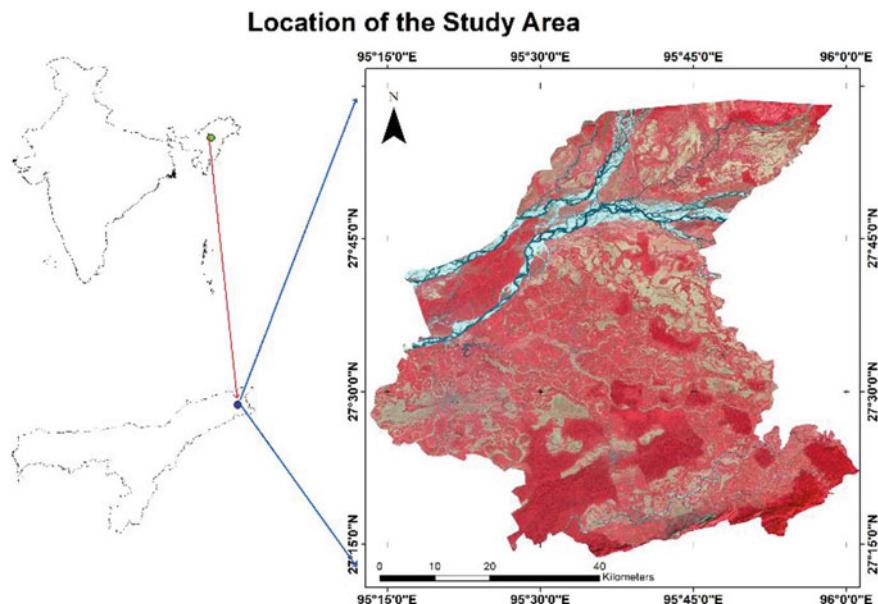


Fig. 1 Location of the study area

and young alluvium. Out of total district area, approximately 25.94% is cultivable followed by 27% under natural vegetation, 33% under orchard/plantation and 4% area under grassland. The primary cultivated crop is paddy and to small extent maize, wheat, pulses, pea and potato. Alfisols, Inceptisols, Entisols and Ultisols are prime identified soils.

3 Materials and Methods

The step-by-step flow of research procedure is depicted in Fig. 2. In order to accomplish the task and achieve the desired objectives at the onset cloud-free satellite imageries pertaining to year 2010 (Landsat 5 TM) and 2020 (Sentinel 2A) were download and atmospherically corrected followed by preprocessing to identify LULC features using visual interpretation keys

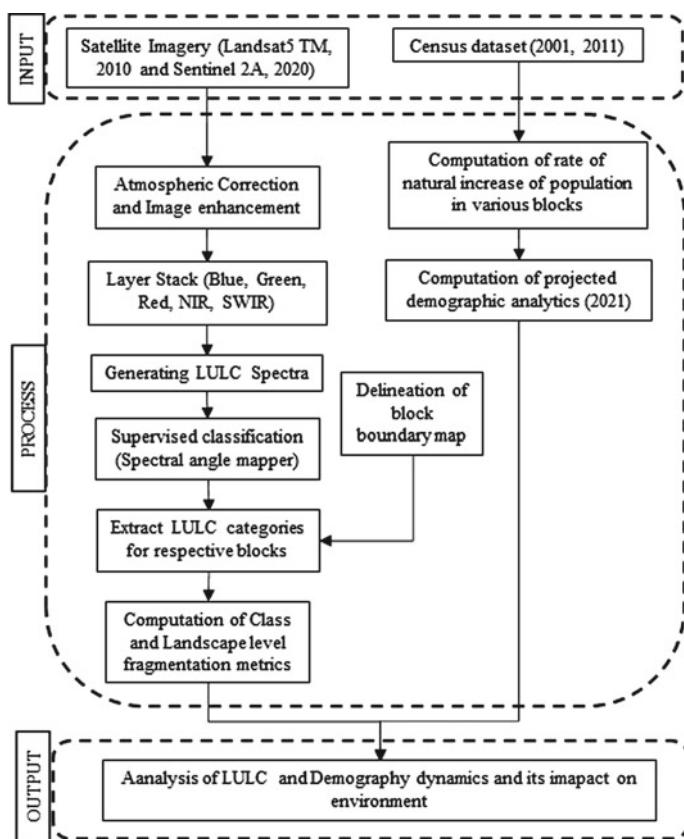


Fig. 2 Methodological flowchart of the present research

ENVI® 5.2 was used for processing, enhancement and classification of satellite imagery to delineate LULC features. The area of interest defining each LULC feature class was delineated using the ROI tool. LULC features, viz. natural vegetation, plantation/cropland, fallow/degraded land, grassland, waterbodies/river, built-up and sand bars were considered for present investigation.

The task of satellite imagery classification was carried out using supervised classification technique involving SAM algorithm, a spectral signature-based technique which utilizes spectral signature in n-dimensional space (n represents number of spectral bands) to match DN values of pixels with feature spectra. Smaller the angles between n-dimensional vectors, higher the similarity and vice-versa, respectively. The algorithm is significantly insensitive toward illumination conditions and albedo effects. The pixel values were used to extract LULC spectra and were saved in spectral library in ASCII format. The spectral angle between n-dimensional vector is calculated using the following equation,

$$\theta = \cos^{-1} \left(\frac{\sum_{i=1}^n t_i r_i}{\sqrt{\sum_{i=1}^n t_i^2 \sum_{i=1}^n r_i^2}} \right) \quad (1)$$

where n represents number of spectral bands, r is reflectance of reference spectra and t denotes reflectance of actual spectra.

The assessment of LULC dynamics was carried out based on change in spatial extent of LULC features and magnitude of landscape and feature class fragmentation, a consequence induced due to human-oriented developments. Fragstats v4.2.1 software was used for the computation of various landscape and class-level fragmentation metrics at block scale for assessing the impact and magnitude of human-induced LULC transformations. The various landscape metrics utilized in present study are presented in Table 1. Furthermore, block-level demographic dataset (total population, total male and female population) was downloaded (<https://censusindia.gov.in/DigitalLibrary/Tables.aspx>) and taken into consideration for assessing the impact of demography on LULC dynamics and environmental sustainability.

The block layer of the study area was prepared from block map obtained from (<http://asdma.gov.in>) in Geographic Information System (GIS) environment using ArcGIS® 10.6 software. Since, the assessment has been carried out for 2010 and 2020 time-period and, the demography dataset pertaining to year 2020 was unavailable therefore, using a basic statistics-based population projection equation, the population count corresponding to year 2021 was projected. The formula used for computation of projected population corresponding to year 2020 can be represented as

$$N_t = P \log_e r * t \quad (2)$$

where N_t represents population at future time-period, P is present population, r is rate of natural increase/100, and t is time-period.

Table 1 Description of landscape and class metrics utilized in the present study

S. No	Landscape/class metrics information	Equation
1	Number of patch (NP): The metrics depicts the total number of patches in the landscape/class	$NP = N$ $N = \text{total number of patches in the landscape}$
2	Patch density (PD): It shows the number of patches per unit area	$PD = \frac{N}{A} (10,000)(100)$ $N = \text{total number of patches in the landscape}$ $A = \text{total landscape area (m}^2\text{)}$
3	Largest patch index (<i>LPI</i>): The metric measures the most dominant patch in a landscape	$LPI = \frac{\max(a_{ij})}{A} (100)$ $a_{ij} = \text{area (m}^2\text{) of patch } a_{ij}$ $A = \text{total landscape area (m}^2\text{)}$
4	Landscape shape index (<i>LSI</i>): A measure of the total edge of the landscape	$LSI = \frac{0.25*E^*}{\sqrt{A}}$ $E^* = \text{total length of edge in the landscape}$, $A = \text{total landscape area (m}^2\text{)}$
5	Interspersion and juxtaposition index (<i>IJI</i>): The metrics signifies the intermixing of patch types in landscape	$IJI = \frac{-\sum_{i=1}^m \sum_{k=i+1}^m [\frac{e_{ik}}{E}] * \ln(\frac{e_{ik}}{E})}{\ln(0.5)(m(m-1))} (100)$ $e_{ik} = \text{total length of edge in landscape between patch type } i \text{ and } k$ $E = \text{total length of edge in the landscape}$ $m = \text{number of patch types present in the landscape}$
6	Shannon's diversity index (<i>SHDI</i>): a patch diversity evaluation metric in a landscape	$SHDI = -\sum_{i=1}^m (P_i * \ln P_i)$ $P_i = \text{Proportion of the landscape occupied by patch type } i,$

4 Results and Discussions

4.1 Analysis of Landuse Landcover Dynamics

Figures 3 and 4 depict the landuse landcover map of the study area corresponding to year 2010 and 2020, respectively, and Fig. 5 presents the statistics associated to LULC dynamics during the time-period.

The analysis of Fig. 4 reveals that during the 2010–2020 time-period, spatial extent of cropland (333.21 km^2) increased largely, whereas fallow land (81.10 km^2) followed by built-up (53.24 km^2) and sand bars (42.08 km^2) show moderate amount of increase in extent. On the contrary, the spatial extent of natural vegetation (-298.49 km^2) and grassland (-154.52 km^2) decreased largely whereas the extent of waterbodies (-56.63 km^2) decreased to moderate extent. The observation of increased extent of cropland and built-up can be attributed to increased population and expansion of urban centers during the time-period. Moreover, the reduction in the

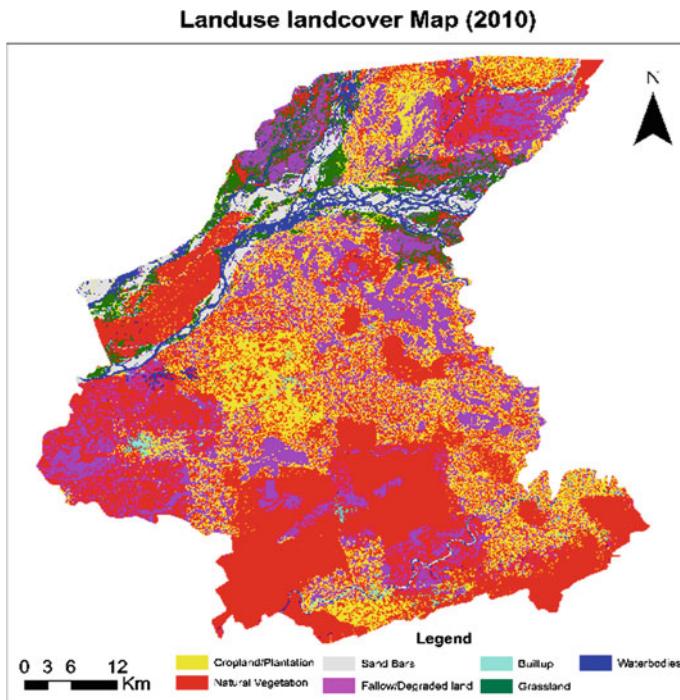


Fig. 3 Landuse landcover map of the study area corresponding to year 2010

natural vegetation cover extent can be ascribed to expansion of cropland category for increased food production to meet the end needs of rapidly increasing human population in the study area. Yearly floods in the Brahmaputra River basin brings large amount nutrient-rich soils and eroded materials which get deposited in flood plains and near river banks, resulting in significant expansion of spatial extent of sand bars. Since grasslands are primarily situated near river channels in the study area, transformation of grassland to sandbars due to yearly floods along with simultaneous conversion of grassland to cropland resulted in decline of grassland extent

The deposition of eroded materials is also the prime reason behind the change in river course on river bed which resulted in decrease in waterbody extent. From the above observations, it can be inferred that due to human development and natural hazards, there occurred significant amount of transformation among LULC categories in the study area. The significant amount of reduction in natural vegetation, grassland along with expansion of cropland and built-up has severe negative impact on human environment (deterioration of water, land and air quality) and biodiversity of the region. Moreover, the LULC dynamics has also significantly affected the land surface processes in the region, namely increased surface runoff and soil erosion [10].

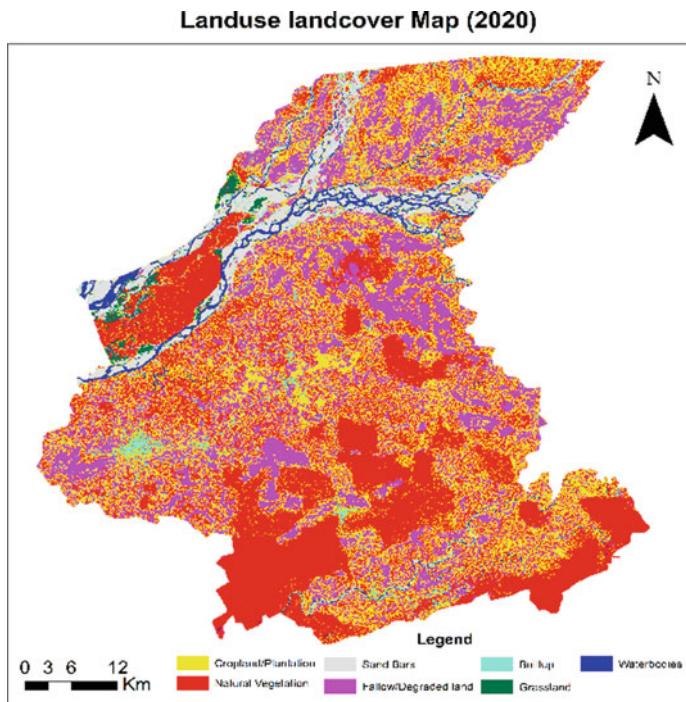


Fig. 4 Landuse landcover map of the study area corresponding to year 2020

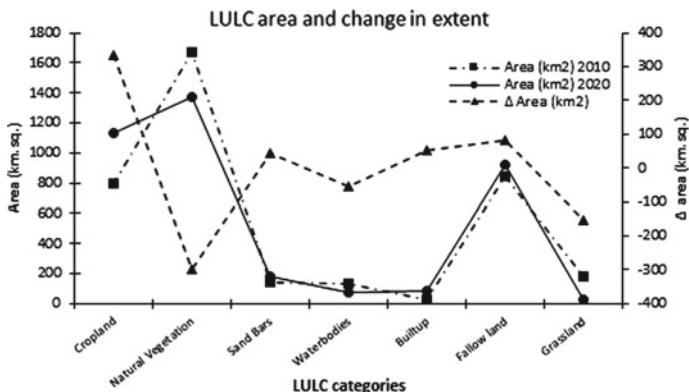


Fig. 5 Landuse landcover and change statistics corresponding to respective time-periods

Table 2 Landscape level fragmentation metrics values corresponding to different blocks in respective time-periods

	Doomdooma 2010	Doomdooma 2020	Mergharita 2010	Mergharita 2020	Sadiya 2010	Sadiya 2020	Tinsukia 2010	Tinsukia 2020
NP	12,037	13,823	13,553	17,622	7699	9188	6001	9779
PD	14.11	16.20	10.23	13.30	15.60	18.61	11.65	18.98
LPI	16.77	9.82	38.97	17.49	10.73	25.43	31.13	8.62
LSI	84.87	91.68	71.67	89.10	62.65	72.41	57.43	78.79
SHDI	1.30	1.23	0.98	1.07	1.43	1.26	1.11	1.20

4.2 Analysis of Landscape Fragmentation and Demography Dynamics

In order to assess the impact and role of human-induced developments on LULC transformation, landscape fragmentation analysis was carried out at block-level using the various landscape and class metrics. The computed landscape and class metrics values corresponding to individual blocks in respective time-periods is presented in Tables 2 and 3, respectively.

Furthermore, the block-level demographic statistics corresponding to 2001 and 2011 along with projected demography statistics for year 2021 is presented in Table 4. From the analysis of fragmentation metrics, it can be observed that the number of patches corresponding to various LULC categories increased significantly, with Mergharita block witnessing highest level of landscape fragmentation followed by Tinsukia, Doomdooma and Sadiya. Moreover, Tinsukia block witnessed highest amount of increase in patch density. Other than Sadiya block, all blocks exhibited a decrease in LPI values (Mergharita > Tinsukia > Doomdooma), whereas the total edge depicted by LSI shows an increasing trend from year 2010–2020 (Tinsukia > Mergharita > Sadiya > Doomdooma). Furthermore, the Sadiya (−18.95) block followed by Mergharita (−6.53) and Doomdooma (−0.91) demonstrates decrease in IJI values and Tinsukia block shows slight increase (0.15). The diversity of the landscape depicted by SHDI values shows decrease in diversification of Sadiya (−0.18) and Doomdooma (−0.07) block landscape. On the contrary, Tinsukia (0.095) and Mergharita (0.08) witnessed minuscule increase in the landscape diversity.

The positive trend of LPI in Sadiya block can be attributed to increase in cropland extent; the largest landuse category of the block, whereas negative trend of LPI values corresponding to other blocks can be corroborated to decreased natural vegetation cover, the largest landcover category.

Furthermore, the analysis of class metrics reveals that the spatial extent of natural vegetation decreased significantly in all the blocks. Moreover, reduction in spatial extent of grassland and waterbodies is also observed in Mergharita and Tinsukia block. The PD and LSI reveal increase in cropland built-up, natural vegetation and fallow land patches in respective blocks, and LPI values show decrease in natural

Table 3 Class-level landscape fragmentation metrics values corresponding to various blocks in respective time-periods

	Cropland 2010	Cropland 2020	Vegetation 2010	Vegetation 2020	Fallow 2010	Fallow 2020	Built-up 2010	Built-up 2020	Water bodies 2010	Water bodies 2020	Sand Bars 2010	Sand Bars 2020	Grassland 2010	Grassland 2020
<i>DOOMDOOMA</i>														
CA	27,678.05	29,082.82	29,669.88	26,727.93	23,631.61	26,963.81	344.06	739.61	1424.54	723.39	858.34	1092.9	1723.98	0
NP	3399	4132	3602	3788	3612	4244	490	1166	640	267	223	226	71	0
PLAND	32.44	34.08	34.77	31.32	27.69	31.60	0.40	0.87	1.67	0.85	1.01	1.28	2.02	0.00
PD	3.98	4.84	4.22	4.44	4.23	4.97	0.57	1.37	0.75	0.31	0.26	0.26	0.08	0.00
LPI	16.77	9.82	5.33	4.16	4.29	4.29	0.02	0.03	0.27	0.10	0.31	0.33	0.71	0.00
LSI	117.92	133.64	94.04	97.21	66.87	74.80	24.33	37.07	27.30	18.45	14.55	15.57	24.04	0.00
III	47.04	52.88	40.35	33.61	52.00	44.17	70.01	83.59	82.40	76.76	83.69	87.22	68.45	0.00
<i>MERGHARTA</i>														
CA	25,185.21	34,714.54	84,058.64	73,246.93	21,015.65	22,259.07	1039.74	1079.06	714.54	714.18	482.40	482.40	0.00	0.00
NP	3376	5593	3768	5061	4660	5159	1112	1173	422	421	215	215	0.00	0.00
PLAND	19.01	26.20	63.44	55.28	15.86	16.80	0.78	0.81	0.54	0.54	0.36	0.36	0.00	0.00
PD	2.55	4.22	2.84	3.82	3.52	3.89	0.84	0.89	0.32	0.32	0.16	0.16	0.00	0.00
LPI	8.10	10.38	38.97	17.49	0.83	0.83	0.10	0.10	0.04	0.04	0.02	0.02	0.00	0.00
LSI	120.35	150.65	64.17	78.40	77.35	82.25	38.83	39.76	27.44	27.41	17.50	17.50	0.00	0.00
III	48.98	51.10	49.89	30.07	50.83	42.69	90.51	85.03	82.66	80.21	96.73	89.25	0.00	0.00
<i>SADIYA</i>														
CA	12,013.64	19,876.79	14,134.16	9371.53	17,484.05	17,618.58	189.45	781.03	1272.32	536.56	679.60	1160.37	3589.05	17.41
NP	1207	1979	2469	3487	2300	2317	167	959	668	136	305	309	583	1
PLAND	24.34	40.27	28.63	18.99	35.42	35.69	0.38	1.58	2.58	1.09	1.38	2.35	7.27	0.04

(continued)

Table 3 (continued)

	Cropland 2010	Cropland 2020	Vegetation 2010	Vegetation 2020	Fallow 2010	Fallow 2020	Built-up 2010	Built-up 2020	Water bodies 2010	Water bodies 2020	Sand Bars 2010	Sand Bars 2020	Grassland 2010	Grassland 2020
PD	2.45	4.01	5.00	7.06	4.66	4.69	0.34	1.94	1.35	0.28	0.62	0.63	1.18	0.00
LPI	10.64	25.43	7.16	1.81	10.73	10.73	0.03	0.30	0.13	0.15	0.33	1.13	0.04	2.58
LSI	69.79	98.62	66.10	83.63	55.77	56.17	16.06	35.65	30.24	18.56	22.12	23.10	43.11	
IJI	51.79	52.66	61.99	25.11	70.14	36.54	83.55	81.05	91.22	72.72	95.88	76.33	59.67	8.95
<i>TINSUKIA</i>														
CA	9726.14	17,794.67	24,437.36	14,141.39	16,574.03	18,110.10	667.86	1371.93	64.46	51.76	59.86	0.00	0.00	
NP	1430	2739	1461	2840	2582	2844	329	1163	101	95	98	0.00	0.00	
PLAND	18.87	34.53	47.42	27.44	32.16	35.15	1.30	2.66	0.13	0.10	0.12	0.00	0.00	
PD	2.78	5.32	2.84	5.51	5.01	5.52	0.64	2.26	0.20	0.18	0.19	0.00	0.00	
LPI	9.39	8.62	31.13	3.12	4.74	4.74	0.76	0.96	0.01	0.01	0.01	0.00	0.00	
LSI	66.38	112.82	64.08	79.55	64.07	67.87	21.39	38.78	10.93	10.42	10.50	0.00	0.00	
IJI	51.66	55.65	49.74	36.19	43.85	44.74	71.23	61.46	75.12	73.63	81.01	76.70	0.00	0.00

Table 4 Demographic statistics of various blocks in respective time-periods

	Total population			Total working population			Main work population			Marginal work population			Non-work population		
	2001	2011	2021	2001	2011	2021	2001	2011	2021	2001	2011	2021	2001	2011	2021
Doomdooma	726,990	833,578	934,853	294,998	346,450	389,623	217,778	249,692	280,026	77,220	96,758	109,674	431,992	487,128	545,275
Margherita	639,770	741,350	832,431	256,454	307,070	346,130	198,636	229,828	258,019	57,818	77,242	88,282	383,316	434,280	486,378
Sadiya	184,024	204,868	228,993	91,338	110,006	124,086	62,916	64,774	71,798	28,422	45,232	53,035	92,686	94,862	105,085
Tinsukia	749,340	876,062	984,711	289,912	350,866	396,006	220,364	276,082	312,931	69,548	74,784	83,274	459,428	525,196	588,800

vegetation and cropland patch. The result obtained revealed large amount of transformation of natural vegetation cover to other LULC categories, mainly cropland and conversion of cropland to built-up and fallow. The rising population and unbalanced development have led to large amount deforestation of natural vegetation cover to timber production, mining and crop cultivation, whereas the unsustainable cultivation technique has resulted in increase in fallow/degraded land area. Furthermore, from the demographic statistics, it can be inferred that the population of the region will increase at substantiate rate; with a smaller number of working/marginal working population, the pressure on natural resources of the study area will increase drastically in near future time-periods. These developments will result in higher degree of landscape fragmentation and transformation of dominant LULC categories in the various blocks of the study area, especially in Doomdooma, Sadiya and Tinsukia blocks where transformation of natural vegetation and expansion of cropland is increasing at very swift pace along with built-up area expansion. This development will drive the landscape transformation to meet the end need of the human population. Therefore, in order to sustain the rising human population, formulation of rigorous environment policies is needed along with block-oriented management strategies for sustainability of natural resources and environment.

5 Conclusion

In recent time periods, transformation of natural vegetation, croplands and urban centers have largely been driven by providing food, shelter and fibers to all human population, globally. The developments being made in order to meet the goals of providing basic amenities to all are augmented by large amount of increase in natural resource consumption, energy and severe damage to flora and fauna all around the world. An understanding LULC dynamics, its transition, its drivers and risk help in formulation of specific management strategies and policies for synergistic development of human with sustainability of natural resources and environment. In the present study, an attempt has been made to assess the role and impact of rising human population on LULC and environment in a rapidly growing hilly landscape of Assam, India. The salient findings of the present investigation highlights:

1. The increasing population and resource consumption in the study area resulted in severe reduction in spatial extent of natural vegetation and grassland, along with higher amount of increase in cropland and fallow land, posing threat to local biodiversity, climate and environment.
2. Of the total population in the various blocks of the study area, large proportion of the population is non-worker and marginal worker. Since the working population is largely situated in and near urban centers, the LULC transformation of cropland, natural vegetation, grassland and fallow land in remote areas can be attributed as a resultant of increased population to meet the end needs of food, fiber and shelter.

3. The results obtained from landscape fragmentation metrics show that in almost all the blocks, the largest patch of the landscape decreased with increase in number of patches and landscape shape index and patch density, and diversity index signifying large amount of fragmentation and transformation occurring among the various LULC categories in the various landscape of the district.

Acknowledgements Authors would like to acknowledge USGS and Census of India for providing satellite imageries and Census data free of cost.

References

1. Braimoh AK, Osaki M (2010) Land-use change and environmental sustainability. *Sustain Sci* 5:5
2. Turner BL II, Lambin EF, Reenberg A (2007) The emergence of land change science for global environmental change and sustainability. *Proc Natl Acad Sci* 104(52):20666–20671
3. Kindu M, Schneider T, Teketay D, Knoke T (2015) Drivers of land use/land cover changes in Munessa-Shashemene landscape of the south-central highlands of Ethiopia. *Environ Monit Assess* 187:452 (2015)
4. Hurni H, Tato K, Zeleke G (2005) The implications of changes in population, land use, and land management for surface runoff in the upper Nile basin area of Ethiopia. *Mt Res Dev* 25:147–154
5. Kathwas AK, Patel N (2021) Geomorphic control on soil erosion—a case study in the subarnarekha basin, India. *Pol. J. Soil Sci.* 54(1):1–24
6. Kicklighter DW, Cai Y, Zhuang Q, Parfenova EI, Paltsev S, Sokolov AP, Melillo JM, Reilly JM, Tchebakova NM, Lu X (2014) Potential influence of climate-induced vegetation shifts on future land use and associated land carbon fluxes in northern Eurasia. *Environ* 9:35004
7. Kathwas AK, Patel N (2019) Effects of landuse and landcover dynamics on the migration and demographic pattern of tribal population using geospatial techniques. *Theor. Empir. Res. Urban Manage.* 14(3):77–97
8. Nwokoro I, Dekolo S. Land use change and environmental sustainability: the case of Lagos Metropolis. *WIT Trans. Ecol. Environ.* **157**–167 (2012)
9. Baruah JJ, Baruah BK, Kalita1 S, Choudhury SK (2016) Impact analysis of open cast coal mining on land use/land cover using remote sensing and GIS technique in Ledo Margherita region of Assam, India. *Imp. J. Interdiscip. Res.* **2**, 671–676 (2016)
10. Das R (2020) Soil loss assessment in Sadiya Region, Assam, India using remote sensing and GIS. *Indian J. Sci. Technol.* **13** (2020)

Modelling Flash Flood Vulnerability and Sensitivity Dynamics of Jiadhal River Basin of Eastern Himalayan Range Using Space Technology and AHP



Rakesh Saur and Virendra Singh Rathore

Abstract Flood vulnerability mapping is significant for flood susceptibility analysis. This study aims to identify and assess the dynamics of flood vulnerability zones of Jiadhal River basin where flash flood is a frequent event in Monsoon. Space technologies such as remote sensing and geospatial technology were employed for investigating flood vulnerability. Influential parameters such as distance from river, elevation, surface roughness, drainage density, slope, soil texture and LULC (land use and land cover) were used in this study. Relative importance of these influential parameters, that is, priority weights were computed using analytical hierarchy process (AHP) method. Next, computed priority weights were used to identify and assess the dynamics of flood vulnerability zones. It has been observed that areas which are near to the river and are located at low elevation areas of river bed are in the high and very high vulnerability zones. Consequently, around 406.7 (35%) and 211.4 km² (18%) of the study area has been severely affected by flash flood.

Keywords Flash flood vulnerability · Analytical hierarchical process · GIS · Sensitivity analysis

1 Introduction

Flood is a common and frequently occurring worldwide event, defined as overflow of water deluging the areas adjacent to river bank. Floods are natural events which cannot be prevented and have the potential to cause fatalities, displacement of people, damage to infrastructure, environment and disrupt the biodiversity of the region [1]. Contrary to floods, flash flood is a short time-period event where channel discharge exceeds rapidly against the channel capacity, resulting in overbank flow [2]. The event is primarily governed by excessive rainfall, dam or levee failure, or sudden release

R. Saur (✉) · V. S. Rathore

Department of Remote Sensing, Birla Institute of Technology, Mesra, Ranchi 834215, India

V. S. Rathore

e-mail: vsrathore@bitmesra.ac.in

of water held by gridlock ice. Since, the destruction and disruption caused by flash floods are high, the researchers worldwide are making great efforts for predicting and modelling the hazard, risk and vulnerability of amenities and environment in the flood plain areas [3–5]. The modelling of flash flood events requires large amount of real-time datasets with various assumptions.

With the advancement of remote sensing and geographic information system (GIS) techniques, the development of hazard, risk and susceptibility models along with collection and management of the climate and geo-environmental datasets has become significantly easier. Worldwide, various flood predictive models have been developed by researchers, namely physical models (HECRAS, MIKE), stochastic and regression models [6, 7]. Contrary to regression, physical models require huge hydrological datasets; however, the newest physically based approach, the on-off classification model, does not require datasets from hydrological events but utilizes historical flood and geo-environmental datasets for delineation of the flood and non-flood categories using machine learning technique.

The present study has been conducted for Jiadhal River basin, a tributary of mighty Brahmaputra River, flowing across Arunachal Pradesh to Assam state of India. Due to heavy rainfall in the basin, there occur numerous flash flood event in the Jiadhal River basin [8, 9] causing severe channel migration and land use and land cover (LULC) dynamics in the region. The above rational became the prime motivation for assessing the flash flood hazard vulnerability which causes large amount of destruction each year in the region. Here, in the present study, an attempt has been made to (1) delineate the flash flood vulnerable zones using the analytical hierarchical process (AHP) in the GIS environment. (2) Assess the impact on the spatial extent under different vulnerability zones. (3) Sensitivity analysis of the vulnerability AHP weights of the parameters in Jiadhal River basin.

2 Study Area

The study area is situated between Arunachal Pradesh and Assam state (Fig. 1), India, spanning over approximately 1173.65 km² area, extending from 27°11'N to 27°4'N Latitude and 94°14'E to 94°38'E Longitude, respectively. The annual precipitation received by the study area ranges between 800 and 3000 mm. The climate of the study area is humid subtropical (<http://cgwb.gov.in/>) and gets flooded every year during the monsoon period due to heavy rainfall events in the upper reaches of the study area. The occurrence of torrential rainfall and the physiographical setup combined makes the study area prone to frequent flash flood events. Physiographically, the basin is divided into two broad categories, viz. the Himalayan hill range (34%) and the fluvial flood plains (66%).

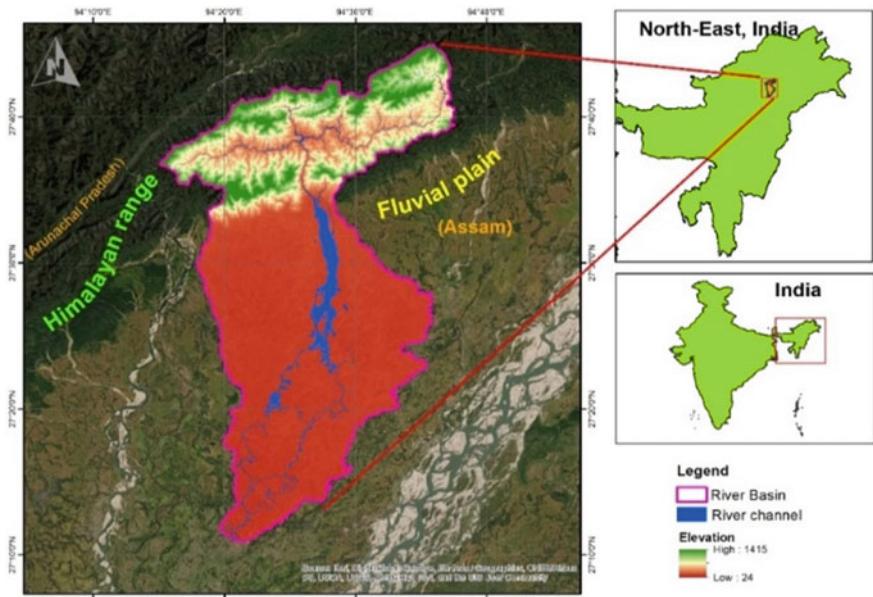


Fig. 1 Location of the study area

3 Methodology

Flash flood is very disastrous event of the study area, and to assess the flash flood vulnerability in a comprehensive way, the following method have been used for the Jiadhal River basin (Fig. 2). In the subsection given below, in-depth detail of the methodology has describe.

3.1 AHP Pairwise Comparison Analysis

At the onset, the satellite imagery pertaining to year November 2010 was downloaded from Bhuvan (<https://bhuvan-app3.nrsc.gov.in/>) and pre-processing and image enhancement was carried out using ERDAS Imagine 2014® software. The enhanced satellite imageries were further used to identify and delineate the major LULC categories (agriculture, inland wetland, river, scrub, settlement and vegetation) of the study area (Fig. 3a). Further, the digital elevation model (DEM) were also obtained from Bhuvan (<https://bhuvan-app3.nrsc.gov.in/>) and were used to derive the elevation, surface roughness and slope parameter. The topomaps (1:50,000 scale) were used for delineation of drainage channels and generation of drainage buffer for computation of distance from river parameter in the GIS environment using ArcGIS 10.0® software. The soil map procured from National Bureau of Soil and Landuse

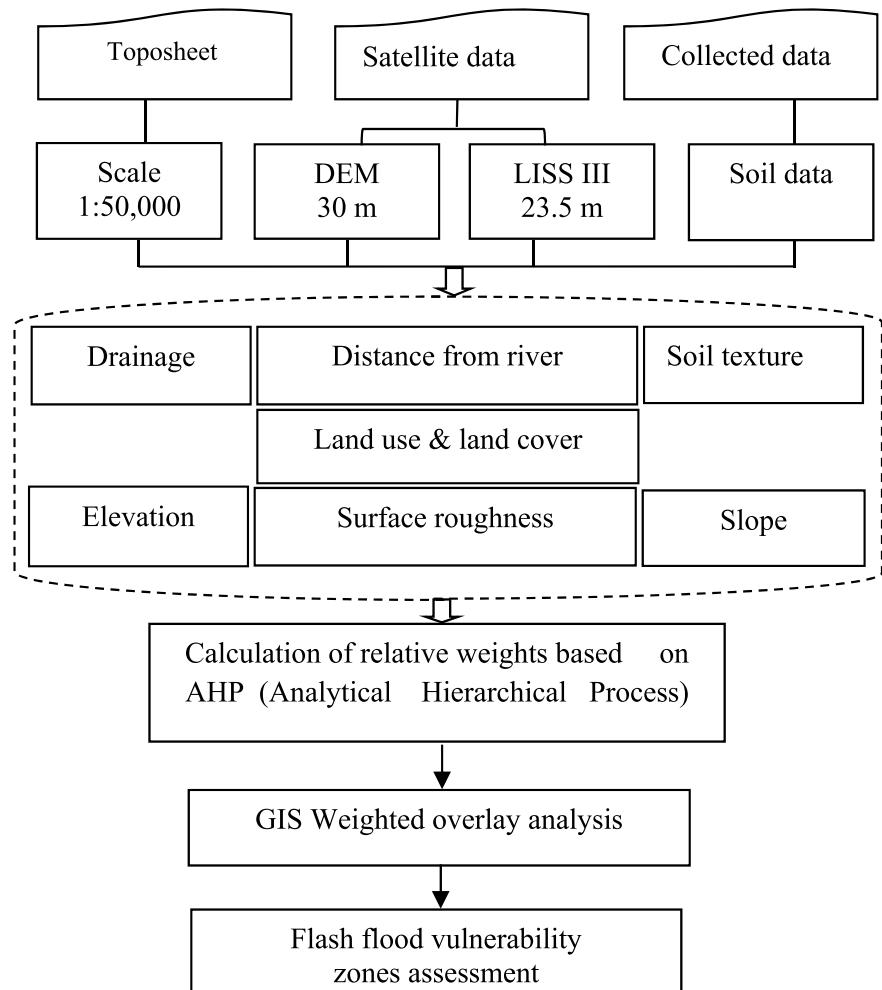


Fig. 2 Flowchart of research methodology used for the analysis

Planning (NBSS & LUP), Assam, was used for identifying and delineating the respective soil texture parcels of the study area. The integration of the respective parameters for determining the relative importance and computation of criterion weights was carried out, implementing the multi-criteria-based weighted overlay technique in the ArcGIS software. The computation of the criteria weights was assessed using AHP technique (Table 1).

The derivation of parameter criteria is a basic requirement of MCDM technique. Since, AHP is significant mathematical method for inspecting complex decision problems [10], here in the present study, the technique was used for assessment of the criteria weights by comparing the relative importance of the respective parameters

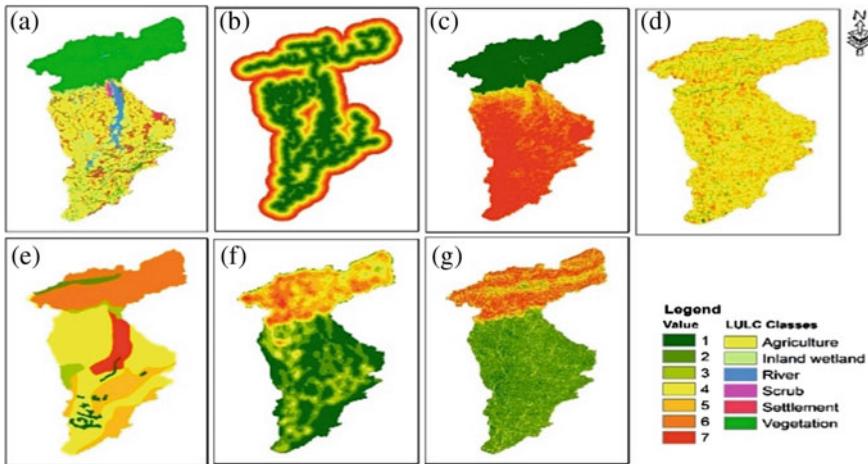


Fig. 3 Parameter layers **a** LULC, **b** Distance from the River, **c** Elevation, **d** Surface roughness, **e** Soil texture, **f** Drainage density, **g** Slope used for Flash flood vulnerability assessment of Jiadhal River basin

Table 1 Normalized flood vulnerability parameters for assessing AHP

Parameters	Dr	Ele	Sr	Dd	Sl	So	Lulc
Dr	1	3	2	3	5	5	9
Ele	0.33	1	3	3	3	3	5
Sr	0.5	0.33	1	3	3	3	5
Dd	0.33	0.33	0.33	1	3	3	3
Sl	0.2	0.33	0.33	0.33	1	3	3
So	0.2	0.33	0.33	0.33	0.33	1	3
Lulc	0.11	0.2	0.2	0.33	0.33	0.33	1

where *Dr* distance from main channel, *Ele* elevation, *Sr* surface roughness, *Dd* drainage density, *Sl* Slope, *So* Soil texture, *LULC* Land use and land cover

using pairwise comparison matrix, considering the Eigenvector corresponding to the largest Eigenvalue of the matrix. It can be represented as to the largest Eigen value of the matrix and then normalizing the sum of the components to unity as:

$$\sum_{i=1}^n w_i = 1 \quad (1)$$

The overall importance of every individual criterion is then calculated. An importance scale is proposed for these comparisons (Table 1). The basic input is the pairwise comparison matrix A, comprised of n criteria, can be established using Saaty [11] AHP scale (Table 2), defined in Eq. (2):

Table 2 Shows priority weights and consistency ration of the AHP analysis

Parameters	Distance from main channel	Elevation	Surface roughness	Drainage density	Slope	Soil texture	LULC	CI	CR
Priority weight	0.35	0.22	0.17	0.11	0.07	0.05	0.03	0.093	0.068

$$A = [a_{ij}], i, j = 1, 2, 3, \dots, n \quad (2)$$

where A is the matrix with elements a_{ij} .

The matrix generally has the property of reciprocity which can be expressed mathematically as:

$$a_{ij} = \frac{1}{a_{ji}} \quad (3)$$

Further, the matrix values are normalized as matrix B (Eq. 4):

$$B = [b_{ij}], i, j = 1, 2, 3, \dots, n. \quad (4)$$

where B is the normalized matrix of A with elements b_{ij} and can be defined as:

$$b_{ij} = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}} = 1, 2, 3, \dots, n \quad (5)$$

Each weight w_i is then computed as using Eq. 6

$$w_i = \frac{\sum_{j=1}^n b_{ij}}{\sum_{i=1}^n \sum_{j=1}^n b_{ij}}, i, j = 1, 2, 3, \dots, n. \quad (6)$$

Equations 7, 8 and 9 present the relationships between the largest Eigenvalue (λ_{\max}) and corresponding Eigenvector (W) of the matrix B [11]:

$$BW = \lambda_{\max} w \quad (7)$$

The CR (consistency ratio) indicates the likelihood of the matrix judgements [10].

$$CR = \frac{CI}{RI} \quad (8)$$

Moreover, the random index (RI) is the average of the resulting consistency index, depending on the order of the matrix given by Saaty [11], and the consistency index (CI) can be expressed as:

$$CI = \frac{(\lambda_{\max} - n)}{n - 1} \quad (9)$$

where λ_{\max} is the largest Eigenvalue of the matrix, and n is the number of parameters. A CR value of 0.10 or less indicates a reasonable level of consistency (Table 3).

Table 3 Shows random consistency ration index (Saaty 1980)

<i>N</i>	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.46	1.49

3.2 Flood Vulnerability Zones Analysis

To assess the flood vulnerability, seven different geospatial parameters (LULC, distance from the river, elevation, surface roughness, soil texture, drainage density and slope) layers were used (Fig. 3). Every parameter plays different significant role in the hazardous scenario. So, according to that, parameters were divided into different categories and given to weight by their level of importance in the flood vulnerability point of view; also, priority weights calculated using AHP methods for every individual parameter was used to analysis the flood vulnerability assessment.

Performing the flood vulnerability zones GIS weighted overlay tool was used in the ArcGIS software. Weighted overlay analysis tool in ArcGIS software allows the user to read the spatial raster data and provide the facility to give weights into the different categories of individual parameter and also AHP priority weights of every sole parameter. To get the final output of the analysis, AHP weights (Table 2) for distance from main channel (0.35), elevation (0.22), surface roughness (0.17), drainage density (0.11), slope (0.07), soil texture (0.05) and LULC (0.03) were used and categorized the results into different flash flood vulnerability zones, viz. low flood vulnerability (LFV), moderate flood vulnerability (MFV), high flood vulnerability (HFV) and very high flood vulnerability (VHFV).

3.3 Sensitivity Analysis

To assess the sensitivity analysis, common methods have been used by changing the input parameter weights and see the effect in output results. This is the logical approach to assess the effects after changing the parameters weights.

There are many ways used to analysis criteria sensitivity [12]. In this analysis, interested in exchanging parameters weights only, within three parameters (1) distance from the river, (2) elevation and (3) surface roughness, which is most influencing parameters in the study area.

To assess the sensitivity, three iteration have run, in iteration 1 exchange the weights (Table 1) of (1) distance from the river and (2) elevation and keep the rest parameters weights as it is. In iteration 2 (2) elevation and (3) surface roughness and iteration 3 (1) distance from the river and (3) surface roughness, respectively.

4 Results and Discussions

Table 1 depicts the pairwise comparison matrix used for determining the parameters weights. Since the criteria weights were based on expert's opinion and knowledge-based weighting, the analysis of Table 1 shows that distance from river (35.17) followed by elevation (21.63) and surface roughness (16.75) are the most influencing flash flood parameter in the study area. On the contrary, slope (7.29), soil texture (5.33) and LULC (3.09) have moderate to low influence in determining the flash flood vulnerable zones. Table 4 presents the spatial extent of study area under respective vulnerable zones.

From the analysis (Fig. 4), it can be deduced that approximately, 44% of the study area falls under moderate vulnerable zone, whereas approximately 54% of the study area comes under high to very high vulnerable zones and nearly 3% area under the low vulnerable zone.

Further, in order to assess the sensitivity of the most influencing parameters, i.e. distance from river, elevation, surface roughness, the weights of the parameters were exchanged and area under respective vulnerable zones were determined. In iteration 1, the weights determined from AHP of parameter 1 (distance from river) and parameter 2 (elevation) were exchanged whereas, iteration 2, the weight for parameter

Table 4 Sensitivity analysis under (a) AHP, (b) Iteration 1, (c) Iteration 2 and (c) Iteration 3 and spatial extent of study area under various vulnerable zones

	LFV	MFV	HFV	VHFV
AHP	31.3	507.7	406.7	211.5
Iteration 1	19.1	525.2	551.9	60.9
Iteration 2	54.2	459.1	433.8	210.1
Iteration 3	26.8	646.3	471.6	12.4

where *LFV* Low flood vulnerability, *MFV* moderate flood vulnerability, *HFV* high flood vulnerability, *VHFV* very high flood vulnerability (unit-km²)

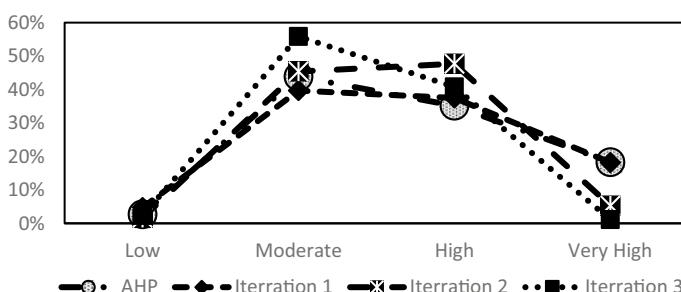


Fig. 4 Spatial dynamics percentage area under different flash flood vulnerability zones determined using **a** AHP, **b** Iteration 1, **c** Iteration 2 and **d** Iteration 3 of Jiadhal river basin

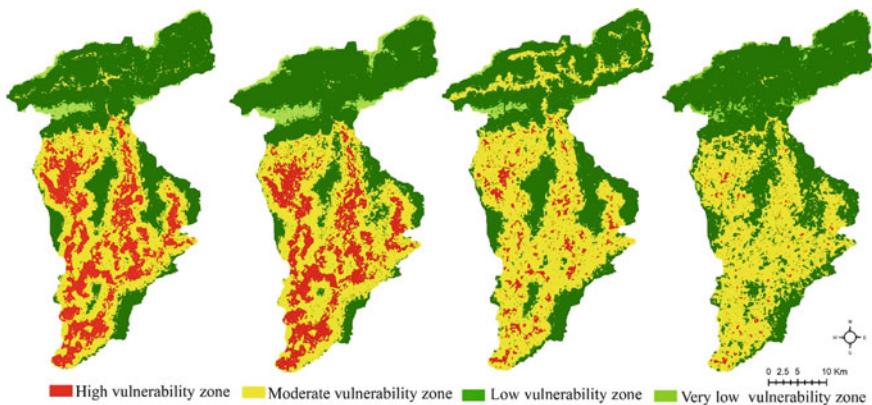


Fig. 5 Present map showing the spatial distribution of study area under respective flash flood vulnerability zones assessed using **a** AHP, **b** Iteration 1, **c** Iteration 2 and **d** Iteration 3 of Jiadhal river basin

2 (elevation) was assigned to parameter 3 (surface roughness), whereas in iteration 3, the criteria weights of parameter 1 (distance from river) and parameter 3 (surface roughness) were exchanged. The results (spatial extent) obtained from the analysis is presented in Table 4. The results from iteration 1 show that majority proportion (48%) of the study area falls under high vulnerable zone. The transformation primarily occurred from the shift of area under very high vulnerable zone to high. On the other hand, from iteration 2, it can be observed that area under very high remains intact, however, significant amount of exchange of area occurred in low to highly vulnerable zone resulting in majority (40%) area falling under moderate vulnerable zone. Furthermore, iteration 3 shows large amount (56%) of increase in spatial extent of moderate and high (41%) vulnerable zones. The transformation in iteration 3 resulted in minuscule proportion (1%) of area under very high vulnerability zone.

From the above observation made from results obtained, it can be deduced that the three parameters have significant amount of impact on flash flood vulnerability analysis and the area of the output results of in the study area (Fig. 5), whereas the other parameters tend to play small role in determining the vulnerable zones. A detailed sensitivity analysis of the respective criterions may show significant insight about the influence of the parameters on the flash floods.

5 Conclusion

The aim of the present study was to determine the flash flood vulnerable zones using the AHP technique. Since the study area is part of the Brahmaputra River basin, the occurrence of yearly flash flood brings havoc in the study area by not only disrupting the natural environment of flora and fauna but also does huge damage to human

infrastructure and environment in the study area. From the above investigation, it is revealed that (1) AHP techniques are very useful to assess the vulnerability zones and observe satisfactory results within the given parameters. (2) Vulnerability analysis using AHP deduced that large part of the study area is under high to very high vulnerability zones, which is about 406.7 (35%) and 211.5 km² (18%), respectively. Also, study shows that (3) AHP parameters and their weights play significant role into the analysis. So, sensitivity analysis results found that distance from river is the most determining parameter for vulnerable zones of flash floods followed by elevation and surface roughness. The study reveals that in order to prevent the major amount of damage and disruption, the strategies should be developed to shift the human infrastructures at significant distance from the river banks, since the floods and flash floods are natural phenomenon and largely cannot be prevented.

Acknowledgements Authors would like to acknowledge USGS and ISRO-Bhuvan for providing satellite imageries free of cost and local people for providing information during field study. Author acknowledge Vice - Chancellor, Birla Institute of Technology, Mesra, Ranchi for encouraging research and UGC-RGNF for providing research fellowship.

References

1. Dawod GM, Mirza MN, Al-Ghamdi KA (2012) GIS-based estimation of flood hazard impacts on road network in Makkah city Saudi Arabia. Environ Earth Sci 67:2205–2215
2. Gomaa MD, Meraj NM, Khalid AA (2011) GIS-based spatial mapping of flash flood hazard in Makkah City, Saudi Arabia. J Geogr Inf Syst 3:225–231
3. Elkhrachy I (2015) Flash flood hazard mapping using satellite images and GIS tools: a case study of Najran City, Kingdom of Saudi Arabia
4. Diakakis M (2011) A method for flood hazard mapping based on basin morphometry: application in two catchments in Greece. Nat Hazards 56:803–814
5. Khosravi K, Shahabi H, Pham BT, et al (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J Hydrol
6. Liu Z, Merwade V, Jafarzadegan K (2018) Investigating the role of model structure and surface roughness in generating flood inundation extents using one- and two-dimensional hydraulic models. J Flood Risk Manage e12347
7. Fekete A (2019) Social vulnerability (re-)assessment in context to natural hazards: review of the usefulness of the spatial indicator approach and investigations of validation demands. Int J Disaster Risk Sci 10:220–232
8. Das PJ (2011) Building communities' capacity for flash flood risk management in the Jia-Dhal river basin Dhemaji district, Assam. Thesis Report, India
9. Das PJ, Bhuyan HK (2013) Policy and institutions in adaptation to climate change, case study on flood mitigation infrastructure in India and Nepal. ICIMOD working paper 2013/4 (2013)
10. Saaty TL, Vargas LG (1991) Prediction, projection and forecasting. Kluwer Academic, Boston
11. Chen Y, Khan S (2010) Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. Environ Model Softw
12. Daniel C (1958) On varying one factor at a time. Biometrics 14:430e431

Automating Vehicle Insurance Process Using Smart Contract and Ethereum



Saroj Kumar Nanda, Sandeep Kumar Panda, Madhabananda Das,
and Suresh Chandra Satapathy

Abstract The existing vehicle insurance system faces number of challenges like frauds, absence of P2P insurance, no timely settlement of claims and absence of on-demand insurance. However, the existing vehicle insurance system is centralized, and not settlement of claims efficiently. Hence, to dissolve this issue, this paper presents a distributed vehicle insurance system. Blockchain is an open, public, cryptographically hashed system that gives immutability and authenticity through digital signature incorporated with hash value. Blockchain exploits hash functions and encryption in the direction of accumulation of incontrovertible records which are more strongly secure over various present very similar cybersecurity resolutions technologies. On the other hand, the existing security deals with a single trusted authority to authenticate information or accumulate encrypted data. In this paper, we proposed a decentralized application (DApp), which deliberates the introduction of blockchain technology followed by Ethereum and smart contracts. Hence, for implementing the smart contract, public blockchain platform used. Finally, the React is used for frontend development.

Keywords Blockchain · Ethereum · DApp · React · Smart contract

1 Introduction

The financial services industry is playing a vital role in commodity transaction and society welfare since it is shaping up future saving and investment [1, 2]. It is providing protection from risks and safeguards to business enterprises. [3] The basis of information technologies has changed the industry over decades. It is enabling an extensive increase in transactions and diversification of products. [4] However,

S. K. Nanda · M. Das · S. C. Satapathy

School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India
e-mail: mndas_prof@kiit.ac.in

S. K. Panda (✉)

Department of Data Science and Artificial Intelligence, IcfaiTech (Faculty of Science and Technology), ICFAI Foundation for Higher Education, Hyderabad, India

the current pace of ICT innovation [5] in business sector has great potential for blockchain. This is mostly happening due to security burden and to decentralize the security burden that are lodged within the industry. The comprehensive view of blockchain technology serves as paradigm shift in security and innovation [6–8].

This paradigm shift benefits all enterprises in the long run. The concern of security and innovation is needed in case of vehicle insurance. Consider a situation when someone sells his or her vehicle, and ownership of the vehicle needs to be changed [9]. The process of transfer of vehicle ownership involves number of authorities, namely Regional Transport Authority [10] and insurance company. It is mandatory to update new ownership at each authority to avoid disputes. Many times, ownership is not updated at all nodes uniformly, especially in insurance company database [11, 12]. In that scenario, insurance company holds details of old owner, and at the time of reinsurance [13], they try to communicate with old owner. This ultimately results in loss to their business.

Blockchain found the solution of these issues, because of its distributed ledger technology network that allows any participants in the network to see the one system of record [14]. Change in the ownership of vehicle at one node results in uniform updating of ownership at all the others nodes present in the network [15].

The existing vehicle insurance system faces many challenges [13, 16, 17]. These challenges are non-proper remote interactions, frauds, absence of P2P [18, 19] insurances, delay in claims processing, hidden terms and conditions in insurance and lack of on-demand insurance. Blockchain technology has the potential to solve the above-mentioned challenges in efficient and profound manner.

2 Literature Survey

It is well known that the existing vehicle insurance system is not efficient. The network is neither customer-centric nor it provides benefits to the companies that provide insurance policies. This is due to the frauds involved in claiming processing. Sometimes the user bribes inspecting officer at the time of claim processing and take false reimbursement from the company [20, 21].

Even the customers face problems like hidden terms and conditions of policy, which are not told to them at the time when the customer avails insurance policy, and no timely settlement of claims; lot of documentation work is required even at the time of renewing the insurance policy, and no P2P insurance facility [22–24]. The existing vehicle insurance system involves brokers who charge commission from both insurance companies and customer. This leads to increase in the overall cost of insurance policy.

Duvivier titled “Is the Blockchain the New Graal of the Financial Sector?” [25] tells how P2P insurance can benefit the customers. Similarly, a paper by Fabrizio Lamberti and Valentina Gatteschi titled “Blockchains Can Work for Car Insurance” [26] tells how blockchains and sensors installed on a vehicle could be combined

to semi-automatically activate/deactivate car insurance coverage in an envisaged on-demand insurance scenario.

In order to solve the above-mentioned problem using robust technology like blockchain, first, we need to understand the underlining concepts of the technology.

A blockchain consists of distributed ledgers where all the transactions are recorded and maintained in a chronological order [27, 28]. All these transactions are wrapped together to form a block, and each block points to its previous block through some hashing mechanism [29–35]. This association is greatly similar to the data structure linked list. In a nutshell, bunch of blocks which contain transactions are chained together to form a blockchain [36–41].

3 System Architecture

In this segment, we put forward a decentralized system model for vehicle insurance process based on blockchain technology.

The proposed model for vehicle insurance system using blockchain technology facilitates various advantages over the traditional existing system. These advantages are discussed below lucidly.

3.1 *Eliminates Brokers*

Blockchain puts an end to intermediaries and brokers. Eliminating intermediaries not only speeds up the entire insurance process but also reduces overall insurance cost by abolishing broker commission charges. This in turn helps the person economically.

3.2 *Reduces Risk and Increases Trust*

Blockchain eradicates fake insurance and increases trust through shared processes and record keeping. Blockchain technology achieved trust through consensus, provenance, immutability and finality. It also ensures that assets in a system have a verifiable audit trail.

3.3 Policy Creation and Claim Processing Through Smart Contract

All the insurance policies are written using smart contract, a piece of code with embedded business logic, rules of an agreement and that stores asset (anything of important value)/currency.

Smart contract brutalizes transaction processing and thus automating the steps involved in vehicle insurance. For example, when a person claims the insurance, the smart contract triggers the respective function and settles the transaction within a fraction of seconds without any delay. In this manner, blockchain provides better solution to existing problem.

3.4 Renew Insurance

Blockchain technology has a prospective to contrivance reinsurance process in much more easy way in comparison to already existing method. Using smart contract, reinsurance premium policies can be shared to both the parties as the same time, eliminating the need of broker and revaluation process of the vehicle. There is no need to complete documentation process again in order to do insurance.

4 System Overview

In this paper, we developed a decentralized application (DApp) to overcome the problems that user faces in existing system.

The decentralized application has participating nodes or peers, as shown in Fig. 1. These nodes are insurance company, police station, repair workshop and insurer.

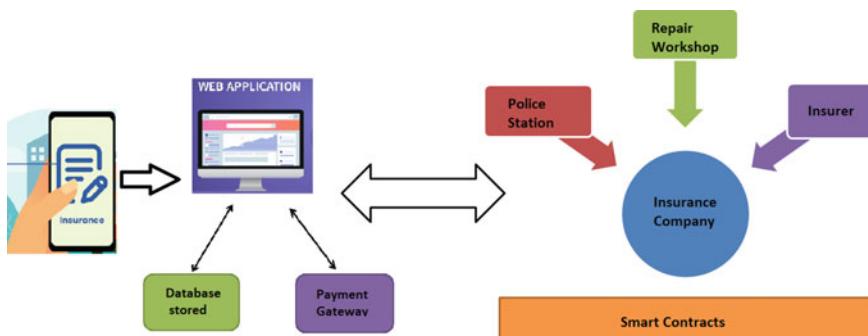


Fig. 1 Smart contract-based vehicle insurance DApp

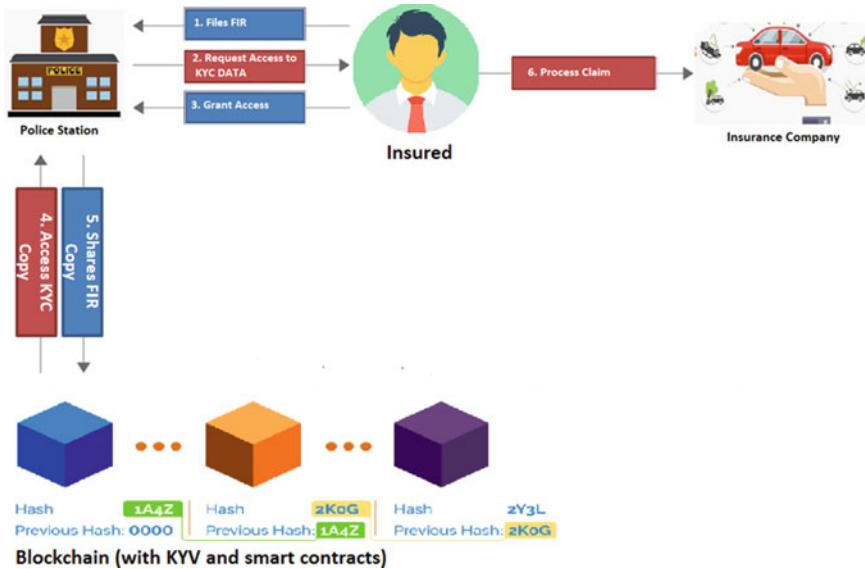


Fig. 2 Steps involved in claim process

Each of the above-mentioned nodes has a functionality associated with it for the effective functionality of the overall network. Insurance company is the node responsible for providing insurance policy for the vehicle of the interested user and is responsible for timely settlement of the claims. The police station node is responsible to authenticate the theft. The repair workshop node is responsible for repairing the vehicle.

Vehicle insurance system consists of many activities. These activities involve searching a relevant insurance policy for the vehicle, buying the insurance policy, renewing the insurance policy and claim process. Each activity involves number of steps in it. For example, when the vehicle of the user is stolen, he first files the FIR at the police station and also submits the necessary evidences like original registration copy issued by regional transport office (RTO), keys of the vehicle etc. For each step, smart contracts trigger the function for the effective working, and finally, timely settlement of claim is processed by the insurance company. Figure 2 shows the use case diagram for the claim process in the vehicle insurance system.

In the next section, implementation of the system model is discussed.

5 Result Analysis

Users retrieved the data from decentralized applications; the home page is available to the user, as shown in Fig. 3. The webpage gives various options to navigate across

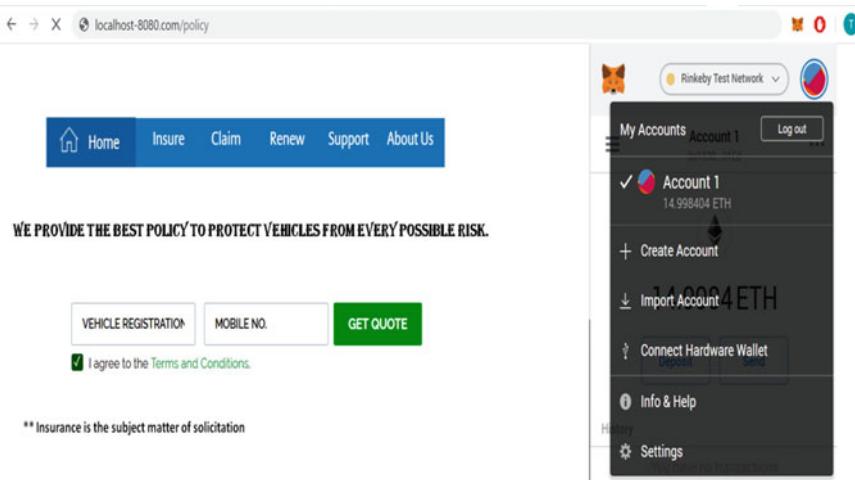


Fig. 3 Home page

the DApp. These options are Home, Insure, Claim, Renew, Support and About Us. User can click on any of these options as per his/her requirement.

The Home Page also gives the option to enter vehicle registration number and mobile number to get quotation of the vehicle insurance (Fig. 3).

If the user is interested to get vehicle insurance policy, then the user needs to click on “Insure” option available on the webpage, as shown in Fig. 4. Further, the user needs to enter the required details related to vehicle. Figure 4 also shows a insure vehicle account.

Fig. 4 Insure vehicle

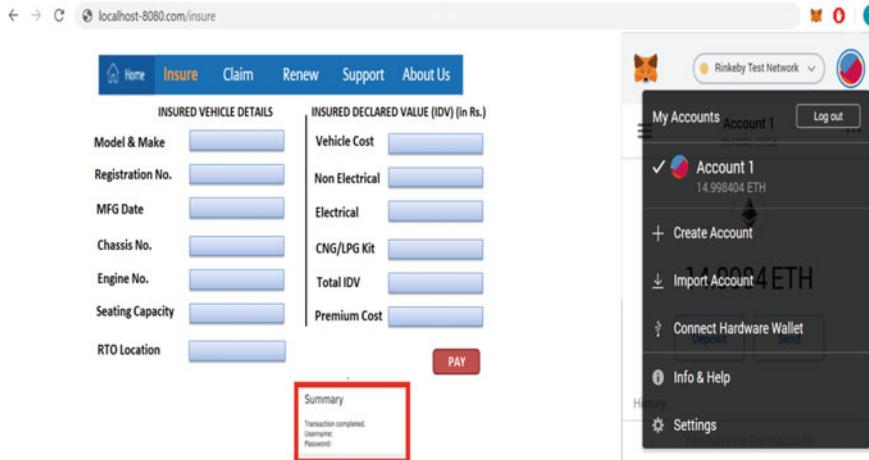


Fig. 5 Insured vehicle details

In existing vehicle insurance system, sometimes, the user claims false reimbursement of the policy after the theft of the vehicle even if the vehicle was not insured under theft protection policy. To avoid such disputes, we designed the smart contract that asks the users whether he/she wants theft protection for the insured vehicle or not, as shown in Fig. 4. This helps to avoid disputes between both the parties at the time of claim processing.

The next step in availing the insurance policy is to enter the details of the vehicle that is used to calculate IDV (INSURED DECLARED VALUE) and premium cost. The function defined in the smart contract takes input value of model and make, MFG date, seating capacity to estimate IDV and premium cost, as shown in Fig. 5.

The declared IDV varies as per the MFG date and seating capacity of the insured value. The IDV value constantly decreases very consecutive year. The webpage gives the option to PAY the premium cost in form of ethers. Once the premium cost is paid, the user gets the summary indicating that transaction is completed. After the transaction is completed, it is added as a new block of record to existing database of the insurance company.

The designed user interface also gives the option to user to renew the policy, as shown in Fig. 6.

The webpage asks the user to enter the last policy no. The function designed in the smart contract automatically retrieves the store data from the database and calculates IDV and new premium cost. If the user is interested to buy policy, the user needs to PAY equivalent ethers for the same.

One of the important features in the vehicle insurance platform is the claim process. The claim process can be either due to accidental damage of the vehicle or due to theft. At the time of claim processing, the user need to click on either DAMAGE or THEFT option, as shown in Fig. 7 depending upon the type of claim.

The screenshot shows a web application for renewing a vehicle policy. At the top, there is a navigation bar with links for Home, Insure, Claim, Renew, Support, and About Us. Below the navigation bar, a section titled "RENEW YOUR VEHICLE POLICY" contains fields for Last Policy No., Registration No., Insured Name, and Mobile No., each with a blue input field. A green "SUBMIT" button is located below these fields. Below the submission area, there are two more input fields: Total IDV Cost and Premium Cost, followed by a red "PAY" button. To the right of the main form, a separate window titled "Account 1" is displayed, showing a balance of 0x1338...31Ed and a "Details" button. The background of the main page has a light gray gradient.

Fig. 6 Renew policy

The screenshot shows a web application for claiming insurance under the "DAMAGE" category. At the top, there is a navigation bar with links for Home, Insure, Claim, Renew, Support, and About Us. Below the navigation bar, a message says "Have you damaged your car? No need to fret, we are with you!" followed by two buttons: "DAMAGE" and "THEFT". The main form contains fields for Policy No., Contact No., Chassis No., Date of Accident, Kilometer Reading, Vehicle Inspection Address, and a "Submit" button. Below the form, a note reads "Claims Discharge Cum Satisfaction Voucher signed across a Revenue Stamp". To the right, a separate window titled "My Accounts" shows "Account 1" with a balance of 14.9998404 ETH. Other options in the account list include "+ Create Account", "- Import Account" (with a balance of 4 ETH), and "Connect Hardware Wallet". The "Info & Help" and "Settings" buttons are also visible.

Fig. 7 Claim under damage category

Figure 7 shows the details which the user needs to fill if he/she chooses insurance claim under DAMAGE category. The function in the smart contract takes the input and cross validates the entered data with the data stored in the blockchain.

The user then gets the message indicating the time at which vehicle inspection would be done by the insurance company. Once the inspection is done, the damaged vehicle photographs are added onto the blockchain by the concerned inspecting officer. The claim is then validated. The vehicle is then sent to the repair workshop for repair. Once the repair is done, the owner of the workshop (who is also of the node of the blockchain network) submits the required details on the Web including the charges involved in repairing the vehicle. Then, the smart contract automatically triggers the function to process the reimbursement within fraction of seconds without any time delay. The insurance company makes payment directly to the repair workshop that repaired the vehicle. This helps in timely settlement of the claims.

Although the designed system model requires a third person to validate the transaction, the entire system design is efficient enough to avoid any type of tampering.



Fig. 8 Claim under theft category

Although we have not considered police station as a participating node when a user claims under damage category, results could be even more commendable if we consider it.

Figure 8 shows the details which the user needs to enter if he/she chooses claim under THEFT category. The user enters the FIR No. of the complaint he filed in police station. The contract cross validates the entered information.

6 Conclusions and Future Scope

Blockchain recently analyses the latest generic solution of every enterprise. The technology is changing the entire gamut of transaction over network. Blockchain projects a scalable security to its end user when it is integrated with other robust technologies like cloud computing or artificial intelligence. Cloud computing is a conglomeration of large network and virtual data store. When we amalgamate blockchain with cloud, it gives a new insight to storage, replication and access to transactional database. It blends the security concept between task user and data collaborator over cloud. The future scope of this paper verifies the research gap exists between blockchain and cloud in terms of data modelling, orientation and artificial intelligence combined with blockchain for better results for vehicle insurance system. To converge all application in blockchain, it incurs high operational cost. We must focus on privacy and availability of data on block-cloud creation in edge computing, where we put resource close to device location. It can effectively utilize all resource over cloud under constrained security. To achieve safe, secure and scalable data sharing over blockchain model, we still wait for a relevant algorithm which will resolve dishonest behaviour of nodes prior to data movement over network.

References

1. Kolb RW (Ed) (2010) Lessons from the financial crisis: causes, consequences, and our economic future (vol 12). John Wiley & Sons
2. Varghese A, Salim MH (2015) Handloom Industry in Kerala: a study of the problems and challenges. *Int J Manage Soc Sci Res Rev* 1(14):347–353
3. Hileman G, Rauchs M (2017) Global cryptocurrency benchmarking study. Cambridge Centre for Alternative Finance, 33
4. Narayanan A, Bonneau J, Felten E, Miller A, Goldfeder S (2016) Bitcoin and cryptocurrency technologies: a comprehensive introduction. Princeton University Press
5. Bardhan A (2008) Of subprimes and subsidies: the political economy of the financial crisis. Available at SSRN 1270196
6. Bouri E, Roubaud D, Shahzad SJH (2019) Do bitcoin and other cryptocurrencies jump together. *Q Rev Econ Fin*
7. Eigner P, Umlauf TS (2015) The great depression(s) of 1929–1933 and 2007–2009? Parallels, differences and policy lessons. *Parallels, differences and policy lessons* (1 July 2015). Hungarian Academy of Science MTA-ELTE Crisis History Working Paper (2)
8. Sodhi NS, Bickford D, Diesmos AC, Lee TM, Koh LP, Brook BW, Bradshaw CJ (2008) Measuring the meltdown: drivers of global amphibian extinction and decline. *PloS One* 3(2):e1636
9. Lewis PH (2009) The agony of Argentine capitalism: from Menem to the Kirchners. Praeger, Westport, CT
10. Temin P (2010) The great recession & the great depression. *Daedalus* 139(4):115–124
11. Sutradhar D US Financial crisis: causes and its impact on India
12. Mukeshchandra PC (2018) Impact of global financial crisis on indian economy
13. Williams M (2010) Uncontrolled risk: lessons of Lehman brothers and how systemic risk can still bring down the world financial system. McGraw Hill Professional
14. Eichengreen B, O'Rourke K (2010) A tale of two depressions: what do the new data tell us? VoxEU, org, 8
15. Panda SK, Satapathy SC (2021) An investigation into smart contract deployment on ethereum platform using Web3.js and solidity using blockchain. In: Bhateja V, Satapathy SC, Travieso-González CM, Aradhya VNM (eds) Data engineering and intelligent computing. Advances in intelligent systems and computing, vol 1. Springer, Singapore. https://doi.org/10.1007/978-981-16-0171-2_52
16. Marr B (2017) A short history of bitcoin and crypto currency everyone should read. Dostupno na <https://www.forbes.com/sites/bernardmarr/2017/12/06/a-short-history-ofbitcoin-and-crypto-currency-everyone-should-read>
17. Chitale R (2008) Seven triggers of the US financial crisis. *Econ Polit Weekly*, pp 20–24
18. Panda SK, Rao DC, Satapathy SC (2021) An investigation into the usability of blockchain technology in internet of things. In: Bhateja V, Satapathy SC, Travieso-González CM, Aradhya VNM (eds) Data engineering and intelligent computing. Advances in intelligent systems and computing, vol 1. Springer, Singapore. https://doi.org/10.1007/978-981-16-0171-2_53
19. Sathya AR, Panda SK, Hanumanthakari S (2021) Enabling smart education system using blockchain technology. In: Panda SK, Jena AK, Swain SK, Satapathy SC (eds) Blockchain technology: applications and challenges. Intelligent systems reference library, vol 203. Springer, Cham. https://doi.org/10.1007/978-3-030-69395-4_10
20. Bayern S (2013) Of bitcoins, independently wealthy software, and the zero-member LLC. *Nw UL Rev* 108:1485
21. Lokre SS, Naman V, Priya S, Panda SK (2021) Gun tracking system using blockchain technology. In: Panda SK, Jena AK, Swain SK, Satapathy SC (eds) Blockchain technology: applications and challenges. Intelligent systems reference library, vol 203. Springer, Cham. https://doi.org/10.1007/978-3-030-69395-4_16
22. Giungato P, Rana R, Tarabella A, Tricase C (2017) Current trends in sustainability of bitcoins and related blockchain technology. *Sustainability* 9(12):2214

23. Foley S, Karlsen JR, Putniņš TJ (2019) Sex, drugs, and bitcoin: how much illegal activity is financed through cryptocurrencies? *Rev Fin Stud* 32(5):1798–1853
24. Agree TTE (2009) Worst financial crisis since great depression. Risks increase if right steps are not taken (<http://ca.news.finance.yahoo.com/s/13022009/34/biz-f-business-wire-three-top-economists-agree-2009-worst-financial-crisis.html>), Business Wire News Database
25. Duvivier PJ Is the blockchain the new grail of the financial sector?
26. Lamberti F, Gatteschi V, Demartini C, Pelissier M, Gómez A, Santamaría V Blockchains can work for car insurance using smart contracts and sensors to provide on-demand coverage
27. Davis EP, Karim D (2008) Could early warning systems have helped to predict the sub-prime crisis? *Natl Inst Econ Rev* 206(1):35–47
28. Sivaraman B (2008) The impact of the US meltdown on the indian economy
29. Bouri E, Gupta R, Roubaud D (2019) Herding behaviour in cryptocurrencies. *Fin Res Lett* 29:216–221
30. Brauneis A, Mestel R (2018) Price discovery of cryptocurrencies: bitcoin and beyond. *Econ Lett* 165:58–61
31. Fry J (2018) Booms, busts and heavy-tails: The story of bitcoin and cryptocurrency markets? *Econ Lett* 171:225–229
32. Yuan Y, Wang FY (2018) Blockchain and cryptocurrencies: model, techniques, and applications. *IEEE Trans Syst Man Cybern Syst* 48(9):1421–1428
33. Corbet S et al (2018) Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Econ Lett* 165:28–34
34. Dierksmeier C, Seele P (2018) Cryptocurrencies and business ethics. *J Bus Ethics* 152(1):1–14
35. https://en.wikipedia.org/wiki/Financial_crisis_of_2007%E2%80%932008
36. Nachane DM (2009) The fate of India unincorporated. *Econ Polit Weekly*, pp 115–122
37. Murphy RT (2008) Asia and the meltdown of American Finance. *Econ Polit Weekly*, pp 25–30
38. Aalbers MB (2015) The great moderation, the great excess and the global housing crisis. *Int J Hous Policy* 15(1):43–60
39. Kharif O, Leising M (2018) Bitcoin and blockchain. *Bloomberg QuickTake*
40. Gatteschi V, Lamberti F Blockchain and smart contracts for insurance: is the technology mature enough?
41. Jena AK, Dash SP (2021) Blockchain technology: introduction, applications, challenges. In: Panda SK, Jena AK, Swain SK, Satapathy SC (eds) *Blockchain technology: applications and challenges. intelligent systems reference library*, vol 203. Springer, Cham. https://doi.org/10.1007/978-3-030-69395-4_1

Energy-Aware Load Balancing in Dynamic Cloud Environment Using Nature-Inspired Technique



Bivasa Ranjan Parida and Amiya Kumar Rath

Abstract Cloud computing has shown its eminence in almost all fields of the technologies spread over the globe. It has enormous benefits in provisioning the cloud resources to the users, being anywhere, at any moment, on a hiring basis as a utility. A few studies regarding dynamic cloud load balancing with energy efficiency have been proclaimed so far. Hence, adequate models in this area are still rare. This paper has proposed a nature-inspired hybridized cuckoo search algorithm (HCS-LB), combining the searching technique implemented by the cuckoo to lay eggs in the nest of another bird and the multi-rumen anti-grazing behavior of the herbivores to balance the load of tasks in multiple cloud datacenters, as a result minimizing the energy consumption in the cloud environment as a whole. This hybrid algorithm uses cuckoo search to allocate appropriate virtual machines to the migrated tasks and the multi-rumen anti-grazing approach to assign tasks to other datacenters as per the requirement. Here, CloudSim is used to simulate the result, and the proposed model is found outperforming over the other existing techniques. This will contribute a bit to the huge requirement to minimize the expenses of the datacenters incurred by the energy consumption and thereby reduce the cost of usage of the cloud resources by the general public.

Keywords Cloud computing · Cloud datacenters · Load balancing · Energy consumption · Cuckoo search · Multi-rumen anti-grazing approach

B. R. Parida (✉) · A. K. Rath

Department of Computer Science & Engineering, Veer Surendra Sai University of Technology, Burla, Sambalpur, Odisha, India

A. K. Rath

e-mail: akrath_cse@vssut.ac.in

1 Introduction

Energy consumption cost is the swiftly growing cost element in the cloud datacenters nowadays. Efficient energy management to control its consumption in cloud environment is becoming a challenge for the cloud researchers. Ideally, power consumption of network devices in cloud datacenters is directly proportional to the workload.

The total energy consumed by the cloud datacenters throughout the world is estimated to be about 1.4% of the global energy consumption. It is gradually increasing with a rate of 12% per year [1]. A Barcelona medium size datacenter consumes energy equal to the power consumption of 1,200 houses annually [2] and pays a bill of £1 million for a year [3]. Environmental Protection Agency (EPA) of USA has noted to Congress that servers consume 40% of the total energy consumed by the datacenter [4]. Other devices in the datacenter like transformers, air conditioners, lighting, pumps and distribution wires, etc., consume the rest of the power. Hence, the servers are the primary consumers of energy in the cloud datacenters. Low server utilization, idle power waste and off course cramped implementation of energy efficient solutions are the prime reasons for such huge energy consumption. Dynamic voltage/frequency scaling (DVFS), powering down of servers or sleep scheduling and virtual machine (VM) consolidation are the various strategies applied in the cloud datacenters to minimize energy consumption [5]. Workload consolidation onto a smaller number of servers in cloud datacenters by selecting the most energy efficient ones is also a key technique of saving power [6]. Efficient load balancing techniques can also help in reducing power consumption in the datacenters. Balancing the load is a process of finding overloaded nodes and then migrating their extra load to the under-loaded nodes, in order to improve the overall resource utilization and performance of the system [7].

This paper has been catalogued as follows. In Sect. 2, the related work based on the research works done on minimizing energy consumption, while balancing the overall load in the cloud datacenters, is presented. Next, Sect. 3 of the paper states the proposed work on energy-efficient load balancing. Further, Sect. 4 contains the simulation and experimental results. Finally, in Sect. 5, the conclusion and the future work of this paper have been discussed.

2 Related Work

Various techniques and models have been introduced to enhance the energy efficiency of the datacenters in past years. Beloglazov et al. [8] identified certain causes of lofty energy consumption in datacenters and illustrated the taxonomy for future improvements. A cost model with measurable metrics for the calculation of total cost incurred in cloud environment has been proposed by Li et al. in [9].

Shang et al. [10] portrayed the DVFS approach for turning servers on/off to adjust the status of servers' power by maintaining the CPU energy consumption according to the workload. Buyya et al. [11] noticed a correlation between the utilization of CPU energy and the workload with respect to time. Bohra et al. [12] also proposed a model to witness the correlation between the components utilization and the system's power consumption. Chen et al. [13] treated a single running task in the cloud as the elementary unit for energy profiling and found out that the total energy consumed by two tasks is merely not equal to the summation of individual consumed energy because of scheduling overhead.

Le et al. [14] implemented a load balancing technique for the dynamic cloud environment by taking care of load placement as well as migration in the cloud datacenters, to save power. Panda et al. [15] mapped the tasks to the VMs and further mapped the VMs to the physical servers to minimize the energy consumption of the cloud datacenters. Panda et al. [16] also proposed an algorithm to balance the energy consumption and make span to resolve the issues associated with task scheduling and consolidation. This paper has considered the cuckoo search and multi-rumen anti-grazing technique to balance work load in dynamic cloud environment both intra-datacenter and inter-datacenter wise, respectively, to minimize energy consumption.

The female cuckoos in some species lay eggs, which resemble the color and pattern of eggs of some chosen bird species. These cuckoos often search and choose a nest of another bird, where the bird just laid its own eggs. They lay their eggs in the nest of the host birds to be hatched and to share food provided by host [17]. Here, in this case to allocate appropriate VM in cloud environment to the tasks provided by the clients as well as the migrated tasks from the overloaded VMs, while balancing the entire system load, the cuckoo search is used as a valid nature-inspired technique. The tasks to be allocated are considered as eggs of the cuckoos, and the VMs of the cloud are treated as the nest of host birds.

Ruminants such as cattle usually have a stomach divided into four compartments to store and chew partially digested food again and again. Although biologically the first chamber of the stomach is called as rumen, but researchers using the concept of multi-rumen as multiple chambers of the stomach treated it as multiple processing units. Sharma et al. [18] illustrated multiple datacenters as multi-rumen to execute huge amount of tasks, without overloading the VMs of a single datacenter. The tasks will be allocated suitably to the different VMs of multiple datacenters in order to make the total system load balanced [19]. But these algorithms are suitable for static environment in cloud only, which is in reality not so feasible. Therefore, in this paper, authors have integrated the concept of multi-rumen anti-grazing with cuckoo search to meet the requirement to balance load in dynamic environment in a robust way.

3 Proposed Work

In a dynamic cloud environment, tasks from the users enter into the datacenter at different instants of time for execution. Hence, the objective is to assign the tasks appropriately, so that all the VMs in that datacenter will be balanced, in order to avoid energy waste due to overutilization or underutilization of the resources. If more tasks are still entering into the datacenter, while the VMs are executing the tasks in their optimum capacity level; then instead of creating more VMs in the same datacenter, the extra tasks can be migrated to the underloaded running VMs of another datacenter. Creating a new VM and initiating execution on it will no doubt take time and will also consume more energy as compared to a running VM.

Here, we have considered a set of m number of tasks $T = \{T_0, T_1, T_2, T_3, \dots, T_{m-1}\}$, where each task is T_k , such that $0 \leq k \leq m - 1$. A datacenter contains a set of n number of virtual machines $\text{VM} = \{\text{VM}_0, \text{VM}_1, \text{VM}_2, \text{VM}_3, \dots, \text{VM}_{n-1}\}$, where each VM is VM_i , such that $0 \leq i \leq n - 1$. We have considered p number of datacenters $\text{DC} = \{\text{DC}_0, \text{DC}_1, \text{DC}_2, \text{DC}_3, \dots, \text{DC}_{p-1}\}$, where each DC is DC_j , such that $0 \leq j \leq p - 1$.

Before allotting the tasks to the virtual machines (VMs) of a datacenter, we should be aware of the capacity of VMs and the capacity of the datacenter as a whole.

Capacity of a VM ($C_{\text{VM}(i)}$) can be calculated as follows:

$$C_{\text{VM}(i)} = \text{VM}_{(i)\text{mips}} + \text{VM}_{(i)\text{bw}} \quad (1)$$

where $\text{VM}_{(i)\text{mips}}$ is the million instructions per second of a VM, and $\text{VM}_{(i)\text{bw}}$ is the communication bandwidth of the VM.

Hence, capacity of a datacenter ($C_{\text{DC}(j)}$) is the summation of the capacities of all VMs in that datacenter.

$$C_{\text{DC}(j)} = \sum_{i=0}^{n-1} C_{\text{VM}(i)} \quad (2)$$

Before allotting the tasks to any VM, we need to calculate the load of the VMs, so that proper load balancing can be done.

Load on a VM is the total length or size of the tasks S_T that are assigned to that VM at a particular instant of time. So, load on a VM can be calculated as per the following formula.

$$L_{\text{VM}(i)} = \sum_{k=0}^{m-1} S_{T(k)}(t) \quad (3)$$

Load of all VMs in a data center is calculated as

$$L_{\text{DC}(j)} = \sum_{i=0}^{n-1} L_{\text{VM}(i)} \quad (4)$$

Standard deviation of load can be calculated as given below

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{L_{\text{VM}(i)}}{C_{\text{VM}(i)}} - \frac{L_{\text{DC}(j)}}{C_{\text{DC}(j)}} \right)^2} \quad (5)$$

If standard deviation (σ) is less than or equal to the threshold value (T_s) [0–1], then the system is balanced. Otherwise, the VMs/datacenters may be overloaded or underloaded.

If $L_{\text{DC}(j)} < C_{\text{DC}(j)}$, then load balancing is possible; otherwise, according to multi-rumen anti-grazing principle, tasks will be migrated to other datacenter.

In a single datacenter, the VMs will be grouped into three categories, such as overloaded (OLVM), balanced (BVM) and underloaded (ULVM). Tasks from overloaded VMs will migrate to underloaded VMs.

Now, we need to evaluate the fitness function of the tasks (cuckoo's eggs) and the underloaded VMs (host nests). Here, we have calculated the priorities of the tasks as the fitness function of them to be allocated in the most suitable VMs.

Priority of a task ($P_{T(k)}$) can be evaluated as follows:

$$P_{T(k)} = \frac{\text{WT}_{T(k)}}{(S_{T(k)} + \text{GR}_{T(k)})} \quad (6)$$

where $\text{WT}_{T(k)}$ is the waiting time of the task in the ready queue before allocation to a particular VM. $S_{T(k)}$ is the length or size of the task. $\text{GR}_{T(k)}$ is the generation rate of tasks dynamically in the cloud environment.

The underloaded VM, which will be the most appropriate one for the task to be tagged with, is decided by the number of prioritized tasks in that VM. If the number of high priority tasks in an underloaded VM is minimum, then that VM will be chosen as the most appropriate one for a migrating task with high priority as it will require less time to be executed.

Algorithm of Hybrid Cuckoo Search for Load Balancing (HCS-LB)

```

Calculate  $C_{VM(i)}$ ,  $C_{DC(j)}$ ,  $L_{VM(i)}$ ,  $L_{DC(j)}$  based on the equations 1,
2, 3, 4 respectively.

Calculate standard deviation ( $\sigma$ ) as given in equation 5

If  $\sigma \leq Ts$ 
    System is balanced
    Exit

If  $L_{DC(j)} > C_{DC(j)}$ 
    Tasks will migrate to another datacenter
Else
    Call Load_Balancing( )

Load_Balancing( )
{
    Group VMs in datacenter DC(j) as ULVM, BVM and OLVM
    For each VM(i) in OLVM group
        Delete k-1 task among k number of tasks in the ready
        queue of VM(i) and add it to open queue for migration
        Calculate  $P_{T(k)}$  of each task k in open queue as in
        equation 6
        Arrange  $P_{T(k)}$  in descending order in open queue
        Calculate number of high priority tasks in VM(i) in
        ULVM
        Find  $P_{VM_{min}}$ i.e., VM(i) in ULVM containing minimum number
        of high priority tasks
        Arrange VM(i) in ULVM group in ascending order
        Migrate each task from open queue sequentially to be
        allocated to each VM(i) in ULVM
}

```

Degree of imbalance (DI) is calculated as

$$DI = \frac{T_{\max} - T_{\min}}{T_{\text{avg}}} \quad (7)$$

where T_{\max} and T_{\min} are the maximum and minimum T_i among all VMs, and T_{avg} is the average T_i of all VMs.

The total energy consumption is the sum of energy consumed in the idle state and active state of the VM. Mathematical representation of the second fitness function in terms of total energy consumption is defined as follows:

$$EC_{DC(j)} = \sum_{i=0}^{n-1} [T_i \times \alpha_i + (MS - T_i) \beta_i] \times PS_i \quad (8)$$

where

T_i = Amount of time i th VM remained in active state.

α_i = Joules/Millions of instructions consumed by i th VM in the active state.

$MS - T_i$ = Amount of time i th VM remained in idle state.

β_i = Joules/Millions of instructions consumed by i th VM in the idle state.

PS_i = Processing time of i th VM in the cloud.

$EC_{DC(j)}$ = Energy consumed by a datacenter.

$$EC = \sum_{j=0}^{p-1} EC_{DC(j)} \quad (9)$$

where

EC = Total energy consumed by all datacenters.

4 Simulation and Result Analysis

We have used CloudSim installed on Intel® Core™ system with i5-4210 M CPU@1.70 GHz processor, 8 GB memory and Windows 7 64-bit operating system for the simulation of the proposed HCS-LB algorithm and some of the existing algorithms, namely dynamic load balancing (DLB) and honey bee behavior-inspired load balancing (HBB-LB) algorithm for comparison.

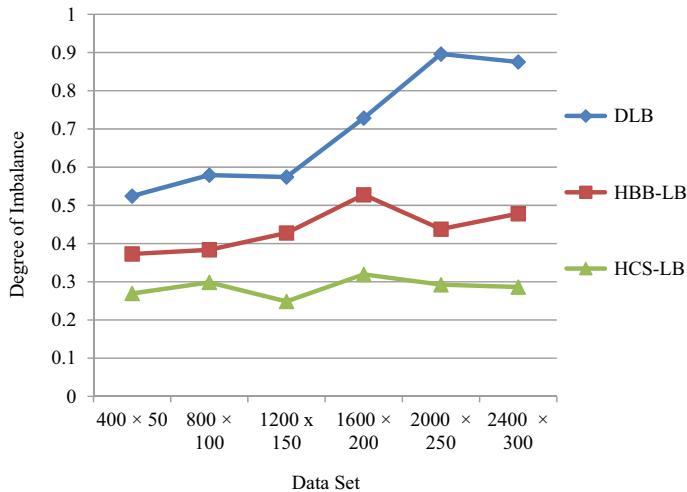
We have generated six different datasets, such as 400×50 (i.e., 400 specifies the number of tasks or user requests, and 50 indicates the number of VMs), 800×100 , 1200×150 , 1600×200 , 2000×250 and 2400×300 . Each dataset contains three instances, namely $i1$, $i2$ and $i3$ of same size.

From Table 1, we observed that the degree of imbalance (calculated as per the Eq. 7) in HBB-LB is less as compared to the degree of imbalance in DLB. As we introduced inter-datacenter task migration to balance the loads on VMs in HCS-LB, the degree of imbalance is reduced further resulting in a more stable system with higher degree of load balancing.

The comparison of degree of imbalance in between the three different algorithms DLB, HBB-LB and HCS-LB is portrayed in Fig. 1, where X-axis is representing the

Table 1 Degree of imbalance in DLB, HBB-LB and HCS-LB for different datasets

Dataset	Degree of imbalance in DLB	Degree of imbalance in HBB-LB	Degree of imbalance in HCS-LB
400×50	0.5243	0.3728	0.2693
800×100	0.5792	0.3837	0.2985
1200×150	0.5743	0.4276	0.2485
1600×200	0.7284	0.5278	0.3193
2000×250	0.8962	0.4379	0.2923
2400×300	0.8754	0.4783	0.2861

**Fig. 1** Degree of imbalance in DLB, HBB-LB and HCS-LB for different datasets

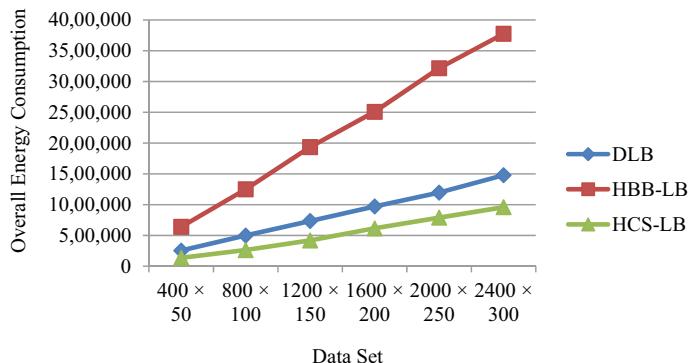
dataset, i.e., number of tasks \times VMs, and the Y -axis is representing the degree of imbalance.

From Table 2, we found the result that the overall energy consumption (calculated as per the Eq. 9) in HCS-LB is comparatively very low with regard to the overall energy consumption in DLB and HBB-LB, as examined over six sets of data on CloudSim.

The overall energy consumption in the three different algorithms, namely DLB, HBB-LB and HCS-LB, is presented in Fig. 2, where X -axis is representing the dataset, i.e., number of tasks \times VMs, and the Y -axis is representing the overall energy consumption. Hence, we found that our proposed load balancing mechanism is reducing the degree of imbalance and overall energy consumption in comparison with existing DLB and HBB-LB algorithm.

Table 2 Comparison of overall energy consumption in DLB, HBB-LB and HCS-LB

Dataset	Overall energy consumption in DLB	Overall energy consumption in HBB-LB	Overall energy consumption in HCS-LB
400 × 50	254,274	639,999	136,467
800 × 100	499,424	1,249,969	262,899
1200 × 150	734,763	1,935,749	418,239
1600 × 200	971,400	2,508,121	614,267
2000 × 250	1,195,547	3,217,328	789,548
2400 × 300	1,478,385	3,776,395	958,762

**Fig. 2** Comparison of overall energy consumption in DLB, HBB-LB and HCS-LB

5 Conclusions and Future Work

The proposed algorithm HCS-LB outperforms over HBB-LB and DLB algorithms in terms of load balancing and energy consumption as well. The cuckoo search mechanism efficiently provisions the appropriate VM to the migrated tasks from the overloaded VMs, on priority basis. The multi-rumen anti-grazing principle helps in allocating the tasks to the VMs in a separate datacenter, when the VMs of the datacenter native to the tasks become overloaded or balanced. Along with balancing the load and making the overall system stable, this approach also proves to be successful in reducing the energy consumption in the datacenters. In future course of action, we will work on designing more energy-efficient load balancing algorithms in dynamic cloud environment to contribute some toward the huge requirement in this field and will move forward to implement renewable energy sources as well.

References

1. BONE project (2009) WP 21 Tropical project green optical networks: report on year 1 and update plan for activities. No. FP7-ICT-2007-1216863 BONE project
2. Kogge P (2011) The tops in flops, pp 49–54. IEEE Spectrum (2011)
3. Google documents. <https://docs.google.com/>
4. Report to congress on server and datacenter energy efficiency, environmental protection agency (2007)
5. Liu J, Zhao F, Liu X, He W (2009) Challenges towards elastic power management in internet data centers. In: Int Conf Distrib Comput Syst workshops. Proc. IEEE, Los Alamitos, CA, pp 65–72
6. Srikanthaiah S, Kansal A, Zhao F (2008) Energy aware consolidation for cloud computing. In: Proceedings of the 2008 conference on power aware computing and systems. HotPower'08. USENIX Association, Berkeley, CA, USA, pp 10–10
7. Kaur R, Luthra P (2014) Load balancing in cloud computing. In: Proc Int Conf on recent trends in information, telecommunication and computing, pp. 374–381
8. Beloglazov A, Buyya R, Lee YC, Zomaya A (2011) A taxonomy and survey of energy-efficient datacenters and cloud computing systems. In: Adv Comput. Elsevier 82:47–111
9. Li X, Li Y, Liu T, Qiu J, Wang F (2009) The method and tool of cost analysis for cloud computing. In: IEEE international conference on cloud computing (CLOUD 2009), Bangalore, India, pp. 93–100
10. Shang L, Peh LS, Jha NK (2003) Dynamic voltage scaling with links for power optimization of interconnection networks. In: 9th international symposium on high-performance computer architecture (HPCA 2003). Anaheim, California, USA, pp 91–102
11. Rajkumar B, Anton B, Jemal A (2010) Energy efficient management of datacenter resources for cloud computing: a vision architectural elements and open challenges. In: Proc. international conference on parallel and distributed processing techniques and applications (PDPTA 2010). Las Vegas, USA
12. Bohra CV (2010) VMeter: Power modelling for virtualized clouds. In: International symposium on parallel & distributed processing, workshops and Phd forum (IPDPSW), 2010 IEEE, vol., no., pp 1–8 (2010)
13. Chen F, Schneider J, Yang Y, Grundy J, He Q (2012) An energy consumption model and analysis tool for cloud computing environments. Green and sustainable software (greens). 2012 first international workshop, pp 45–50 (2012)
14. Le K, Bianchini R, Zhang J, Jaluria Y, Meng J, Nguyen TD (2011) Reducing electricity cost through virtual machine placement in high performance computing clouds. In: Proceedings of 2011 international conference for high performance computing, networking, storage and analysis ACM, p 22
15. Panda SK, Jana PK (2017) An efficient request-based virtual machine placement algorithm for cloud computing. In: International conference on distributed computing and internet technology. Springer, pp 129–143
16. Panda SK, Jana PK (2019) An energy-efficient task scheduling algorithm for heterogeneous cloud computing systems. Cluster Comp 22(2):509–527
17. Yang XS, Deb S (2009) Cuckoo search via Lévy flights in nature & biologically inspired computing, NaBIC 2009, World Congress, 2009. IEEE
18. Sharma SCM, Rath AK (2017) Multi-rumen anti-grazing approach of load balancing in cloud network. International Journal of Information Technology, Springer
19. Kulkarni K, Annappa B (2015) Load balancing strategy for optimal peak hour performance in cloud datacenters. In: Proceedings of IEEE international conference on signal processing, informatics, communication and energy systems

Analyzing Intake of Water by a Human Using Meteorological Conditions



M. Rekha Madhuri, T. Sriya Datta, M. Nithisha, and Y. Sandeep

Abstract Water is a vital and essential compound to run a human body. It regulates human body temperature and also performs different types of operations in the human body. The water level in the human body depends on weight, height, gender, and age where there would be a difference in the quantity of water in human bodies. The dehydration level for the human body is different for everyone where it changes according to the work done (like daily workouts), body conditions, and atmospheric conditions. The main goal of the project is to determine the intake of water for human in a day where it takes the inputs as a person age, weight, gender, height and also takes inputs of atmospheric conditions like humidity and temperature where it changes from place to place, and it depends on the hydration in the human body. By taking these inputs, it calculates water by using water balance cognition methods. The temperature and humidity are calculated by the IoT device called DHT11 sensor and Arduino Ethernet shield and stored in Xampp server for further processing. After calculating temperature and humidity, all the inputs are processed using water balance cognition methods and then it gives the quantity of water to be consumed by a human as the output.

Keywords IoT devices · DHT11 sensor · Arduino Ethernet shield · Xampp server · Water balance cognition methods

M. R. Madhuri · T. S. Datta (✉) · M. Nithisha · Y. Sandeep
Y. Sandeep Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College,
Kanuru, Vijayawada, India

1 Introduction

The main idea of the problem raised when there is an online webinar on drinking water by make a difference (MAD) organization in the college. They have spoken about many diseases related to improper intake of water and also suggested drinking more and more water. That is the starting step of the project and decided to design the project which will give an accurate amount of water to be consumed by an individual human on a particular day also based on different conditions. From this project, many people who do not know how much water to be consumed will get awareness to drink water so that they can get rid of many diseases.

2 Basic Definitions and Background

There are many technologies used to predict the consumption of water, but the easiest and effective way is using IoT devices for atmospheric conditions and using different water balance cognition methods for water intake.

2.1 *Internet of Things*

Internet of things (IoT) focuses on network of physical objects called “things” which are embedded with software, sensor, and many other technologies to rope and traffic data with other systems over the Internet.

Many technologies nowadays are evolved due to the combination of multiple technologies and real-time analysis. Major fields like automation, control systems, embedded systems contribute to IoT [1].

In the current market, the smart home is a concept based on IoT which automates the devices and all other electrical appliances. This smart home includes security cameras, automated lights, and these can be controlled via devices associated with them, such as smartphones.

IoT can also be used in healthcare systems. Several issues are arising in the market due to the rapid growth in IoT technology mostly in the security and privacy sector.

2.2 *Water Balance Cognition Methods*

There are many methods in which the amount of water taken is calculated by taking different inputs; but in this method, the inputs slightly differ from others and this makes the cognition methods more flexible to use for different scenarios. Here

the inputs of the person like age, weight are taken, and inputs of atmosphere like temperature and humidity are taken and are processed using these methods.

This empirical method focuses on the previous day's intake of water and recalculates for every user differently according to their given inputs and their atmospheric conditions. This also contains many sub-methods which calculate the intensity of a neuroendocrine response employed by the brain to calculate total body water concentration and volume [2].

2.3 Problem Statement

This project focuses on the water consumption of an individual using IoT devices like the DHT11 sensor. It also provides knowledge on the consumption of water based on meteorological conditions and factors that are necessary for a human body. Improper water consumption may cause a heavy burden on human body performance, energy, weight, functioning of the different organs, and overall health and also leads to hyponatremia [2]. Therefore, the balanced consumption of water could be very essential for the body. The main goal of this project is to give inputs to a person on the quantity of water consumed on daily basis.

2.4 Real-Time Applications of Proposed Work

There are many applications where this water intake is extremely useful for every individual human for enjoying their life in a better way. Some are mentioned below:

- Persons who are suffering from typical diseases like urinary tract infections can be cured by drinking an ample amount of water.
- Persons who are affected with kidney stones and other related diseases related to kidneys are also be cured by sequentially consuming water.
- Physical exercises can make the body fit also while doing exercises we need to consume more water.
- Drinking an ample amount of water makes a person stronger both physically and mentally.

3 Summary of Literature Survey

From the literature survey, we observed the following. All the different papers we have referred for our project which have guided us on only one side, that is either on the weather monitoring or in the water intake of the person. Our project is the combination of both parameters. Some of the papers only show it is hydrated or dehydrated but does not show any quantity of water. In some papers, they have

used machine learning, and some used more tools like electrodes to connect that are connected to the body [3] that is difficult to make it where we have used the minimal technology for better efficiency.

We have used water balance cognition methods for better calculation of water quantity in the human body and also used the sensors for better calculation of temperature and humidity which gives different values from place to place. By taking all these inputs, it gives the better result of quantity of water that needs to be consumed by a human in a day. We also give notifications to their devices for taking water at particular intervals of time.

4 Proposed Method

4.1 Design Methodology

Figure 1 diagram shows that the user gives the input like age, gender, weight, height from the user, and those are processed with the help of Xampp server. Later we take inputs from the DHT11 sensor which calculates the temperature and humidity from the atmospheric conditions that are stored in the Xampp server for further process [4].

The values that are stored in Xampp are used for calculating water level using the water cognition method it calculates and gives the result as the quantity of water needs to be taken by a human in a day.

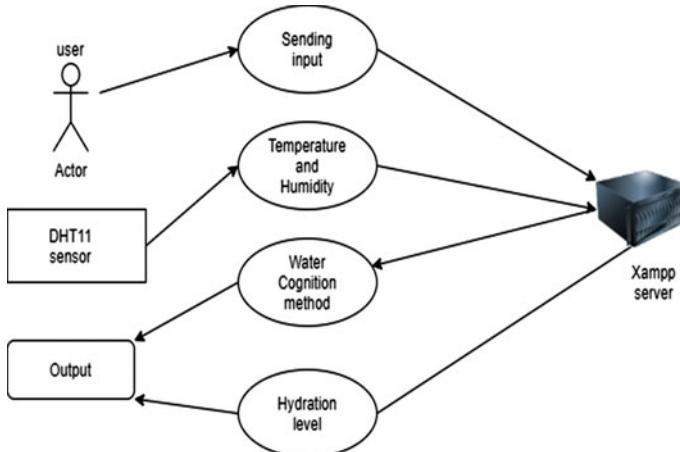


Fig. 1 UML representation

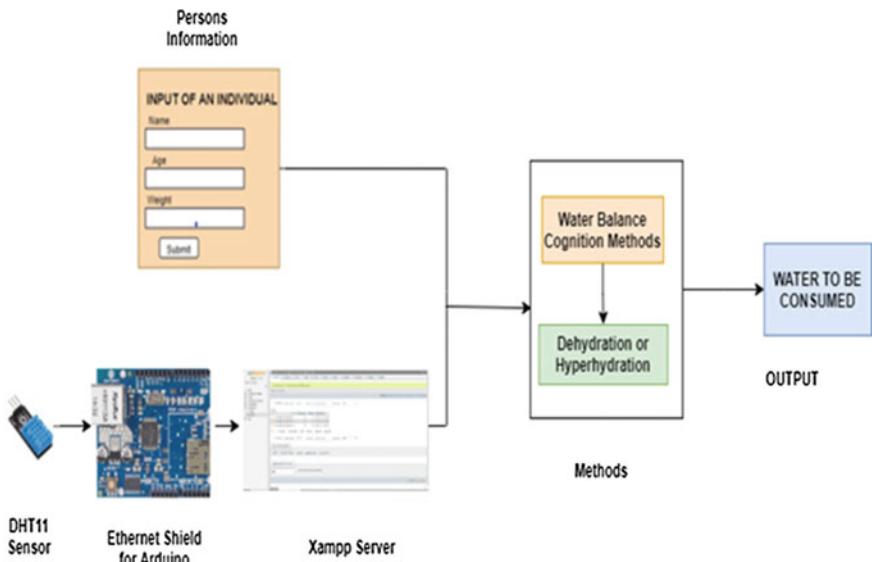


Fig. 2 Architecture diagram

4.2 System Architecture Diagram

In Fig. 2 diagram as shown in Fig. 1, we take inputs of a person like age, weight, height, gender. Later we collect information of atmospheric conditions like temperature and humidity using DHT11 sensor [5], and Ethernet shield for Arduino these IoT devices gives accurate results of temperature and humidity, and these values change from place to place and time to time these values are taken for every 1000 s.

These values are stored in the Xampp server the stored values of these are used for calculating using water cognition methods these helps in calculating the water levels in the human body and also used for hydration and dehydration levels. The final verdict is by taking all these factors, and the output is shown as some quantity of water needs to be consumed by a person in a day.

4.3 Description of Algorithm

How much quantity of water is to be consumed daily by a person? It is one of the most asked questions with not a simple answer. There are many studies, but none gave the correct answer and also gave varying recommendations over all these years [6]. But depending on one's individual health factors and activity level water should be consumed. This project is providing water need to be gulped with the following parameters (Table 1).

Table 1 EFSA and IOM reference for water intake[7, 8]

Age group	ESFA		IOM	
	Total water intake	Fluid intake	Total water intake	Fluid intake
0–6 months	0.68 (milk)	0.68 (milk)	0.70	0.70
6–12 months	0.80–1.00	0.64–0.80	0.80	0.80
1–2 years	1.10–1.20	0.88–0.90	–	–
2–3 years	1.30	1.00	–	–
1–3 years	–	–	1.30	0.90
4–8 years	1.60	1.20	1.70	1.20
9–13 years Boys	2.10	1.60	2.40	1.80
9–13 years Girls	1.90	1.50	2.10	1.60
Boys 14+ and adult men	2.50	2.00	3.30	2.60
Girls 14+ and adult women	2.00	1.60	2.30	1.80
Pregnant women	2.30	1.84	2.60	1.90
Lactating women	2.60	2.10	3.40	2.80

1. Gender

The Institute of Medicine agency has set daily adequate intake levels (AI) for fluids. The AI for adult men is 3 L approximately 13 cups a day. The AI for adult women is 2.2 L approximately 9 cups a day. The AI values include all types of fluids such as juices and shakes. All these values are applicable for a healthy person with no long-time diseases and at normal temperature. The intake will vary depending on factors [7].

2. Age

European Food Safety Authority (ESFA) and US Institute of Medicine (IOM) data to take reference water intake are in L/day (liters per day) [7].

3. Weight

One easiest way to judge your fluid needs is by weight. Every kilogram requires nearly 30 to 40 mm of water.

If you weigh 50 kg, then you need at least 1.5–2 L of water per day.

For fulfilling the fluid balance in the body, one needs to drink water that matches their individual weight, environmental condition, and activity level in the same day.

4. Activity Level

While doing workouts, a general rule is to consume 500 ml of water or approximately two cups, before half an hour to get start your workout and drink 250 ml of fluid for every 15 min during the workouts.

The main idea here is to get completely hydrated before the start of workout. So that you have a good electrolyte balance. Drinking a little amount of water during the workouts helps in maintaining body hydration. The motivation behind getting hydrated before a workout is to help to recover hydration. Without proper rehydration, after the workout there is a risk of becoming dehydrated and delaying recovery [8].

5. Weather

Climate is also one of the factors to be considered when it comes to hydrating. In temperatures like moist and hot weather conditions, you'll have more sweat, which eventually you need to intake extra fluid [8]. Higher altitudes may lead to dehydration. Overview of health also plays a major role in your requirement to water intake. Diarrhea, vomiting, and also fevers will result in the loss fluid, which requires repossession through the additional fluid intake.

- First, determine one's weight. Based on the weight, the water intake drastically varies. For example, if you are obese, then you need to drink more water compared to others [9].
- Second, how much water you need to drink is calculated by multiplying your weight by 2/3 [9].
- Third, next important aspect is your activity level. The more you work out, the more you will be dehydrated [9].
- Lastly based on the climatic conditions mentioned above, need to increase or decrease the water intake by at least 10% [9].

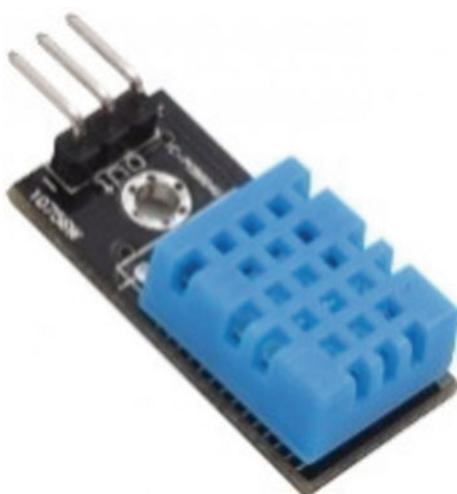
4.4 Description of Tools

1. DHT11 Sensor

The DHT11_sensor in Fig. 3 is a temperature, humidity sensor that comes with negative temperature coefficient (NTC) which is used to measure temperature and 8-bit microcontroller and gives the output temperature and humidity [10]. The sensor measures temperature in the range of $0^{\circ}\text{C} \geq t \leq 50^{\circ}\text{C}$ with an accuracy of $\pm 1^{\circ}\text{C}$ and humidity in the range of $20\% \geq h \leq 90\%$ with an accuracy of $\pm 1\%$. So, this sensor provides us with accurate room temperature and humidity.

2. Arduino UNO

The Arduino UNO in Fig. 4 is a microcontroller with ATmega328 as a controller. This Arduino board is mainly used for electronics project; but in this project, it will help to retrieve the sensor data [10]. The Arduino UNO board is commonly called as Arduino board. The board consists of 14 digital (input/output) pins, where 6 are analog input pins, USB connector, one power jack, ICSP header, one reset button, and other components.

Fig. 3 DHT11_sensor**Fig. 4** Arduino UNO

All these components are fixed subsequently to the Arduino board to function well. The board is charged with the help of USB port or directly charged with the DC supply to the board.

3. Ethernet Shield

Ethernet shield in Fig. 5, which is used with Arduino technologies, is a computer component. With the help of Ethernet cable and LAN wire, these components help a device to connect with internet. It will help us to send and receive information over the Internet.

Fig. 5 Ethernet Shield

It contains the main element called as jack with which an Ethernet shield be used to plug into any other device. Along with few other circuits, this jack is also attached to a small board [11]. The circuits connected make the Ethernet shield to communicate to the attached device.

Most of the Ethernet shields are comprised of a standard feature that lets them reset along with their normal power cycle of this device. Some of the Ethernet shields come with storage device slots.

4. Xampp Server

XAMPP server is an open-source as well as cross-platform web server that mainly contains Apache HTTP server. It interpreters for scripts that are covered by the PHP and Perl programming languages [12].

XAMPP server is adjusted to transmit from a local test server to a live server possible since most actual web server deployments mainly use the same components as XAMPP.

The DHT11 is a temperature and humidity sensor. Using Arduino, the values of temperature and humidity are listed in Arduino IDE. To store the values obtained, an Ethernet shield for Arduino is inserted on top of the Arduino board [13].

Using the XAMPP server, start the apache and MySQL to store the values from the Ethernet shield into the database [13]. Values from the IoT are stored in the MySQL database with the help of PHP. Collecting the information of an individual is using a form to calculate the water intake.

5 Results

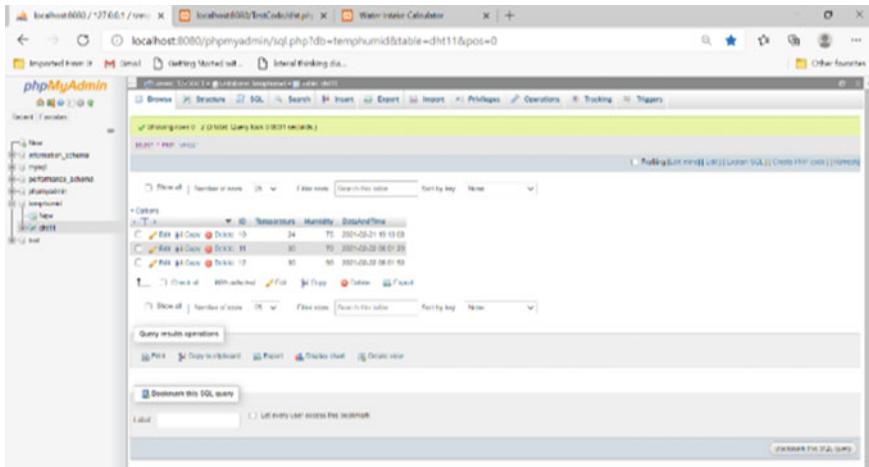


Figure 6 depicts the temperature and humidity values collected into the Xampp server with the help of Ethernet shield, Arduino board, and DHT11 sensor.

In Fig. 7, the inputs of an individuals are collected through a form with the help of Xampp server.

1. Input–Output(1)

Figures 8 and 9 explain that if gender is male, age about 25, weight is up to 54 and temperature is 24 deg centigrade, then the liters per day need to take will be 2.35.

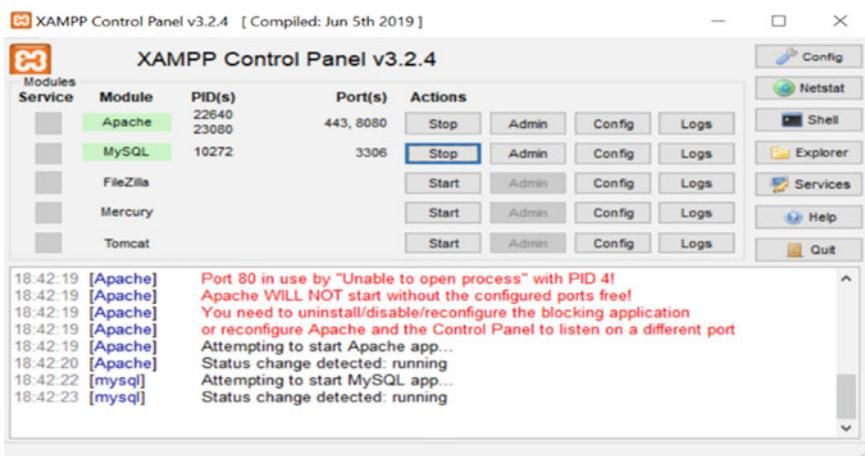


Fig. 6 Xampp server

The screenshot shows a web browser window with the URL `localhost:8080/WATER1.html`. The title of the page is "Water Intake Calculator". The form contains the following fields:

- Gender:** Male (selected)
- Age:** (empty input field)
- Weight:** (empty input field) kg (dropdown menu)
- Activity Level:** Sedentary (selected)
- Sweat:** Light Sweat (selected)
- Calculate:** (button)

Fig. 7 Inputs of an individual

The screenshot shows the same web browser window with the URL `localhost:8080/WATER1.html`. The "Age" field now contains "30". The "Weight" field contains "90" with "kg" selected in the dropdown. The "Activity Level" is set to "Light Active" and "Sweat" is set to "Moderate Sweat". All other fields remain the same as in Fig. 7.

Fig. 8 Input-1

The screenshot shows a web browser window with the URL `localhost:8080/try.php`. The main content of the page is the text "2.35lts".

Fig. 9 Output-1

2. Input–Output(2)

Figures 10 and 11 explain that if gender is male, age about 13, weight is up to 52 and temperature is 30 deg centigrade, then the liters per day need to take will be 2.1.

3. Input–Output(3)

Figures 12 and 13 explain that if gender is male, age about 30, weight is up to 90 and temperature is 30 ° centigrade, then the liters per day need to take will be 3.9 L.

localhost:8080/WATER1.html

Imported From IE Gmail Getting Started wit... lateral thinking dia...

Water Intake Calculator

Gender: Male

Age: 13

Weight: 52 kg

Activity Level: Sedentary

Sweat: Moderate Sweat

Calculate

Fig. 10 Input-2

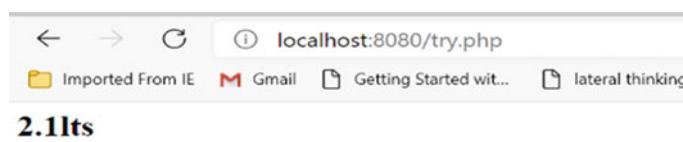


Fig. 11 Output-2

localhost:8080/WATER1.html

Imported From IE Gmail Getting Started wit... lateral thinking dia...

Water Intake Calculator

Gender: Male

Age: 30

Weight: 90 kg

Activity Level: Light Active

Sweat: Moderate Sweat

Calculate

Fig. 12 Input-3

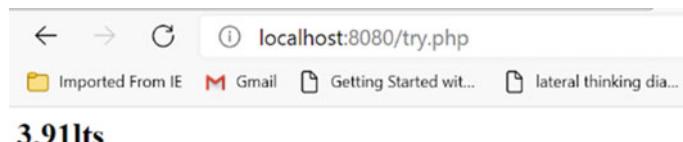


Fig. 13 Output-3

Water Intake Calculator

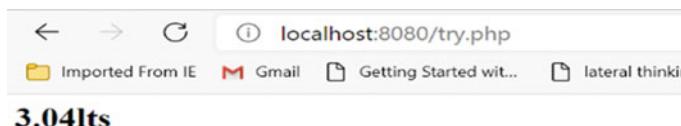
Gender
Male

Age
25

Weight
54 kg

Activity Level
Sedentary

Sweat
Light Sweat

Fig. 14 Input-4**Fig. 15** Output-4

4. Input–Output(4)

Figures 14 and 15 explain that if gender is male, age about 35, weight is up to 70 and temperature is 32 ° centigrade, then the liters per day need to take will be 3.04 L.

6 Analysis and Discussion

This simple water intake calculator will help you to approximate the quantity of liquid water that a body should be consumed in a day based on meteorological condition, your age, weight, and activity level. In classic way, this project gives you an idea, how many ounces of water you should drink a day with the help of this calculator to steer clear of dehydration.

So, the best practice is to follow the water intake calculator. The absolute hydration specification is provided with this water intake calculator to steer clear of dehydration. Drinking an adequate amount of water makes you to get the right nutritional balance [14].

Always keep in mind that drinking an ample amount of water will assist one's body to have a perfect nutritional balance, stay hydrated, and allow body to process smoothly.

Water in the human body helps to digest and absorb the food, transport nutrients, maintains one's body temperature, and helps in good circulation [15]. Hence, to boost

your body function to get how much water you need to gulp in a day by entering the parameters needed by our page and click on calculate button at the bottom.

7 Conclusion

The main objective of the project is to develop awareness among every individual in consumption of water for better life also directing them in the right path on how much quantity needed to be consumed for them on daily basis. The water consumption is calculated on the basis of the person's factors like age, weight, gender, height, and atmospheric conditions like temperature, and humidity.

References

1. Joseph J, Joe F (2019) IoT based weather monitoring system for effective analytics. *Int J Eng Adv Technol* 8(4)
2. Bera TK (2019) Bioelectrical impedance methods for noninvasive health monitoring. *J. Med. Eng.* 14(7)
3. Yao A, Yang CL, Seo JK, Soleimani M (2017) EIT-based water pressure sensing. *Comput. Math Methods Med* 9(3)
4. Soleimani M, Seo JK (2016) Electrical impedance spectroscopy for electro-mechanical characterization sensor. *Int J Scent Res Comput Sci* 14(6)
5. Shirode M, Adaling M, Biradar J, Mate T (2018) IOT based water quality monitoring system. *Int J Scent Res Comput Sci Eng Inform Technol* 3(1)
6. Abubaker A, Kavya CR, Thomas T, Joseph N, Begum S (2018) A study on IOT approach for monitoring water quality using MQTT algorithm. *Int Res J Eng Technol* 5(3)
7. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2010.1459>
8. <https://ncbi.nlm.nih.gov/pmc/articles/PMC4473081/>
9. VivekBabu K, Anudeep Reddy K, Vidhyapathi CM, Karthikeyan B (2017) Weather forecasting using raspberry pi with internet of things (IOT). *ARPJ Eng* 12(17), September 2017
10. Sandeep V, Gopal KL, Naveen S, Amudhan A, Kumar LS (2015) Globally accessible machine automation using Raspberry pi based on Internet of Things. *Adv Comput Commun Inform (ICACCI)* 19(2), 2015
11. Hoffer EC, Simpson DC, Meador CK () Correlation of whole-body impedance with total body water volume. *J Appl Physiocol* 27(4), October 2018
12. Stewart R, Reed J, Zhong J, Morton K, Porter T (2014) Human hydration level monitoring using embedded microcantilever sensors. *Med Eng Phys* 9(8), April 2014
13. Daniels MC, Popkin BM (2015) Impact of water intake on energy intake and weight. *Int J Sci Res* 16(5), September 2015
14. Sen D, Dey M, Kumar S, Boopathi CS (2016) Smart irrigation using IoT. *Int J Appl Sci Technol* 29(4), June 2016
15. Palma D, Agudo JE, Sánchez H, Macías M (2014) An internet of things example: classrooms access control over near field communication 7(5), April 2014

Driver Drowsiness Detection and Object Detection for Vehicles Using Yolov3



D. Sampath, V. V. Krishna Chaitanya, B. Dhanvanth Kumar, and G. Geetha

Abstract The objective is to detect objects using You Only Look Once (YOLO) approach. This method has several advantages as compared to other object detection algorithms. Far-off detecting information is a significant hotspot for making or refreshing geological data framework dataset and are utilized in various sorts of territories. Programmed division of distant detecting information into semantically classes has been a traditional issue from past many years and has numerous application zones like guides, route, checking of climate, arranging of metropolitan and country region, yet after such extraordinary advances, it actually needs more precision and effectiveness. The difficult assertion for this work is essentially playing out the scene division which can be an elevated picture of metropolitan and provincial region from far-off detecting information, so various constructions can be recognized like vehicles, streets, structures and so on This assignment is testing a result of the presence of profoundly various items or design or shapes which calls for earlier information apart from standard smoothness and co-occurrence assumptions. In this paper, we discuss a model that detects when the driver is getting into drowsiness and alerts him. It also allows him to visualize the lane on which vehicle is moving on and also deviations in the lane.

Keywords Remote sensing · Image segmentation · Pattern recognition · Landscape · YOLO dataset · Lane detection · Drowsiness detection

D. Sampath · V. V. K. Chaitanya (✉) · B. D. Kumar · G. Geetha
V.R. Siddhartha Engineering College, Kanuru, Vijayawada, Andhra Pradesh, India

1 Introduction

Vision-based driver-assistance systems (VBDAs) have emerged as a significant contribution for the improvement of advanced driver assistance systems (ADAS) and have had gradually gained a substantial meaning in people's life. Every year, approximately 1.35 million people die as a consequence of road traffic crashes. Also, around 20–50 million people undergo non-fatal injuries, including, sometimes, disabilities as an outcome. Takizawa et al. [1] Nowadays, we find car accidents and related deaths every day because of other people's carelessness (driver's drunkenness, drowsiness or sudden illness). Also according to several studies, 90% of traffic accidents are caused by a human error [1, 2]. However, machines when programmed properly are more reliable and consistent, and they commit far less mistakes. Therefore, they can prevent most of those mentioned fatalities.

This paper covers designing, developing and manufacturing a smart camera system for a Ford Focus that will be able to monitor both outer and inner conditions. Development includes researching the topic to design and produce a competent solution for a camera system that is a consumer-level product, which means that mountings and wiring are implied to car's original wiring as good as possible. An onboard computer for the system is located in the trunk, and wiring is done in a way that all the cameras are plugged in with a single universal cable. The monitoring system is capable of detecting 360° surroundings of outside of the car (tracking other traffic, pedestrians, etc.), as well as inside of the car by monitoring the drivers state (drowsiness, sudden illness, etc.) and objects not supposed to be in car. Therefore, the motivation of this project is the strong willingness of assuring safety and comfort behind the wheel that concerns both drivers and passengers, and vulnerable road users: pedestrians, cyclists and motorcyclists. Article discovery is an innovation that recognizes the semantic objects of a class in computerized pictures and recordings. One of its ongoing applications is self-driving vehicles. In this, our assignment is to recognize various items from a picture. The most well-known item to distinguish in this application is the vehicle, bike and passerby. For finding the items in the picture, we use object localization and need to find more than one article progressively frameworks. There are different methods for object location; they can be separated into two classifications: first is the calculations dependent on classifications. There have been parcel of works done in the space of pixel-wise arrangement of far-off detecting symbolism (i.e., low goal pictures). Nowadays, more consideration is going toward the item or construction identification in metropolitan regions in situations as high goal pictures. So for this, there can be a potential arrangement which is rule-based methodologies which plan redid rules to encode an earlier information for some particular classes. In this methodology, when an information picture is given, it characterizes picture pixels into various class names, then an over-complete arrangement of streets and building up-and-comer are produced. Then, these produced applicants are pruned to an ideal subset. Drowsiness and also uneven/unclear roads are some of the major reasons that lead to road accidents. So we came up with a model that detects when the driver is getting into drowsiness and alerts him. It also allows him

to visualize the lane on which vehicle is moving on and also deviations in the lane. This approach worked so well; now we came up with the extension of detecting the vehicles that come by the way the vehicle is moving on. For that, YOLO dataset is used. YOLO dataset is to find out the different objects that come along on the road and gives suggestions to the driver in which direction he has to move to pass that object that is coming by his way. This increases the accuracy of autonomous vehicles to drive safely even in a road that has various number of vehicles. So, this helps enhancing the road safety and prevent accidents.

In the previous work, the output was identifying the road lanes along the way. In the present work, the extension added was detecting the objects or obstacles and giving the direction to move. No dataset was used in the previous work, YOLOV3 dataset was used to recognize the objects in the present work. In the previous work, modules of canny were used. The efficiency is also increased comparatively since the obstacle is detected and the safe way is predicted. In real time, our algorithm processes 45 frames per second. YOLO algorithm makes localization errors but predicts less false positives in the background.

During the driving operation, humans use their optical vision for vehicle maneuvering. The road lane marking is considered as a constant reference for vehicle navigation. The development of an automatic lane detection system using an algorithm is one of the prerequisites to have in a self-driving car. Computer vision is a technology that can enable cars to make sense of their surroundings. It is a branch of artificial intelligence that enables software to understand the content of image and video. In this proposed topology, YOLO enables it to recognize different objects in images by examining and comparing millions of examples and cleaning the visual patterns that define each object. While particularly effective for grouping assignments, profound taking in experiences genuine impediments and can fizzle unpredictable. This implies that a driverless vehicle may collide with a truck without trying to hide, or more awful, inadvertently hit a walker. The current PC vision innovation utilized in self-governing vehicles is additionally powerless against ill-disposed assaults, by controlling the AI's information channels to compel it to commit errors. For example, specialists have shown that they can deceive a self-driving vehicle to abstain from perceiving stop signs by staying high contrast marks on them. You Only Look Once: Unified, Real-Time Object Detection, by Joseph Redmon. Their earlier work is on distinguishing objects utilizing a relapse calculation. To get high precision and great forecasts, they have proposed YOLO calculation in this paper [1]. Comprehension of Object Detection Based on CNN Family and YOLO, by Juan Du. In this paper, they for the most part clarified about the item discovery families like CNN, R-CNN and analyzed their productivity and acquainted YOLO calculation with increment the proficiency [2]. Figuring out how to Localize Objects with Structured Output Regression, by Matthew B. Blaschko. This paper is about Object Localization. In this, they utilized the bounding box technique for limitation of the items to beat the downsides of the sliding window strategy [3].

2 Literature Survey

Traffic Sign Classifier [1]: In this module, we investigated how traffic signs can be detected and recognized by the convolutional neural networks (CNN) model. CNN is a multi-layer neural network model, which is typically made up of two main parts. The first part contains alternating convolution and pooling. The second part is made up of connections and softmax classification. Increased accuracy in vision tasks and improved learning ability is one of the properties of CNN. Another factor in evaluating the effectiveness of this technology is its performance under strict time constraints.

Traffic signal recognition [2] is a central feature of autonomous vehicles, and considering the risks of a late classification, the time it takes for this process to be completed and the accuracy of the process itself are given equal importance. However, many papers that present research about their model do not provide the time taken for the process. In this paper, we observed a concise survey of recent research endeavors in traffic sign recognition (TSR), which implemented CNN architectures. The TSR systems consist of two parts: traffic sign detection and classification. The regions of interest are first extracted in detection and then identified correctly or rejected in classification.

In this paper [3], we observe how the CNN model has been employed in TSR systems. The recognition task consists of two parts: detection and classification, so we will divide the sections in this way: (1) CNN models used for detection in [1], the algorithm which is proposed is based on deep visual feature. This combines convolutional neural networks (CNN) and support vector machine (SVM). The characteristics of images can be extracted in many cases, and the four types of arrows can be identified accurately using this. It is because the process of recognition is based on the deep visual features; the detection process is performed in two steps: (1) collect and process the images, then determine the key frames of the images, and (2) the convolutional neural network (CNN) is applied to extract the deep visual features of the images. There are two benefits for extraction by CNN: (1) the local perceptual vision, and (2) weights of shared. The extraction of features is done by a series of convolutional layers. The features are then used to train the SVM classifier in a one-dimensional vector. The accuracy for this method is 71.42%.

2.1 Object Detection

The input is live stream video from the camera. Next, the input video is split into number of frames, i.e., for every minute of video, nearly 2000 to 3000 frames would be generated. Every object in each frame is recognized with the labels that were given to the objects as per the YOLO dataset. Now, position of these recognized objects from the frames have been calculated by considering them to be in a 2D plane. With

this, we get converse side (opposite side to the object) as the output, i.e., if the object is detected in the right side, the output generated would be left side.

2.1.1 Driver Drowsiness Detection

Implemented using computer vision, detect the facial motion and appearances changes during drowsiness. Video imaging systems have been used in the vehicles to monitor the driver movements, percentage of closure of both eyes of the driver.

- Two techniques are used.

1. Facial landmark detection

Facial milestone identification is a cycle of restricting key facial designs on face. Incorporates eyes, nose, mouth.

2. Eye aspect ratio

Eye aspect ratio decides whether the eyes are shut for long measure of time. We can accept the client is in danger of nodding off and sound an alert to command notice.

2.1.2 Lane Detection

A video is given as input in the lane detection from the resource paper lane detection for autonomous vehicles using open CV library which is stored in the CAP variable. We split the taken input video into frames. Then, each frame is sent to the canny image function to detect the edges using canny edge detection algorithm. We crop the part of the image where lanes are present using reason of interest (ROI) into which the canny image is sent as parameter. We create a mask which stores the part of the road in between the lanes. Hough lines give approximately the presence of lanes from the image.

2.2 *Architecture of Research Methodology*

Engineering for scene division for metropolitan situation from far off detecting information is appeared in Fig. 1.

Stage 1: Data assortment and dataset planning.

Stage 2: Dataset is isolated into preparing dataset for making the model and test dataset for testing the dataset.

Stage 3: Developing a model for scene division for metropolitan situation from far off detecting information.

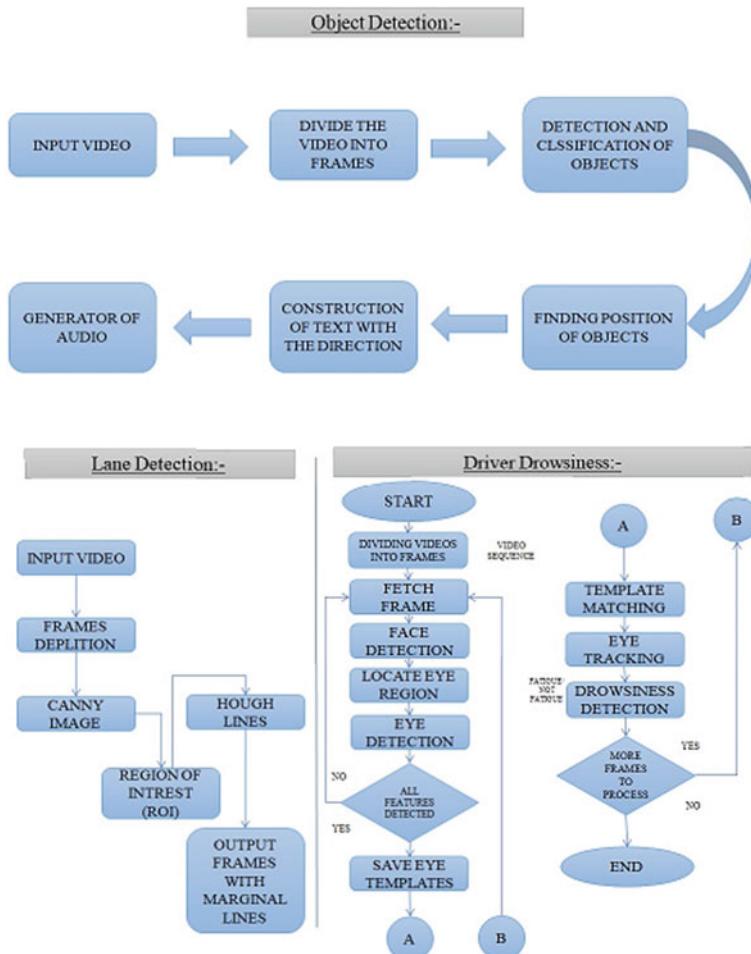


Fig. 1 System architecture

Stage 4: Training and experimentation on datasets.

Stage 5: Testing the model exhibition as far as precision and running it on ongoing information.

The project involves detection of lane lines in an image, object detection and drowsiness detection using Python and OpenCV. OpenCV means “Open-Source Computer Vision,” which is a package that has many useful tools for analyzing images.

3 Procedure of the Proposed Method

Figure 1 describes flowchart diagram of the research methodology. The proposed method describes object detection, drowsiness detection and line detection analysis.

3.1 Procedure for Lane Detection

Input: Input video of road.

Output: Detected lane on the road.

Step 1: Video having road lanes.

Step 2: Splitting the video into frames.

Step 3: Converting images into grayscale image for canny edge detection.

Step 4: Canny edge detection algorithm to detect the lane edges in the frame.

Canny edge detector: input = frame, output = frame converted into gray scale with detected edges.

Step 5: Hough Lines technique: Input = Frame, Output = Frame with detected edges.

Step 6: ROI: Detecting the lanes in the frame.

Step 7: The weight added frames are then made into video as final output.

3.2 Procedure for Drowsiness Detection

Input: Live video of the driver.

Output: Detecting the drowsiness of the driver by making an alert.

Step 1: Capturing the video.

Step 2: Detecting the face in the video.

Step3: Spotting the eyes by using shape_predictor_68_face_landmark.dat file.

Step 4: The eye_aspect_ratio function which is used to compute the ratio of distances between the horizontal eye landmarks.

Step 5: Alerting when eyes are closed.

3.3 Procedure for Object Detection

Output: Detecting the objects around the surroundings.

Step 1: The live input stream of video is given as input.

Step 2: Splitting the image frames into cells typically using 19x19 grid.

Step 3: For more than one object it uses 1805 boundary boxes, for predicting the class of an object.

Step 4: Formula for Bounding Boxes:

$$Y = (P_c, B_x, B_y, B_h, B_w, C)$$

B_x, B_y = Center of a Bounding Box.

B_w = Width of the bounding Box.

B_h = Height of the Bounding Box.

C = Class of an Object.

P_c = Probability that there is a
object in the Bounding Box.

Step 5: Non-max suppression technique:

Input: Input frame consisting of
multiple Bounding boxes.

Step 6: Identifying and predicting the class of different objects and the bounding boxes specifying object location.

Step 7: The detected objects with different class frames and converted into video as final output.

4 Experimental Results

Dataset YOLO dataset can be used for this work. **Evaluation Measures** Comparison with other algorithms on the same dataset for pixel-wise classification accuracy in precision and recall for object and lane detection and drowsiness detection.

Software and Hardware Requirements Anaconda Python sensorflow.

Drowsiness and also uneven/unclear roads are some of the major reasons that lead to road accidents. So we came up with a model that detects when the driver is getting into drowsiness and alerts him. It also allows him to visualize the lane on which vehicle is moving on and also deviations in the lane. Fig. 2 describes output result of lane detection on highway roads. Fig. 3 describes output result of canny image on highway roads. Fig. 4 describes output result of lane detection on country side roads. Fig. 5 describes input diagram of driver drowsiness image. Fig. 6 describes output diagram of driver drowsiness image alert system.

Figure 7 shows object detection model; proposed method was trained on different vehicle classes that run on traffic roads. A result of up to 75% MAP was achieved

Fig. 2 Output for lane detection on highway roads

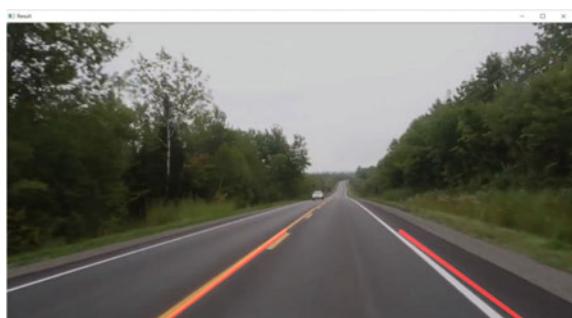


Fig. 3 Output for canny image on highway roads



Fig. 4 Output for lane detection on country side



Fig. 5 Input for driver drowsiness image

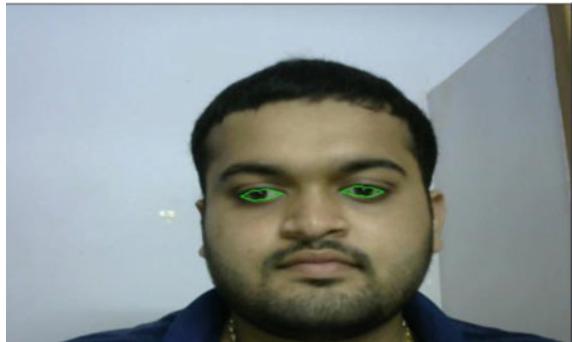


Fig. 6 Output for driver drowsiness alert system

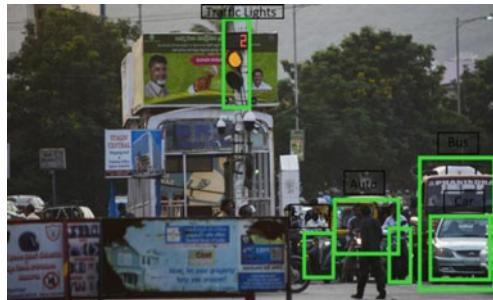


when the model was trained on 5562 video frames using 80–20 ratio of training–testing data. The YOLO CNN model was used, and it was fine-tuned by training the model on YOLO dataset, and some custom-tuned datasets before it was put to test. It was found that high intra-class variation such as bicycle, cars, bus, van truck and traffic lights with less quantity of training samples were reducing the accuracy for some classes. So, different labels were recommended for different looking objects in the same class to reduce the intra-class variation. Training this model to different cities traffic such as Bangalore and Mumbai will enhance the performance. The observations of installation and maintenance costs, traffic measurement effectiveness and rule violation would be shared with the traffic authorities. Above Fig. 7 are the sample images of the model detecting objects.

5 Conclusion and Future Work

This approach worked so well; now we came up with the extension of detecting the vehicles that come by the way the vehicle is moving on for that, YOLO dataset is

Fig. 7 Output for object detection image on traffic



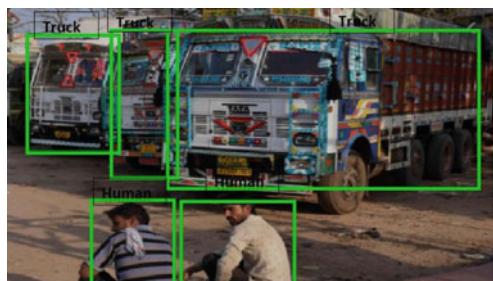
The camera module is attached to the dashboard that detects the traffic lights and vehicles etc.



Output for live Objection detection



Output for Objection detection rough roads



used. YOLO dataset is to find out the different objects that come along on the road and gives suggests the driver in which direction he has to move to pass that object that is coming by his way. This increases the accuracy of autonomous vehicles to drive safely even in a road that has various number of vehicles. So, this helps enhancing the road safety and prevent accidents. In the present work, the extension added was detecting the objects or obstacles and giving the direction to move. No dataset was used in the previous work, YOLOV3 dataset was used to recognize the objects in the present work. In the previous work, modules of canny were used. The efficiency is also increased comparatively, since the obstacle is detected and the safe way is predicted.

References

1. Takizawa H, Kuramochi Y, Aoyagi M (2019) Kinect cane system :recognition aid of available seats for the visually impaired. In: Proceedings of the IEEE 1st global conference life science technology (LifeTech), March 2019, pp 189–193
2. Shaikh F, Meghani MA, Kuvar V, Pappu S (2018) Wearable navigation and assistive system for visually impaired. In: Proceedings of the 2nd international conference trends electronics information (ICOEI), May 2018, pp 747–751
3. Zhang H, Ye C (2017) An indoor wayfinding system based on geometric features aided graph SLAM for the visually impaired'. IEEE Trans Neural Syst Rehabil Eng 25(9):1592–1604

Traffic Management System Using AI and IoT



Priya Geda and Sumitra Motade

Abstract Traffic congestion has become a major issue nowadays. Not only in metro cities even in small cities this traffic congestion is a big problem. Therefore, we have required an intelligent traffic control system that can decrease traffic. Our current traffic control system is not flexible and adaptable, and it is time-based and is unaffected by heavy traffic. Because of its static nature, it is unreliable, unpredictable, and noisy. The proposed system is the approach for traffic control. In the proposed system the image processing and IR sensors are used. IR sensors are installed at the roadside from where they detect vehicles and send results to Arduino UNO. On the other side, image processing is used to count vehicles by using camera surveillance and sends the result to the microcontroller through a serial port. Arduino UNO then analysis both the results and update green path time according to the traffic density on the road. Using technologies separately may have its disadvantages which can be overcome by considering a combination of both results for a better result. This method increases the accuracy and productivity of the system. Apart from this, the project gives prioritization to the emergency vehicles like fire trucks, ambulances. Also, this project saves basic information about vehicles, i.e. speed, vehicle colour, type, etc. This information can be used for further analysis, and as a result, it can forecast traffic congestion and road condition at various times.

Keywords Traffic management system · Image processing · CNN · IoT · IR sensor · Arduino UNO · Prioritization to emergency vehicle

P. Geda (✉) · S. Motade

School of Electronics and Communication Engineering, Dr. Vishwanath Karad MIT-WPU, Pune, Maharashtra, India

S. Motade

e-mail: sumitra.motade@mitwpu.edu.in

1 Introduction

With each passing day, technology in almost every sector emerges and vanishes. When it comes to the rising pressure on our roads and traffic congestion problems, road traffic has become an unavoidable reality in the world's rising megacities. The increase in traffic is due to an increase in the proportion of cars on the roads and insufficient infrastructure to manage it. Increased traffic congestion wastes fuel, time, and thus raises the overall cost of transportation [1]. According to the researcher, it is the basis of the global economy. Almost every country is confronted with this inevitable issue, which is accompanied by other related issues. On the other hand, it not only harms the environment but also harms the productivity of industries.

The paucity of a complex traffic management device that chooses correctly has been identified as the primary cause of road traffic. The current traffic control system has a preset green path clearance time that is not affected by traffic volume [9]. Even if there are not many cars on a specific roadway, the green signal period remains constant, causing vehicles in other lanes to wait unnecessarily. Furthermore, there is no prioritization for ambulances and fire trucks that can be shown to be life-threatening. As a result, resolving this issue has become very crucial.

This area is immense. A lot of researches are going on in this field. The problem can be overcome by using the technologies such as the IoT and image processing. These techniques have been used in our approach to design a suitable and efficient traffic management system. Each approach has its own set of flaws. As a result, the proposed system employs a hybrid model that increases the accuracy and is more effective, thus resolving the difficulties that arise. Other than managing traffic, this system gives prioritization to the emergency vehicles. This model can be used for further analysis related to traffic jams.

2 Related Work

There are so many technologies designed for traffic management. This would include technology such as traffic alert systems, ZigBee, ultrasonic sensors, RFID, video detection, Lorawan, GPS, and WSN.

In paper [1] congestion level is developed by using the Internet of things. The researchers have used ultrasonic sensors to manage the timings of road signals depends on the present jams at the intersections. The drawback of this paper is that it did not propose an effective simplification for EV's which can often be combined into this controlling jams schemes via IoT.

Article [2] proposed the model which will improve the traffic signal's duration period is dependent on the number of vehicles on the road. In paper [2] the drawback of it is that IR-frequencies are affected by hard objects, smoke, dust, fog, sunlight, etc. This model cannot able to communicate the ambulance to the entire base station

to obtain a simple free lane to hurry up to the hospital on time for the person who is in need.

In paper [3], RFID tag and RFID reader are used to count no. of automobiles on the roadway. In this system, installation of an RFID tag is necessary. Vehicles have to be in a range of the RFID reader. The drawback of the system is that the EMV has to come in the range of RFID reader to be detected and once it is detected it will have to wait for some time as the time will not be sufficient to make the changes in the traffic signal. The cost of initial installation is high.

Paper [4] is built to manage traffic on roads by the sensor, video surveillance, and RFID tag and reader installed on the roadside. It adjusts signals time based on heavy traffic on a specific roadside and controls traffic by interacting with a central computer.

Paper [5] describes the study of using the data-mining tool to assist in the production of street lights that schedules signals. This method can be used to develop a time of day (TOD) signal control device by using the data which already exist in the database. The major drawback is that it uses historical data for the prediction of the traffic jams on the road, and hence, accuracy may vary in this model.

The prediction of traffic flow for a short period method was also proposed for the people to choose their path and predict time by using spatiotemporal analysis and GRU [6].

Another method established is wireless communication by using the transceiver RF module installed at every traffic junction and tollgates and programmed to connect each automobile.

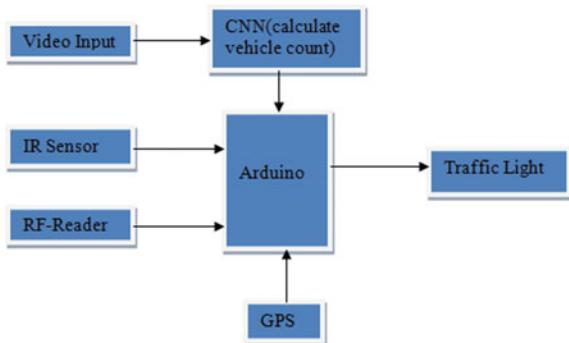
3 Proposed System

So far, numerous techniques have been introduced for the traffic management system. Techniques such as inductive loop counter and weight sensor need an underground construction, which raises implementation and maintenance costs. If techniques have been used separately then there must have been flaws that can be resolved to increase system performance. In this model, we have offered an approach to managing traffic congestion. Apart from this, the project gives prioritization to the emergency vehicles like fire trucks, ambulances. This project is designed to increase accuracy and more efficiency.

The key benefit of this strategy is that it is more efficient due to the hybrid approach. Image processing and IoT if used alone may have some drawbacks like variations in light or poor environment which may vary the result. Similarly, if an IR sensor is used alone may get an inaccurate result when the distance between two vehicles is more. Thus, proposed model can overcome these drawbacks by using both technologies together. In this model, if one system gets failure there still have another active one.

The picture of vehicles can be taken by the camera which is located near to the traffic lights at the crossings which are possible by image processing. After getting the vehicle count from the picture it further sends to Arduino UNO. Simultaneously,

Fig. 1 Block diagram of proposed system



installation of the IR sensors on the roadside detects the no. of vehicles and send them to the Arduino UNO. Apart from this when the emergency vehicle is detected by the RF-reader it gives priority to it and switches to a green signal directly.

Figure 1 depicts the block diagram of the system. Here Arduino UNO is the IoT gateway that performs all the tasks of the system. A pair of IR sensors are used to count the no. of cars on the road. And the images are captured by the camera and then analysed with OpenCV to determine the traffic on the road. Also, the emergency vehicles can be detected by the RF-reader whenever it passes. One of the major disadvantages of the current traffic light control system is that it is unchanged. The proposed system uses the SSD model, for detecting the vehicle, which gives the required accuracy and processing speed, with mobilenet used to produce high-level features. After detecting the vehicles and count frames, the results send to Arduino UNO by serial port. Simultaneously, there are IR sensors on the other side which also counts the vehicles and send results to Arduino UNO. Now, we have both the results from the IR sensor and image processing. The Arduino connected to the traffic lights which update the clearance time based on how many vehicles are counted by the system. The RF-reader is connected to Arduino UNO to detect emergency vehicles. GPS is used to give global positions by calculating latitude and longitude.

IR sensor is used to count the cars on the roadway. In it, the IR transmitters release IR beams and the IR receiver receives the reflected IR rays. If any vehicle passes out from these sensors, can be detected by the IR receiver. Let x is the distance, in metres, where the first IR sensor is placed. Let say other IR sensors are placed at some interval let us say y metre. So the second sensor will be placed at $(x + y)$ and third would be at $(x + 2y)$ and the fourth will be at $(x + 3y)$ and so on. Now these IR sensors are further connected to Arduino pins, where they send the information about the count of vehicles. Table 1 shows the status of the IR sensors in different situations. Their results are as follows:

In this table 1 and 0 are the binary no. where 1 represents high and 0 denotes low and IR1, IR2, IR3, and IR4 are the 4 infrared sensors. The Arduino receives the result from these sensors and processes accordingly and controls the traffic.

Table. 1 Status of the sensor

Sr. no	IR1	IR2	IR3	IR4	Traffic status
1	1	0	0	0	Low
2	1	1	0	0	Medium
3	1	1	1	0	High
4	1	1	1	1	Extreme high

Simultaneously, the image can be taken by the camera which is located at the crossings. These images are further processed by the OpenCV which includes the following steps:

1. **Video Analysis:** The input video is analysed by Python code using OpenCV module functionality where it reads video frames. These frames then send to TensorFlow object detection API. The input video has frame width \times height is (640×352) and the total bitrate is 510 kbps. I have used an SSD model here for vehicle detection. After downloading the model, here we have loaded the label map. Label maps are generally used for mapping the genre names.
2. **Resize Image:** After capturing the image, it sends the image piece to the colour recognition module.
3. **Extract features:** Extracting the features is done by using the colour histogram method. This is the easiest and instinctive way to retrieve all features out of the image's different colour channels. Features can be extracted by plotting the histogram of the different colour channels and then collecting the data from the histogram. This helps in extracting features with peak pixel values for R, G, and B.
4. **Training Classifier:** In TensorFlow, two modules are included, one is the vehicle detection image by colour recognition module where the training dataset is giving to train the model for colour classification by using a KNN classifier with histogram. And other is the speed and direction prediction module. After extracting the feature, the KNN classifier is used in different classes for colour recognition. The classification is done by the mass voting of its neighbour. The data is set to the class with the most nearby neighbour. The accuracy depends on increasing the value of the nearest neighbour, i.e. the value of k . This model also gives information about every vehicle that passed through the ROI line.
5. **Predict vehicle:** For vehicle detection, a customized SSD model was trained that can detect vehicles of various categories like motors, bicycles, vans, taxis, auto-rickshaw, etc. SSD model is a smart CNN for real-time vehicle detection. Single-shot detection (SSD) applies to take one-shot to recognize many vehicles by the image. The SSD model is evolved from a feed-forward convolution network, which provides a set of bounding boxes of a certain size.

After the predicted vehicle, the ROI line is set. Whenever the vehicle is passed over the line it makes the colour of ROI from red to green. As Fig. 4 shows if ROI is greater than the threshold, the vehicle count increases and if not it will read the video frame again. After counting all the vehicles, the frame will be zero and then send to the Arduino by a virtual serial port driver. Apart from counting the vehicle this system also gives information about the vehicle such as speed of the vehicle, the colour of it, vehicle type and direction of it which can be used for further analysis. Figure 2 shows the flowchart of the image processing.

After all these steps the virtual serial port drive is used to send the results of image processing to Arduino. Now we have both the results from the IR sensor and image processing which is used to calculate the final traffic density. Arduino UNO is connected to the traffic lights. The clearance time of the traffic lights will depend on how many vehicles are counted by both systems. Let us say if there is low traffic density detected the clearance time will increase by 1 s. to the initial count value, and if medium traffic is detection there will be adding 5 s. to the initial count. Similarly, if high traffic is detected 10 s will be added to the count and so on. The following Table 2 can be used for the mapping to calculate the time:

Apart from traffic control, the proposed system can also be used for emergency vehicles. To give prioritization to the emergency vehicle, the RF-reader has been used which are also connected to Arduino and located on the road. It is understood that an emergency vehicle has an RF transmitter so that we can detect the vehicle whenever it passes near the RF-reader. The communication will then start between the transmitter and receiver. As soon as the reader detected the emergency vehicle it sends the information to the Arduino and it instantly turns on the green signal. The signal will stay green until the receiver receives it.

4 Simulation

Figure 3 shows the simulation of the model. For the simulation, here one pin of IR sensor is connected to the pins A3, A4, A5, 13 respectively of Arduino UNO microcontroller, and the other two pins are grounded and connected to the VCC supply. The output pins, i.e. A0, A1, and A2 of the microcontroller, are connected to the traffic lights diodes, i.e. D1, D2, D3 or LED-Red, LED-Yellow, and LED-Green. One pin is connected to the serial port, which is linked to the image processing. The PD1/TXD pin is connected to a virtual terminal which is used to see the result virtually. And 12 pin is connected to an RF-reader which is used to give prioritization to an emergency vehicle. When the key is active the traffic light directly turns on the green signal to clear the traffic. GPS is also connected to give a global position. The transmission pin of the GPS is connected to the virtual serial port. The output of it is done by python code. It shows the global position by calculating latitude and longitude. GPS shows the traffic density on Google maps. If the traffic density is low, it will show a green path, medium then it shows a blue path and if there is high traffic then it shows a red path on Google map.

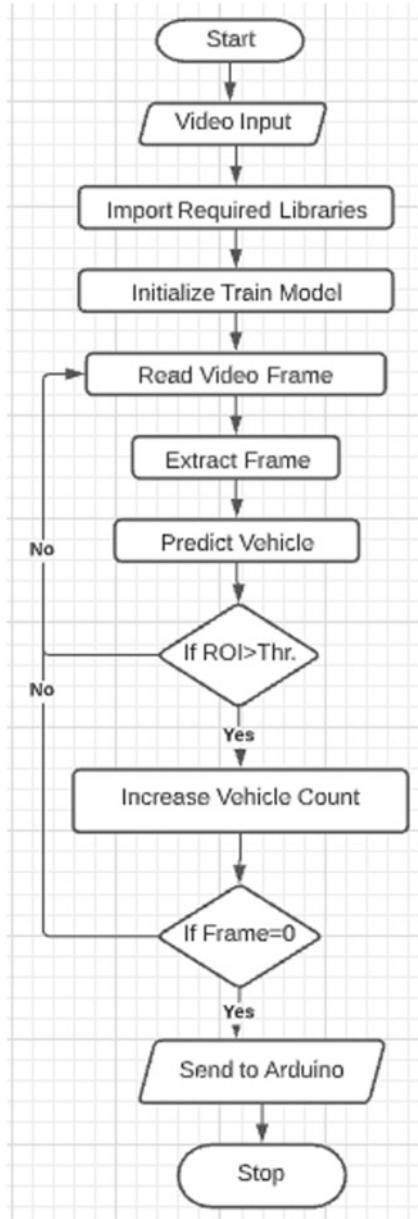


Fig. 2 Flowchart diagram of image processing

Table 2 Plotting traffic status and time interval

Sr. no	Traffic status	Time interval (s)	Weighted avg
1	Low	Count = count + 1	1
2	Medium	Count = count + 5	2
3	High	Count = count + 10	3
4	Extreme high	Count = count + 15	4

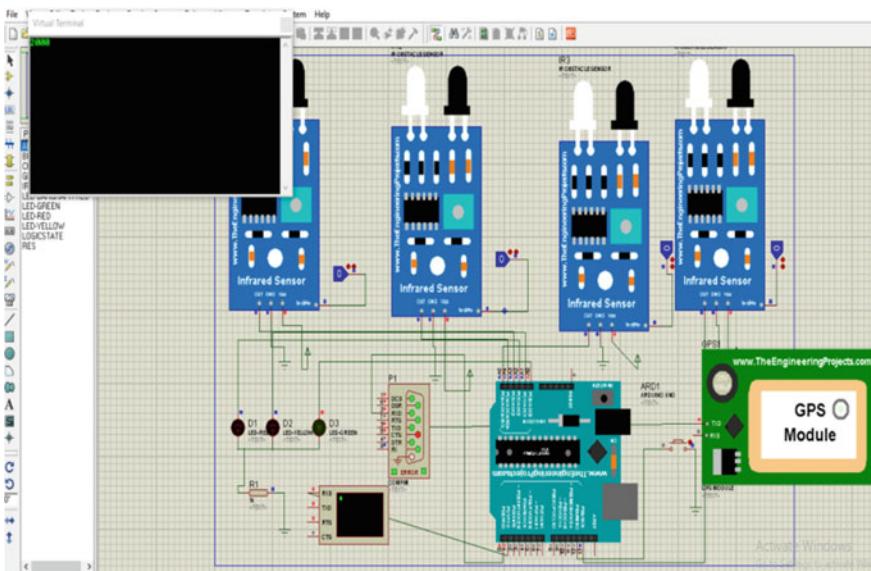
**Fig. 3** Simulation of the project

Figure 6 indicates the simulation of the traffic signal. The simulation part is done by the Pygame library, which is generally used for developing video games. It also provides computer graphics and audio packages that are intended for python programming. Here Pygame library is used to simulate the real-time traffic signal. This is a four-way intersection with two-lane. There are four traffic light signals, one for each direction and each one has a count-down timer right above it. The no. of vehicles that have cross-intersections displayed beside the signal. When the red signal reaches 5 s. a snapshot of a vehicle at the signal is taken and the vehicle detection algorithm is run which returns the no. of vehicles of each class. This is inclined to set the green signal timings. The red signal time of all other signals will be changed accordingly (Fig. 4).



Fig. 4 Simulation of the traffic signal

5 Result

Figure 5 shows the combined result of both the technologies, i.e. Artificial Intelligence and IoT. On the image processing side, the video input is given to the OpenCV module. After that image processing using OpenCV with `ssd_mobilenet_v1_coco` algorithm is being applied. After getting the frame = 0 the total vehicle count data send to Arduino UNO via. serial port. Initially, the count delay is set to 2000 ms.

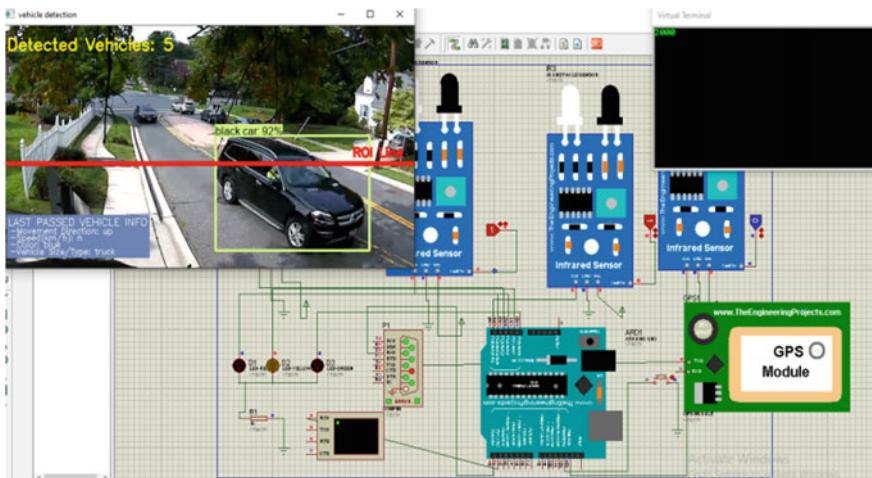


Fig. 5 Combine result of both technology

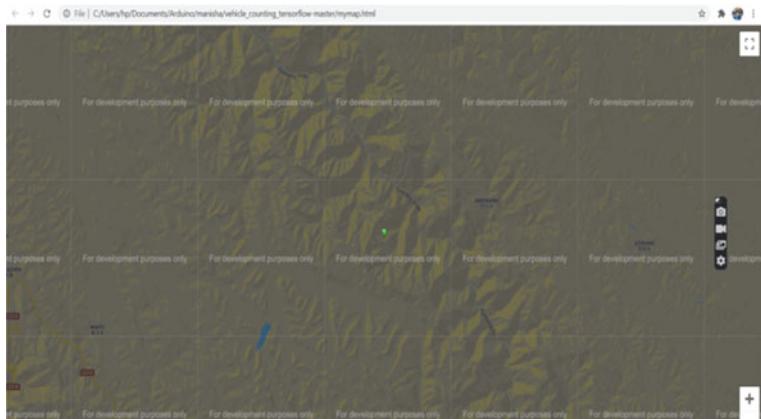


Fig. 6 Output of GPS

while detecting the traffic density the count delay increases accordingly. Like if the traffic density is low the 1000 ms will add to the count delay, if medium it will increase by 2000 ms and so on. Similarly, if one IR sensor is active and others are inactive then the count delay will increase by 1000 ms and so on. After calculating all the parameters accordingly traffic light will turn on. If an emergency vehicle is detected to RF-reader then the controller gives priority to the emergency vehicle and immediately turns the green light on.

The system also gives information about the last passed vehicle such as vehicle type, colour, speed, and direction which can be saved to the database for future analysis. Table 3 has drawn to show the details about the vehicles.

Figure 6 shows the output of GPS. It gives the global position on the Google map.

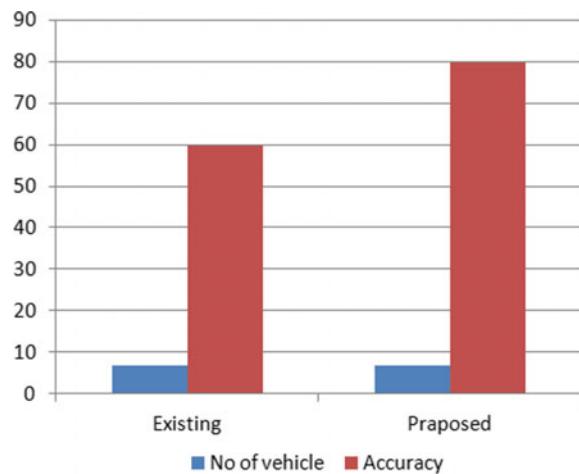
Figure 7 shows the comparison of two models that is YOLO and SSD. The proposed system used an SSD model because of its high speed for detecting the vehicles.

Figure 8 shows the existing vs proposed system in terms of accuracy and no. of vehicles.

Table 3 Details about the last passed vehicle

S. no	Vehicle type	Vehicle colour	Vehicle direction	Speed (km/h)
1	Car	Red	Down	79.58
2	Surfboard	Blue	Down	69.56
3	Truck	White	Down	65.04
4	Truck	White	Down	43.27
5	Truck	Blue	Up	n.a
6	Car	Black	Down	28.37
7	Car	Green	Down	39.80

Parameters	YOLO	SSD
Model Name	You only Look once	Single Shot Detector
Speed	Low	High
Accuracy	80.3%	72.1%
Time	0.84-0.9 sec/frame	0.17-0.23 sec/frame
Frame per second	45	59
Mean Average Precision	0.358	0.251

Fig. 7 Comparison between two algorithms**Fig. 8** Existing versus proposed traffic system

6 Conclusion and Future Work

Conclusion: The traditional traffic congestion signalling systems operate on a scheduled time at every intersection for every ongoing traffic. The major issue is that if there are no vehicles on an ongoing road, traffic on another road must have to wait until the timer of that incoming road reaches 0. This issue creates an undesirable delay on the road which causes heavy traffic jams, human time, money, and increases pollution. Thus, the offered system gives the solution to the increasing issue of road traffic and can efficiently replace the current conventional design or traffic control systems. Because of its consistency and reliability, this technology has become a new frontier of vehicle detection. The clearance time can be varied by varying the traffic volume and circumstances. By controlling the traffic congestion in addition this model gives the priority to the EVs such as hospital wagons, fire brigades makes it more intelligent. Also, this model can be used for further analysis for controlling future traffic jams. In case of one system get failure the system gets active to control the traffic which makes it more efficient and effective.

Future Scope: In future work, live data of Google maps can be streamed to identify the congestion parameter of traffic and the timings and direction of traffic lights can be adjusted accordingly to divert the oncoming traffic through an alternate route.

Also, future work may involve establishing a connection between the proposed systems to generate an accidental warning. The inter-connected mobile unit in an ITS can communicate with each other to send an alert of a unit that has sensed an event. This could interact with the vehicles system to warn the driver of a nearby accident. Cars can share traffic jams and roads conditions to create an updated loads condition database through which the best route to a destination can be determined.

References

1. Akhtar M, Raffeh M, Ul Zaman F, Ramzan A, Aslam S, Usman F (2020) Development of congestion level based dynamic traffic management system using IoT. In: 2020 international conference on electrical, communication, and computer engineering (ICECCE), pp 1–6. doi:<https://doi.org/10.1109/ICECCE49384.2020.9179375>
2. Harshitha D, Harshitha SP, Anjum I, Indushree VP (2019) IoT based smart traffic signal monitoring system. Int J Eng Res Technol (IJERT) NCRACES 7(10)
3. Javaid S, Sufian A, Pervaiz S, Tanveer M (2018) Smart traffic management system using internet of things. In: 2018 20th international conference on advanced communication technology (ICACT), pp 1–1. doi:<https://doi.org/10.23919/ICACT.2018.8323769>
4. Zamani Z, Pourmand M, Saraee MH (2010) Application of data mining in traffic management: case of city of Isfahan. In: 2010 2nd international conference on electronic computer technology, pp 102–106. doi:<https://doi.org/10.1109/ICECTECH.2010.5479977>
5. Dai G, Ma C, Xu X (2019) Short-term traffic flow prediction method for urban road sections based on space-time analysis and GRU. IEEE Access 7:143025–143035. <https://doi.org/10.1109/ACCESS.2019.2941280>
6. Faldu P, Doshi N, Patel R (2019) Real time adaptive traffic control system: a hybrid approach. In: 2019 IEEE 4th international conference on computer and communication systems (ICCCS), pp 697–701. doi:<https://doi.org/10.1109/CCOMS.2019.8821749>

Autonomous UAV with Human Detection



Kshitij Karnawat, Nihar Asare, Sumit Anilkumar Singh, and Anju Kulkarni

Abstract The use of Unmanned Aerial Vehicles (UAV) is increasing tremendously and with it the need for autonomous flight. In this paper, it is discussed how an autonomous UAV is built with the help of a cascaded Proportional Integral Derivative controller and Simultaneous Localization and Mapping (SLAM). The simplistic design and easily available components make a light weight quadcopter style UAV. A Human detection neural network is also added to the UAV which can be used for many applications such as aerial photography, search and rescue, defense. In previous papers the UAVs processed data on board which in turn caused the UAV to use huge amounts of memory and processing power, thereby increasing the power consumption and cost. Whereas the UAV system proposed in this paper transmits raw data to the main control station for processing in turn reducing the on board processing and avoiding the use of costly controllers.

Keywords Autonomous UAV · Drones · PID · Convolutional neural network

1 Introduction

An unmanned aerial vehicle (UAV) is an aircraft without a human pilot on board. The flight of UAVs may operate with various degrees of autonomy: either under remote control by a human operator or autonomously.

While drones originated mostly in military applications, their use is rapidly finding more applications including aerial photography, deliveries, agriculture, surveillance and racing [1].

K. Karnawat (✉) · S. A. Singh · A. Kulkarni

Electronics and Telecommunications Department, Dr. D. Y. Patil Institute of Technology, Pimpri, India

N. Asare

Electrical Department, Dr. D. Y. Patil Institute of Technology, Pimpri, India

The ability of a UAV to take off, execute a mission, and return to its base without significant human intervention promises to reinforce UAV deployment in many application domains.

Building an autonomous UAV is not as easy as building an autonomous car. An autonomous car moves in a two dimensional space and has a lot of constraints on its movement. Whereas an UAV moves in three dimensional space and has six degrees of freedom. This increase in degrees of freedom and the number of variables that need to be calculated cause an increase in the complexity of the -autonomous navigation of the UAV.

The levels of autonomy can be classified into four levels in the lowest being full manual control to a fully autonomous system.

Level 0: requires full control from the pilot as the UAV has no support systems.

Level 1: involves basic assistance features. Typically, Level 1 UAVs are equipped with a hover control. The pilot is still in full control of the UAV but gets some forms of help.

Level 2: corresponds to semi-autonomous UAVs. The UAV can take off and land hover, maintain its roll, pitch and yaw autonomously.

Level 3: At this level, the UAV can dynamically plan a path and fly to a specific location but needs to be monitored all the time by a pilot who can take over in cases of emergency [2].

In this paper, it is discussed how to develop a fully autonomous UAV that can autonomously take off, control its attitude, altitude, hover, land and dynamically plan its path to a given GPS coordinate with the help of the peripherals onboard the UAV.

2 Methodology

2.1 UAV Design

2.1.1 The Microcontroller

The brain of the UAV is a Raspberry Pi controller. Traditionally most drones are equipped with a STM32 based microcontroller instead of the Raspberry Pi. The Raspberry Pi availability in the market and its capability to run the Robot Operating System (ROS) are the two main reasons to choose it over other controller. The higher processing power is also a key factor as this reduces the response time of the system and allows us to add many different modules for further enhancement. In the case of the STM32 based flight controllers available in the market are designed such that the motors and a few other peripherals can be added where as in the case of Raspberry Pi the 40 GPIO pins and the different kinds of input output ports allows the addition of several modules.

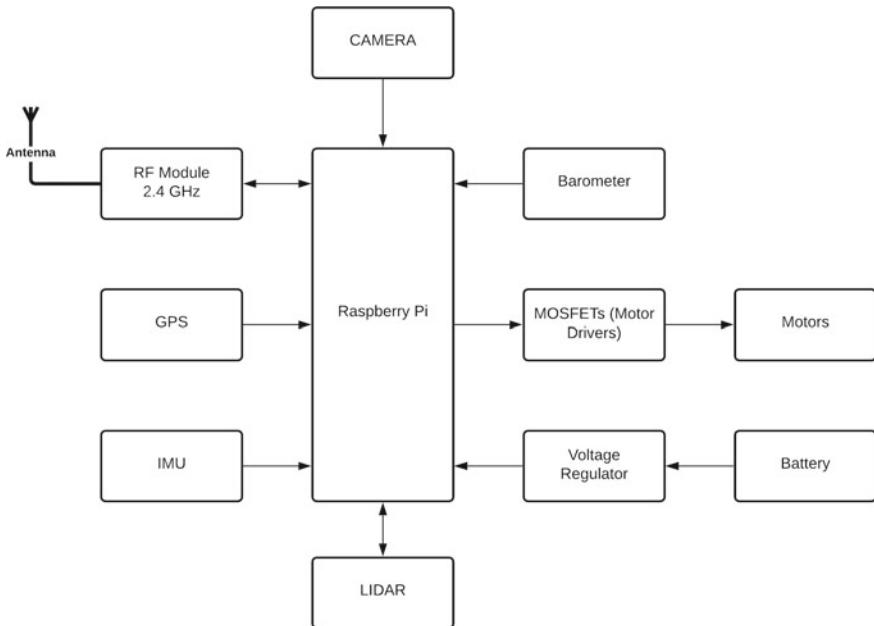


Fig. 1 UAV design block diagram

2.1.2 Motor and Motor Control

The UAV is a Quadcopter style design. This design allows the UAV to move swiftly and with agility in a terrain with a lot of obstacles. Four micro coreless motors are used for the same. Coreless motors are used for their light weight and high revolutions per minute (RPM). These light weight motors reduce the overall weight of the UAV. The high RPM allows the propellers of the quadcopter to achieve high amounts of thrust in turn increasing the load carrying capacity of the UAV. These motors are controlled via a custom built MOSFET motor driver. As the coreless motors do not require a large amount of current to generate high RPM a simple MOSFET is enough to supply power to it. The motor controller controls the speed of the motor just like a traditional motor controller using pulse width modulation (Fig. 1). The custom built motor driver is smaller, lighter and cheaper than the ones available in the market. This helps us to reduce the overall weight of the UAV and also increase its load carrying capacity.

2.1.3 Sensors

During the flight, the UAV may encounter several obstacles in its path. To overcome these obstacles, and avoid collisions it is necessary to detect them as fast as possible. For this purpose, a number of sensors are on board the UAV. To determine the position

of the UAV in three dimensional space a GPS is used to track the location of the UAV in latitude and longitude or the x and y coordinates. It is also the main component of autonomy. The altitude or the z coordinate is given by a highly accurate barometer or an altimeter. Next to determine the orientation of the UAV a nine axis Inertial Measurement Unit (IMU) is used. It comprises of a Gyroscope, Accelerometer and a Magnetometer. These three sensors provide the roll, pitch and yaw or the telemetric data of the UAV thus completing all the six degrees of freedom that are required to represent a UAV.

For avoiding obstacles, a LIDAR is being used. The point data produced by the LIDAR is used for mapping the environment around the UAV. If any obstacles are detected in this environment the UAV then dynamically alters its path to reach the set point. A LIDAR is used instead of an array of infrared sensors as the LIDAR gives a 360-degree view of the surroundings whereas the infrared sensors would have some blind spots between them.

A high resolution camera is also placed on the UAV. The main use of this camera is to relay live feed to the control station. This feed is then used to detect human beings.

2.2 UAV Working

The UAV achieves its autonomy with the help of a Cascaded Digital PID controller as shown in Fig. 2. The PID Controller helps in determining if the UAV is moving in the correct direction or not. The PID controller is a Digital controller as this reduces the number of components that are to be used. A digital PID controller also is easier to tune as no physical components are required giving higher accuracy than a physical PID Controller.

The attitude controller is responsible for handling the roll, pitch and yaw of the UAV. The IMU is used as feedback in this PID controller. The IMU provides real time values for the roll, pitch and yaw of the UAV. These values are then used to generate the error by comparing them to the input values. The attitude controller then

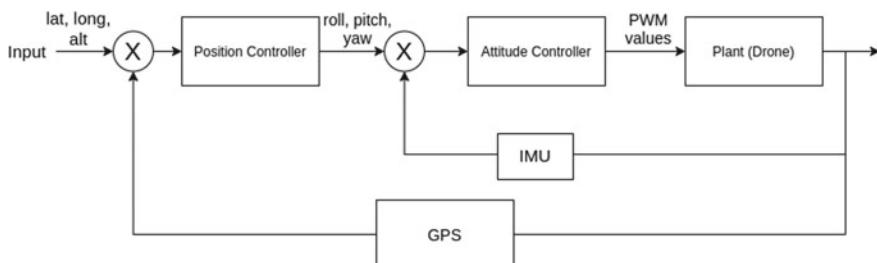


Fig. 2 Block diagram of the PID controller

generates output in the form of PWM values. These PWM values are used to control the speed of the motors on the UAV. Thus changing the orientation of the UAV.

The outer loop of the cascaded PID controller is the position controller. The position controller uses the GPS and altimeter for feedback. The GPS provides the real time latitude and longitude of the UAV which act as the x and y coordinates in the system. The altimeter provides the altitude or the z coordinate. These values are then used to generate the error by comparing them with the location provided by the user from the control station. For a quadcopter style UAV the movement is produced by changing the roll, pitch and yaw of the UAV. Therefore, the position controller uses this error to generate the corresponding roll, pitch and yaw values to minimize the error. These are the same roll, pitch and yaw values that are fed to the attitude controller.

The UAV's LIDAR is used to detect obstacles in real time. Simultaneous Localization and Mapping (SLAM) is used to map the terrain and the obstacles detected by the UAV. SLAM also helps the UAV to plan its path and dynamically alter it if any obstruction has been met with by the UAV.

2.3 Convolutional Neural Network

The high resolution camera onboard the UAV is used for the detection of human beings. Since the UAV does not have the processing power to run a Convolutional Neural Network (CNN) and detect human beings only raw video is transmitted by the UAV to the control station. The control station then passes this video feed via the CNN to detect human beings.

The CNN converts is trained with over 6000 images of human beings and uses techniques like Image Augmentation to improve its results. The CNN is made using TensorFlow 2.0 an open source end to end machine learning platform.

The architecture comprises of 17 layers including four convolutional layers, four max pooling layers, five rectifier linear unit layers, a flattening layer, two fully connected layers and a sigmoid layer. The input layer contains an image with 150*150 pixels. The first convolutional feature map is 148*148. The second is 72*72. Then, the third one is 34*34. The layers of max pooling have 2*2. The first max pooling produces 74*74 feature maps. The second one gives 36*36 feature maps. Then, the third one has 17*17 feature maps. The five Rectified Linear Unit Layers (ReLU) are used between layers to clear the negative values. The last layers are fully connected with 512 nodes. 1 node is connected to the sigmoid layer to give two classes [3, 4].

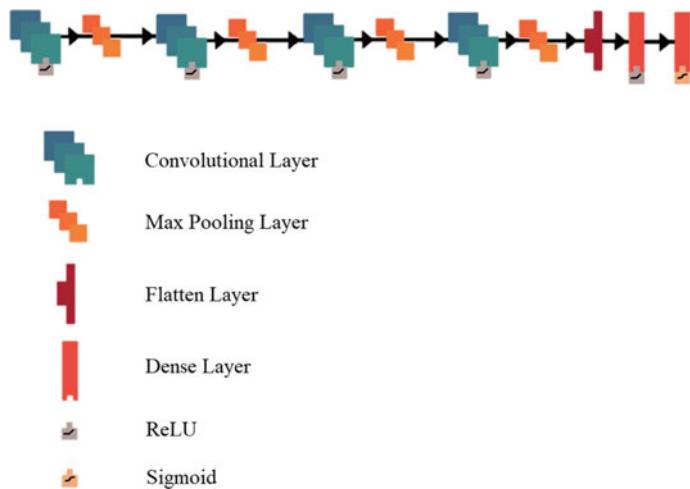


Fig. 3 CNN architecture

3 Results

The CNN trained for detecting human beings was run for a hundred epochs. The accuracy and losses for each epoch were noted and the data was plotted as shown in Figs. 4 and 5, respectively. On the *X* axis of Fig. 4 are the number of epochs and on the *Y* axis the accuracy ranging from 0 to 1 is plotted. As seen in the plot the accuracy at the start was 65% this increased after each epoch. The accuracy after a 100 was in the range of 90–92%. Along with this the validation accuracy is also plotted in red. The validation accuracy is not as stable as the training accuracy but is increasing overall giving an accuracy of around 85%. Plotting the accuracy graph helps us understand if the CNN is working properly or not.

Fig. 4 Training accuracy achieved by the CNN

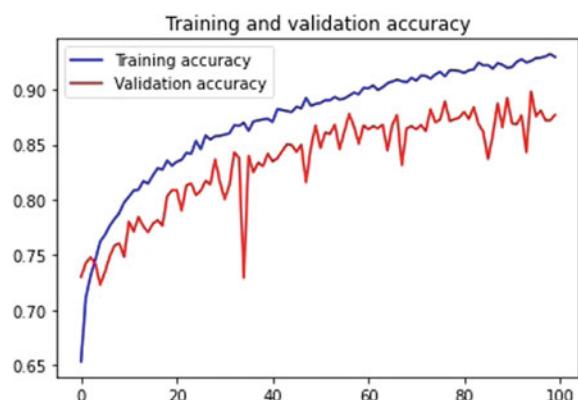
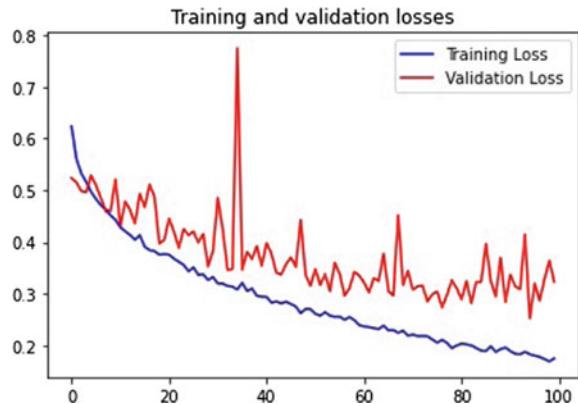


Fig. 5 Training losses for the CNN



In Fig. 5 the training losses have been plotted similar to training accuracy the training losses help in determining the efficiency of the CNN. The lower the loss is the better. As seen in Fig. 5 there is a steady decrease in the loss. The losses nearing the end are lower than 0.2. The validation losses are fluctuating and can be improved in the future (Fig. 5).

The CNN gave an accuracy of 72% on the test dataset with an F1 score of 0.83.

References

1. Yahyanejad S, Rinner B A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale UAVs
2. Kirkland G (2019) What are the levels of autonomy for self-driving vehicles? Robot Bus Rev (26 July 2019)
3. AlDahoul N, Sabri AQM, Mansoor AM Real time human detection for aerial captured video sequences via deep models
4. Krishna SL, Chaitanya GSR, Reddy ASH, Naidu AM, Poorna SS, Anuraj K Autonomous human detection system mounted on a drone
5. Tathe SV, Narote SP Real-time human detection and tracking

Team Sports Result Prediction Using Machine Learning and IoT



Lydia D. Isaac and I. Janani

Abstract The introduction of different new technologies in the field of sports has brought in a lot of modernization into many games. Such technologies like goal line tracking systems, player performance tracking systems and video-assisted smart referee systems involve mounting huge number of cameras around the sports ground and other electromechanical devices. Tracking player performance and other stats related to the game was a difficult task for the team managers and the coaches few years ago. But the advent of technologies like artificial intelligence (AI) and Internet of things (IoT) has made a lot of changes to the way games are being played and been enjoyed by the fans. IoT makes the data collection task convenient as it allows collection of large real-time data by employing numerous sensors, and AI makes predictions easier with efficient machine learning algorithms. The proposed work, hence, focus toward exploring how these technologies help in predicting the results of team sports by quantifying the team's performance. This research takes advantage of the wearables, sensors, tracking systems and integrating technology that help gather relevant data related to the teams for analysis. This data can then be applied to various machine learning algorithms that work better for predictions on game outcomes.

Keywords IoT · Machine learning · Wearable technology · Team performance tracking · Sports result prediction

1 Introduction

Innovations in the recent Internet of things (IoT) technology have led to the modernization of industries driving to the conception of Industry 4.0. Other pioneering paradigms such as data analytics and cloud computing has led to a quite significant

L. D. Isaac (✉) · I. Janani
Sona College of Technology, Salem, Tamilnadu, India
e-mail: lydia@sonatech.ac.in

I. Janani
e-mail: jananii@sonatech.ac.in

development in several application domains. Since sports industry chip in globalized recognition and a sense of respect internationally, it prevails as the most important sector for every nation. Forbes survey done in 2020 revealed the fact that investments made in sports was approximated to 74 billion dollars globally. Added to its numerous smart equipments, even tiny gadgets and devices being developed with respect sports are now enabled with some technology. With this prominence of the sports industry, it has become mandatory to build up novel frameworks for the fruitful apprehension of the vision of smart sports.

Result prediction in sports is a fascinating and the most challenging task as the inherent nature of sport is unpredictable, and also the results of a match depend on incessant potential factors. Although it seems difficult, predicting the outcome of sport is more seemingly sought after by different stakeholders. Social media and news groups rely mostly on the sport experts for publishing latest news on predictions of forthcoming matches. Also the bookmakers, gamblers and other sports betting platforms show huge interest in result predictions as quite a large monetary amount is poured in the betting process. This theme may also be of importance to the players, the team managers and more relevantly to the performance analyst in finding out the significant factors using which some tactics could be applied to win matches.

Basketball being one among the most well-liked and fashionable sports played across the globe is being extensively investigated by the research community in the recent years. The National Basketball Association (NBA) located in the USA is the premier professional basketball league. Quite a number of issues related to this sport have been addressed like player assessment and ranking, importunate behavior of the teams and outcomes prediction. Many statistical models have been proposed by the researches on predicting the outcomes of a match. Wearable devices, motion-capturing systems and other advanced biometric data have started shifting the landscape of this sport. As the game has become more modernized with new technological innovations, this paper focuses towards deriving an architecture where the power of wearable technology could be imbibed in sport result prediction by integrating IoT and machine learning.

2 Literature Survey

Driven by the availability of enormous comprehensive sports datasets, researchers across the globe have come up with several frameworks that could be used for data acquisition and relative decision making with respect to sports. Partha Pratim Ray [1] has proposed an IoT-based standard architecture for sports where a detailed study on different technologies involved has been done. A multi-layered approach to facilitate IoT enablement in sports, and also, other frivolous activities has been suggested. The focus toward many technological issues, interactions, devices involved and data to be considered for analysis are envisaged. Bunker et al. [2] have proposed a novel framework for sport prediction on critical analysis of machine learning literature focusing on applying ANN for result predictions. Luca et al. [3] have come up with

an IoT-aware architecture for improving safety in sport. Ikram et al. [4] reveal IoT to be a novel paradigm which integrates different technologies to support our lives. The paper proposes an IoT-based architecture to protect footballers from serious health problems. The authors focus towards embedding sensing devices like sensors and RFID tags and perform cloud computing in football to monitor health status of the players and in turn reduce happenings of bad health conditions.

A remarkable number of researches have been devoted towards sports prediction on various sports like football, baseball, basketball, cricket and many more. Thabtah et al. [5] have constructed prediction models by applying naïve Bayes, decision trees and ANN methods of machine learning to different datasets. The prediction accuracy rates have been compared promoting selection of the best model for predicting outcomes of the sport.

Razali et al. [6] has employed Bayesian networks for predicting win on the basis of home/away matches and the draws. Valero [7] has worked with the datasets of major league baseball over 10 years and have tried predicting the baseball match outcome based on home team win/lose by using data mining techniques. The work of Pathak et al. [8] was to predict One Day International cricket match outcomes using naïve Bayes classifier, SVM classifier and random forest-based classifier. Loeffelholz et al. [9] have applied neural networks in order to forecast the outcomes of basketball matches. Importance of feature selection has been highlighted, and role of expert opinion has also been discussed. A hybrid fuzzy-based SVM model has been proposed by Jain et al. [10] for analyzing the outcomes of basketball matches. Gobinath et al. [11] have provided a detailed review on the significant role of wearable technology in the consumer sports sector. They have brought out how sensors are beneficial in collecting data that could be useful in plenty of ways with regard to attributes related to the domain of sports. They have provided a detailed study on which sensor technology could be used for sport applications. With such technological improvements, the proposed research focuses on integrating wearable technology through IoT frameworks in the field of sports.

3 IoT-Based Approach

Internet of things (IoT) for sports is frameworks envisioned to provide solutions to the problems related to sports through real-world applications of IoT. Figure 1 shows a generic sensor-fueled analytical architecture that can be used in used for NBA match predictions.

This IoT-based approach to sports data analytics is a multi-layered approach which has a multiple sensors layer, gateway communication layer, sports management layer and sports application layer. This section details on the concepts involved in all the layers, and hence, it is purely theoretical.

1. **Multiple Sensor Layer:** This layer is vital as it involves the inclusion of respective sensors to the body of the players mostly in the form of wearables, sensors

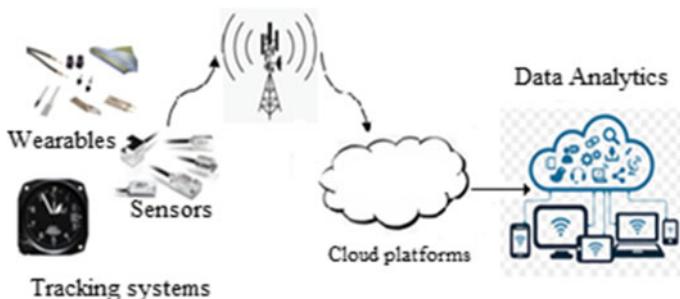


Fig. 1 Generic view of IoT-based architecture for sports

and tracking systems and to the sporting equipments. Wearables are almost attached to the belts, wrist bands and helmets of the players where the possibility of getting busted is the minimum. Many other sensors like body sensors, sports equipment sensors and tracking systems can be placed in and around the sports equipments or the stadium or ground to fetch or track the most essential data required for any predictions to be done. The analog or digital signals received from these sensors are further passed to the communication layer via a network layer which is sensor-based. 610wPAN is an IPv6 personal area network accommodate even the smallest devices.

2. **Communication Layer:** This layer incorporates microcontrollers or digital signal processing units that process the received data from the sensor layer based on embedded/RTOS. The network technologies deployed in this layer should involve long range. Some of such technologies include ANT an proprietary open-access multicast wireless network, industrial scientific and medical radio band that works usually at 2.4 GHz, low-power wide area network, WiMAX, GPRS, LTE supporting 4G technology and 3GPP, Ethernet, HAN, SAN (storage area network), GSMUMTS, Wi-Fi and RF. Satellite and fiber optical communication offered above it leverage high speed connectivity.
3. **Sports Management Layer:** This layer involves statistical analysis, data mining and visualization, predictive and equipment analysis. This layer mainly includes cloud services. The cloud services need to investigate or examine all the data that is received from the wearables and other sensor devices so as to provide accurate feedback reports to the team coach or the team managers. It must also enable the QoS and the security capabilities. Safe storage of data and compatibility with every software platforms is most importantly necessary with respect to the cloud.
4. **Sports Application Layer:** Many key insights are gathered on performing descriptive and inferential statistical analysis on the data that has been received from the lower layer. 3D visualizations of the obtained values help in understanding the real-time characterization of the sport. Rapid changes have been brought into the field of sports due to developments in science and technology.

Sports organizations employ data mining techniques to draw up statistical analysis, player performance and game outcome prediction, pattern discovery, need for training analysis and almost in every strategy planning. This section will briefly introduce some important machine learning algorithms that could be used in predicting the results of the sport. This involves SVM, logistic regression, naive Bayes and artificial neural networks.

Figure 2 shows the flow diagram of the machine learning side of the proposed approach. The main objective of this section is to exploit patterns in training data and use them to analyze unknown data. This case is a multiclass classification problem as the classification is being done into more than two classes' viz. win, lose and draw.

The most essential and important step before the process of training any model is data cleaning and extracting the necessary features. This is necessary for obtaining more accurate results. And also to prevent ending up with poor prediction outputs, it is indeed necessary to work upon the missing data and the outliers. Thus, this step contributes in removing redundant and irrelevant attributes from our data that can deteriorate the model's accuracy (Fig. 2).

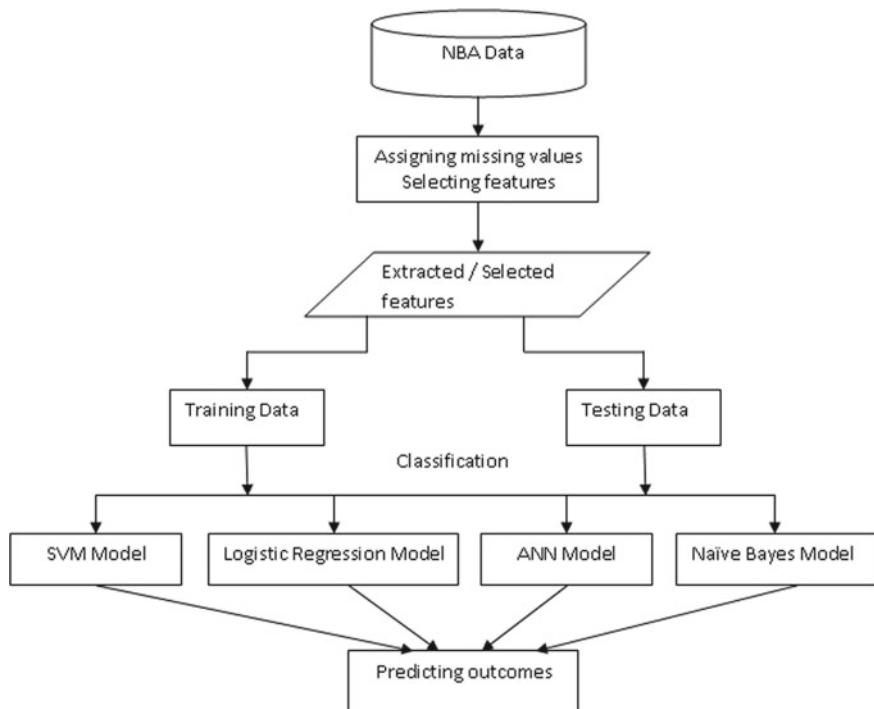
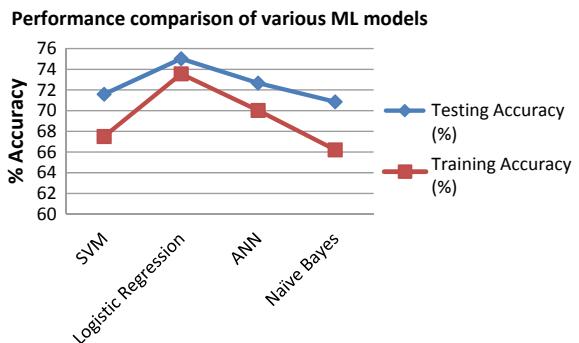


Fig. 2 Machine learning architecture for basketball outcome prediction

Fig. 3 Comparison between various ML models used for NBA outcome prediction



3.1 Support Vector Machine

Providing with features having nonlinear relationships make a support vector machine do well. SVM classification models work by discovering a maximal margin hyperplane.

To put in its most simple way, SVMs are used for binary classification tasks. For multiclass classification, the problem is broken down into multiple binary classification tasks generally termed as one-vs-one, and therefore, the required number of classifiers for multiclass classification is obtained by the formula:

$$\frac{n * (n - 1)}{2} \quad (1)$$

SVM tries fitting a model by maximizing the geometric margin separating positive and negative samples. Since optimization of the geometric margin and assurance that all the samples are separated by the margin has to be done simultaneously, this introduces a constrained optimization problem. The Karush–Kuhn–Tucker in short called as KKT condition by Boser et al. 1992 provides a solution to this constraint. And also, kernel trick can be practical in transforming nonlinearly separable features into the HD feature space. Though several kernel choices seem to exist, the most common ones used frequently are polynomial kernels and Gaussian kernels.

3.2 Logistic Regression

This is a generalized linear regression model. The target variables are discrete class labels. It has a sigmoid curve and the equation of sigmoid function is

$$S(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

This model always takes care to that the generated number falls in the range 0–1. This is because the numerator is smaller than the denominator by 1. Multiclass classification here can be employed using one-vs-rest scheme where, a binary classification problem is done for each class or by altering loss function as cross-entropy loss. Although logistic regression models are most popularly used in sports domain, they fail in identifying nonlinear relationships among the features.

3.3 Artificial Neural Network

The artificial neural networks are nothing but mathematical models that resemble the natal neural network. It comprises of three layers of neurons: input layer, hidden layer and output layer. It has the advantage of modeling even complex relationships between input and the output. Constructing neural networks to associate the input vector to the output of the network that has more than two classes is a multiclass neural network classification. Two architectures have been proposed by researchers for building neural networks for multiclass classification. The first one involves building multiple binary ANNs such that each one is trained with different features, but this may produce overlaps in the classification boundary zone. The second one is implementing a single ANN but with multiple outputs. Though the complexity is high, it defines sharp classification boundaries.

3.4 Naïve Bayes

The naïve Bayes classifier is the fastest when it is for classification tasks. It assumes the attributes to be conditionally independent of the class label. It is based on Bayes theorem that works on conditional probability as shown below

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)} \quad (3)$$

where $P(H)$: the probability of hypothesis H being true, $P(E)$: the probability of the evidence, $P(E|H)$: the probability of the evidence given that hypothesis is true and $P(H|E)$: the probability of the hypothesis given that the evidence is true.

The class with maximum is selected to be the suitable class. The most common application domains where naïve Bayes classifiers are employed include classifying texts, spam detection and applications that involve real-time predictions.

3.5 Boosting

A very powerful learning can be established with the concept of boosting. It forms a strong classifier committee by combining many weak classifiers in order to provide solutions to classification tasks. This approach would outperform most strong monolithic classifiers like SVMs and ANNs. The most commonly available boosting algorithm includes discrete AdaBoost, LogitBoost, Real AdaBoost and Gentle AdaBoost.

3.5.1 Algorithm

The proposed work aims in predicting the results of a match, i.e., win, lose or draw by employing classification models in machine learning.

- Importing the required Packages and Datasets.

Last 10 years NBA data is taken here to train the model.

- Performing Feature extraction and Data Cleaning.
- Performing Classification to predict match results (Win/Draw/Lose).

Dataset used:

Season	Day	WTea mID	WScore	LTea mID	LScore	WLoc e	WMO T	WFG M	WFGA 3	LFM	LFM	LFTA	LDR	LAst	LTO	LSI	LBk	LPF		
0	2003	10	1104	68	1028	62	N	0	27	58	10	16	22	10	22	8	18	9	2	20
1	2003	10	1212	70	1033	63	N	0	26	62	24	9	20	20	25	1	12	8	6	16
2	2003	11	1256	73	1431	61	N	0	24	58	26	14	23	31	22	3	12	2	5	23
3	2003	11	1295	56	1457	50	N	0	18	38	22	8	15	17	20	9	19	4	3	23
4	2003	11	1400	77	1208	71	N	0	30	61	16	17	27	21	15	12	10	7	1	14
5	2003	11	1458	81	1106	55	H	0	26	57	11	12	17	6	22	8	19	4	3	25
6	2003	12	1161	80	1236	62	H	0	23	55	15	20	28	9	21	11	30	10	4	28
7	2003	12	1106	75	1457	61	N	0	28	62	17	17	23	8	25	10	15	14	0	18
8	2003	12	1164	71	1156	66	N	0	28	58	18	12	27	13	26	13	25	8	2	18

Classifier makes the decision based on 13 features which are traditional basketball metrics:—wfgm: field goals made, wfgm3: three pointers made, wfga: field goals attempted, wfga3: three pointers attempted, wftm: free throws made, wfta: free throws attempted, wor: offensive rebounds, wdr: defensive rebounds, wast: assists, wto: turnovers, wstl: steals, wblk: blocks, wpf: personal fouls and 3 engineered features:—fgp: field goal percentage, 3 pp: three point percentage, ftp: free throw percentage.

- Added to the common features, the player statistics is being taken to analyze success of the particular shoot. The following features are taken to analyze

Table 1 Accuracy % obtained

Classifier	Testing accuracy (%)	Training accuracy (%)
SVM	71.59	67.5
Logistic regression	75.02	73.56
ANN	72.67	70.03
Naïve Bayes	70.86	66.21

whether the shoot is successful or not: Combined shot type, Location X, Location Y, Minutes_remaining, Shot_type, Shot_zone_Area, Shot_zone_basics, Shot_zone_range and opponent.

5. From the dataset, 70% is taken as training data and 30% as testing data.
6. Training the model using various classification algorithms.
7. Comparing the accuracy of the models, the best model is being chosen.
8. Predicting the outcome of the match using the best model.

By using data samples extracted from the data mart, the specified models were trained and tested for the purpose of predicting the NBA game outcome. The accuracy of the models has been shown below in Table 1.

Simple logistics classifier yielded good prediction results with 73.56% accuracy compared to that of naïve Bayes with lowest accuracy of 66.21%. The comparison between the performances of various ML models used for analysis has been represented in Fig. 3.

4 Conclusion and Future Scope

As a research of investigating integration of IoT and artificial intelligence (AI) for predicting outcomes of team sports, this work carries certain reference value in the future for related research work and implementations. IoT has been used to add up to the accuracy of the collected data as they are real-time data. Wearable technology is used for data acquisition and machine learning is used for result prediction. Several issues could be faced during implementation as addressed below:

- Establishing IoT-based infrastructure in the stadiums could be expensive.
- Every player may not agree upon on using wearables for collecting data as they may feel as privacy breach.
- Interoperability is vital for the success of the entire system as this is an interconnection of many devices working with different technologies.
- Technology advancements may make the system obsolete after a particular period; hence, maintenance cost could be high.
- Technological issues may affect data transmission and hence may introduce delays.

This work assumes the power of network intelligence, coordination between the devices involved and analytics to have a dedicated connection between the players and equipments used by them. In the near future, physical implementation of the entire system would be done and the related issues would be addressed for further research.

References

1. Ray PP (2015) Generic internet of things architecture for smart sports. In: 2015 international conference on control, instrumentation, communication and computational technologies (ICCICCT), 978-1-4673-9825-1115
2. Bunker RP, Thabtah F (2019) A machine learning framework for sport result prediction. *Appl Comput Inform* 15(1)
3. Catarinucci L, De Donno D, Mainetti L, Patrono L, Stefanizzi M, Tarricone L (2017) An IoT-aware architecture to improve safety in sports environments. *J Commun Softw Syst* 13(44). <https://doi.org/10.24138/jcomss.v13i2.372>
4. Ikram MA, Alshehri MD, Hussain FK (2015) Architecture of an IoT-based system for football supervision (IoT Football). IEEE 2nd World forum on internet of things (WF-IoT), Milan, pp. 69–74
5. Thabtah F, Zhang L, Abdelhamid N (2019) NBA game result prediction using feature analysis and machine learning. *Ann Data Sci* 6:103–116
6. Razali N, Mustapha A, Yatim F, Aziz R (2017) Predicting football matches results using Bayesian networks for English premier league (EPL). *IOP Conf Ser Mater Sci Eng* 226:012099. <https://doi.org/10.1088/1757-899X/226/1/012099>
7. Valero CS (2016) Predicting win-loss outcomes in MLB regular season games—a comparative study using data mining methods. *Int J Comput Sci Sport*, pp 91–112
8. Pathak N, Wadhwa H (2016) Applications of modern classification techniques to predict the outcome of ODI cricket. *Procedia Comput Sci* 87:55–60
9. Loeffelholz B, Bednar E, Bauer KW (2009) Predicting NBA games using neural networks. *J Quant Anal Sports* 5:1–17
10. Jain S, Kaur H (2017) Machine learning approaches to predict basketball game outcome. In: Proceedings of the 3rd international conference on advances in computing, communication & automation (ICACCA) (Fall). Dehradun, India, pp 1–7
11. Aroganam G, Manivannan N, Harrison D (2019) Review on wearable technology sensors used in consumer sport applications. *Sensors* 19:1983. doi:<https://doi.org/10.3390/s19091983>

A Hybrid SVM–ABC Model for Monthly Stream Flow Forecasting



Ujjawal K. Singh, Baidyanath Kumar, Nabin K. Gantayet, Abinash Sahoo, Sandeep Samantaray, and Nihar Ranjan Mohanta

Abstract Developing reliable estimates of stream flow forecasting is crucial for water resources management and flood forecasting purposes. The study investigates accuracy of hybrid support vector machine (SVM) integrated with artificial bee colony optimisation (ABC) algorithm for monthly stream flow forecasting. ABC algorithm is applied for improving performance of SVM model by helping in selection of optimal SVM parameters. Monthly flow data from two stations (Salebhata and Sundargarh) on Mahanadi River basin, India, are used in the study. Results of hybrid model are compared with results of conventional SVM model, and quantitative statistical measures are used for validating both models. Obtained results prove that hybrid SVM–ABC model processes complex hydrological data better and have better generalisation capability with higher prediction accurateness.

Keywords ABC · Mahanadi river · Stream flow · SVM

1 Introduction

Planning and management. However, because of complexity of atmospheric processes, stream flow is one of the most difficult and complex components of hydrological cycle for understanding and modelling. Data-driven models have been broadly utilised for modelling many hydrological parameters, like prediction of extreme events [1], stream flow [2], suspended sediment forecasting [3], reservoir inflow forecasting [4], precipitation prediction [5], groundwater forecasting [6, 7] and rainfall-runoff modelling [8–10]. Most popular and widely utilised data-driven

U. K. Singh · A. Sahoo · S. Samantaray (✉)

Department of Civil Engineering, NIT Silchar, Silchar, Assam, India

e-mail: sandeep1139_rs@civil.nits.ac.in

B. Kumar · N. K. Gantayet

Department of Civil Engineering, GIET University, Gunpur, Odisha, India

N. R. Mohanta

Department of Civil Engineering, NIT Raipur, Raipur, Chhattisgarh, India

models in the field of hydrology are artificial neural network (ANN) and SVM. SVM is an eminent classification tool recommended by Vapnik (2000). It has proven to have good performances in several applications [11]. A number of related studies have grown over recent past, with contributions in field of science and engineering [12].

Guo et al. [13] proposed an improved SVM model to predict monthly stream flow in Yichang station of Changjiang River basin, China, and assessed the model performance with conventional ANN and SVM techniques. They found that improved SVM model gave better results with high accuracy. Kisi et al. [14] studied accuracy of ANN, adaptive neuro-fuzzy inference system (ANFIS) and SVM models for forecasting daily intermittent stream flows of Babaeski and Uzunkopru stations, in Thrace province situated in Turkey. Comparison outcomes showed that SVM performed better than other proposed models in daily intermittent stream flow forecasting. Shabri and Suhartono [15] investigated potential of least-squares SVM (LS-SVM) model for improving accurateness of stream flow forecasting of Tg Rambutan and Tg Tulang of River Kinta in Perak, Malaysia. Assessment of results indicated that LS-SVM model is a convenient and promising technique for stream flow forecasting. Mandal et al. [16] proposed hybrid SVM-ABC technique for overcoming problem of false leak detection. Kuo et al. [17] applied a C5 decision tree (DT) for extracting rules from SVM results and ABC algorithm for selecting important features. Result showed that classification accurateness and complications of final decision tree can be simultaneously improved. Farfa' and Cea [18] implemented and evaluated a calibration method for hydrological models, combining advantages of ABC with ANN. Yildiz and Açıkgöz [19] proposed a hybrid model, integrating extreme learning machine (ELM) technique and ABC algorithm for forecasting small hydropower plant generations.

2 Study Area

Mahanadi River emerges near Pharsiya village in Raipur district, Chattisgarh, at an elevation of about 442 m above MSL, having a catchment area of 141,589 km². Major portion of Mahanadi flows through Chhattisgarh (75,136 km²) and Orissa (65,580 km²) states with a total length of around 858 kms before discharging into Bay of Bengal. Key streams are Tel, Jira, Ib and Ong. Basin is characterised by a tropical monsoon climate (June to September), receiving average precipitation of 1352 mm (Fig. 1). Monsoon season receives in excess of 80% of annual runoff with average yearly flow of 1896 m³s⁻¹.

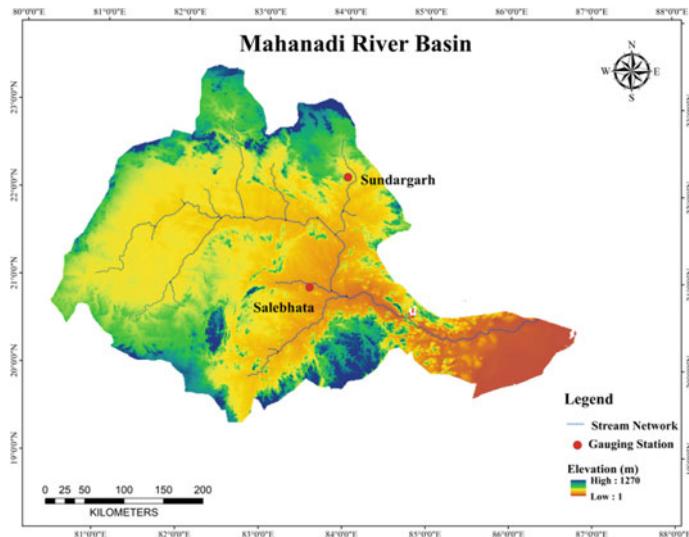


Fig. 1 Proposed study area

3 Methodology

3.1 SVM

SVM is a new supervised learning technique governed by principle of structural risk minimisation which is known for their application in regression and classification processes (Vapnik 2000). It is fast gaining popularity because of its several attractive characteristics and encouraging empirical performances. Previous investigations have revealed that SVM can illustrate good and comparable generalisation behaviour [3, 20–22]. Here, input data are mapped to high-dimension feature space, utilising a non-linear mapping function \emptyset , also known as kernel function. Considering a given training data $a\{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq X \times R$, where X signifies input pattern space (Smola and Scholkopf 1998). In a higher-dimensional feature space, hypothesis space of SVM is restricted to linear functions which can be formulated by following equation (Haykin 1999):

$$\min \emptyset(w, \xi) = \frac{1}{2} w^2 + C \sum_{i=1}^n \xi_i + \xi_i^* \quad (1)$$

Subject to: $y_i(w^T x_i + b) > 1 - \xi_i; \xi \geq 0; i = 1, \dots, n$

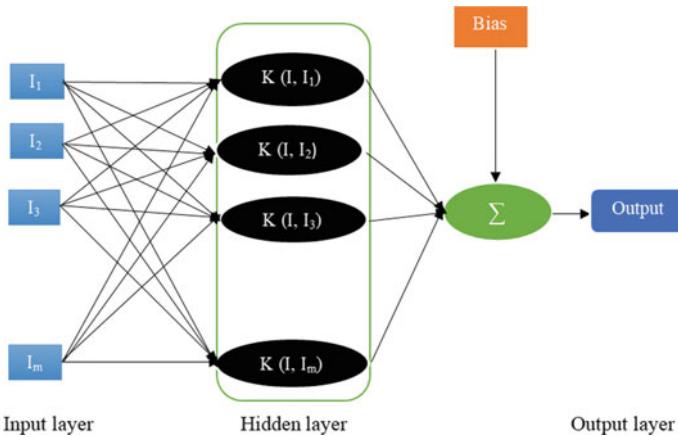


Fig. 2 Architecture of SVM

where \emptyset —objective function, w —weight vector, b —bias, ξ_i —slack variable and C —regularised parameter determining trade-off amid empirical error and regularised term (Fig. 2).

3.2 Artificial Bee Colony

Karaboga (2005) initially proposed ABC algorithm. In ABC, location of each source of food signifies a possible solution of optimisation problem. Enhancement of each source resembles to suitability of each solution. Firstly, initialisation of possible solution x_i ($i = 1, 2, \dots, N$) is done in D-dimension space, and every source of food entices one employed bee so that location of source is same as that of employed bee location. Onlooker bees find location based on selection probability of fitness values which was remembered by employed bees:

$$P_i = \frac{\text{fit}_i}{\sum_{n=1}^N \text{fit}_n} \quad (2)$$

where fit_i —fitness of i th food source; P_i —probability of selection. Subsequently, onlooker bees search neighbourhoods in accordance to following formula:

$$x_i^{k+1} = x_i^k + r \times (x_i^k - x_i^j) \quad (3)$$

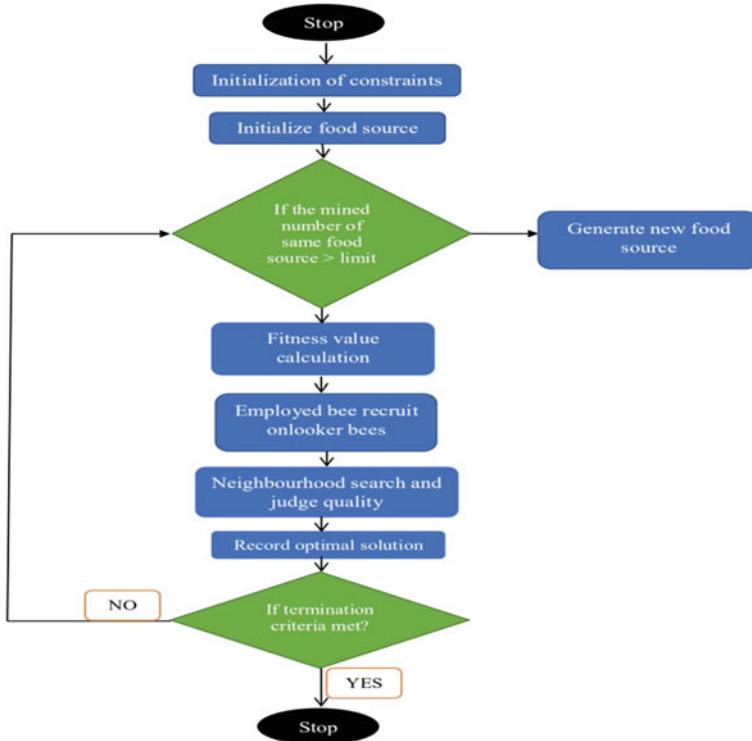


Fig. 3 Flow char of ABC

where $i \neq j$, $j = 1, 2, \dots, N$, x_k^i —position of present food source, and r —arbitrary number lies between (-1 and 1) range. Then, comparison is done between fitness values before and after searching, and we choose the superior one (Fig. 3).

3.3 Model Preparation

Monthly stream flow data with up to six lag time is collected from IMD, Pune, for a period of 1984–2010. Data values from 1984 to 2010 are used for training, and those from 2011 to 2019 are used for testing model effectiveness. RMSE, WI and R^2 are applied for performance assessment of models.

4 Results and Discussions

The performances of SVM and SVM-ABC model for different input are given in Table 1. When SVM6 model was considered, principal valuation of R^2 and WI is 0.9467 and 0.9501, 0.9462 and 0.9486 for testing phases at Sundargarh and Salebhata stations, respectively. Correspondingly at Sundargarh and Salebhata stations, the prominent value of WI and R^2 are 0.9813 and 0.977, 0.9767 and 0.9717, respectively, during testing phases for SVM-ABC6 model. The complete results of both the gauge stations are shown in Table 1.

Actual verses predicted stream flow of Sundargarh and Salebhata watershed using SVM-ABC and SVM methods for training and testing stages are shown in Fig. 4. Deviation of actual and predicted stream flow is revealed in Fig. 6. Anticipated crest

Table 1 Outcomes of SVM

Station	Model	Training			Testing		
		RMSE	WI	R^2	RMSE	WI	R^2
Sundargarh	SVM1	14.54	0.9197	0.92376	15.73	0.9153	0.91019
	SVM2	17.03	0.9265	0.92895	18.34	0.9238	0.91948
	SVM3	18.79	0.9342	0.93208	19.06	0.9299	0.92502
	SVM4	19.84	0.9399	0.93894	20.58	0.9364	0.93107
	SVM5	21.87	0.9457	0.94786	22.69	0.9429	0.93846
	SVM6	23.45	0.9554	0.95643	24.47	0.9501	0.94673
	SVM-ABC1	9.06	0.9594	0.95997	10.34	0.9565	0.95274
	SVM-ABC2	8.25	0.9685	0.96674	9.11	0.964	0.95997
	SVM-ABC3	6.99	0.9723	0.97159	7.68	0.9689	0.96453
	SVM-ABC4	5.47	0.9776	0.97751	5.99	0.9736	0.96996
Salebhata	SVM5	2.75	0.9812	0.98043	3.36	0.9778	0.97328
	SVM6	1.86	0.9867	0.98406	2.63	0.9813	0.97709
	SVM1	15.82	0.9195	0.91891	16.9	0.9164	0.91004
	SVM2	17.74	0.9276	0.92308	18.74	0.9223	0.91709
	SVM3	19.05	0.9301	0.92992	21.03	0.9278	0.92163
	SVM4	20.56	0.9389	0.93205	22.75	0.9341	0.92996
	SVM5	22.9	0.9457	0.94386	23.67	0.9399	0.93702
	SVM6	23.89	0.9528	0.95264	24.74	0.9486	0.94625
	SVM-ABC1	10.57	0.9583	0.95583	11.6	0.9554	0.95004
	SVM-ABC2	8.68	0.9642	0.96286	9.89	0.9588	0.95764
	SVM-ABC3	7.72	0.9698	0.96705	8.32	0.9663	0.96218
	SVM-ABC4	6.19	0.9725	0.97108	6.78	0.9685	0.96476
	SVM-ABC5	3.98	0.9791	0.97694	3.97	0.9751	0.97003
	SVM-ABC6	2.37	0.9839	0.98007	3.22	0.9767	0.97179

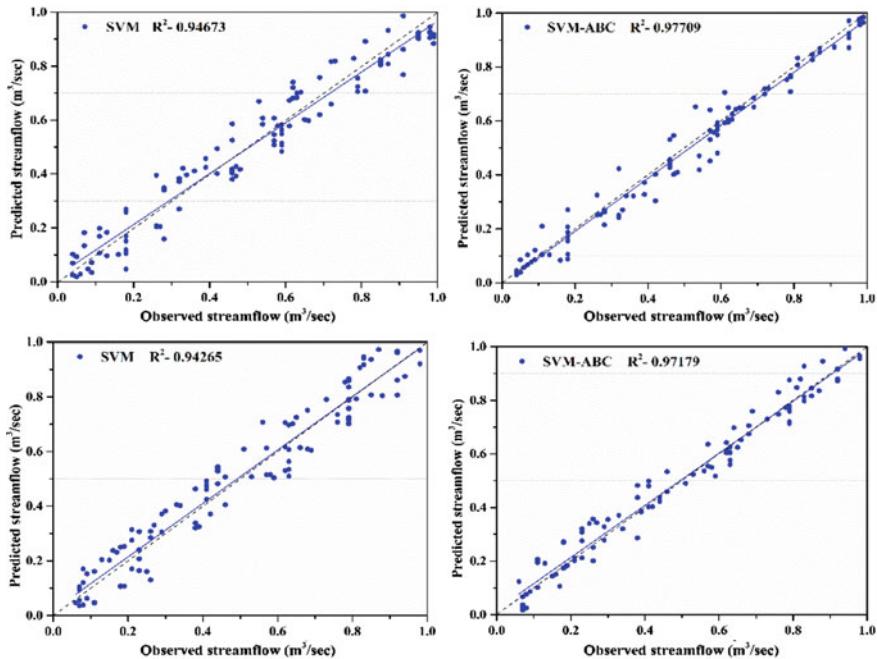


Fig. 4 Scatter plot of observed vs. predicted stream flow model at Sundargarh and Salebhata during testing phase

stream flow for Sundargarh watershed are $4837.76 \text{ m}^3/\text{s}$, and $4688.25 \text{ m}^3/\text{s}$ mm for SVM–ABC and SVM against real peak stream flow, $4950.64 \text{ m}^3/\text{s}$ mm respectively. Similarly for Salebhata gauge station, the peak stream flow value is $5567.64 \text{ m}^3/\text{s}$, whereas for SVM–ABC and SVM, peak stream flow is 5412.303 and $5251.39 \text{ m}^3/\text{s}$. Also, comparison of various models for propose station is mentioned via box plot (Figs. 5 and 6).

5 Conclusion

Monthly stream flow forecasting is vital in hydrological practices. In present study, accuracy of SVM–ABC model has been investigated for forecasting monthly stream flow at proposed gauge stations. The SVM–ABC model was obtained by integrating SVM with ABC optimisation algorithm. According to results, it was found that SVM–ABC models for all different input combinations provided better prediction results compared to standalone SVM models for monthly stream flow prediction. SVM–ABC model improved average error value (RMSE) with respect to the conventional SVM model by 3.2–2.8% for the training and testing data sets, respectively. The results indicated that SVM–ABC outperformed SVM model with better accuracy.

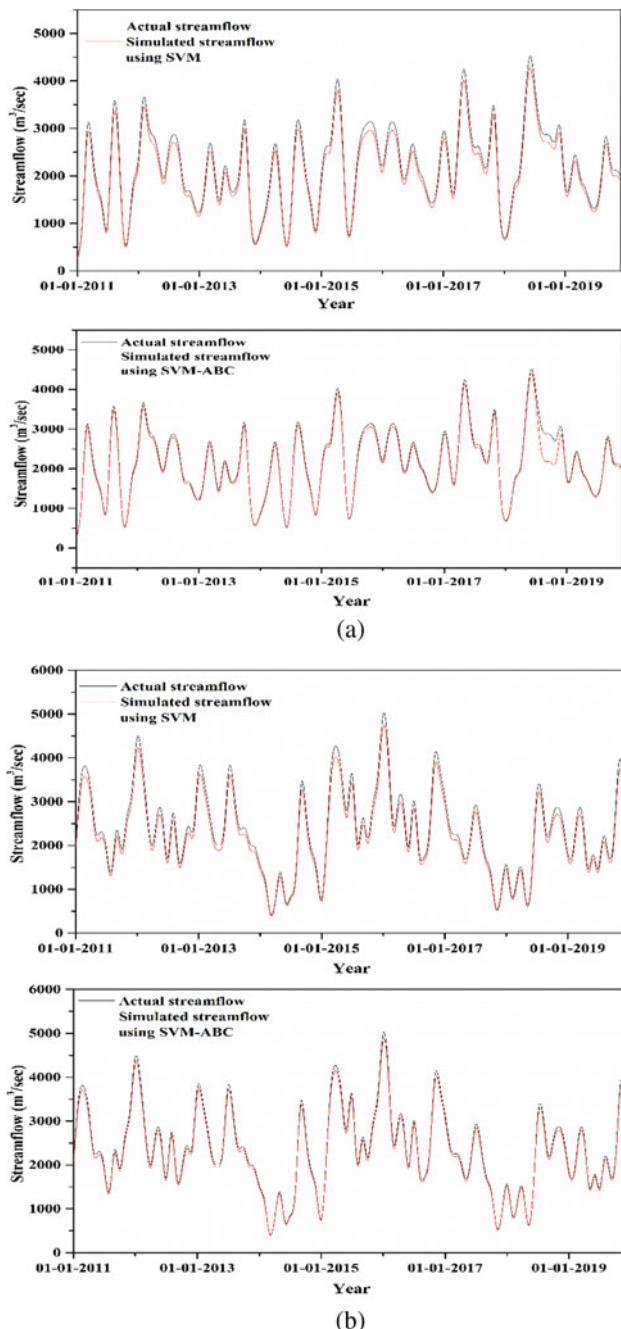


Fig. 5 Actual vs. simulated steam flow using SVM and SVM-ABC at Sundargarh and Salebhata station

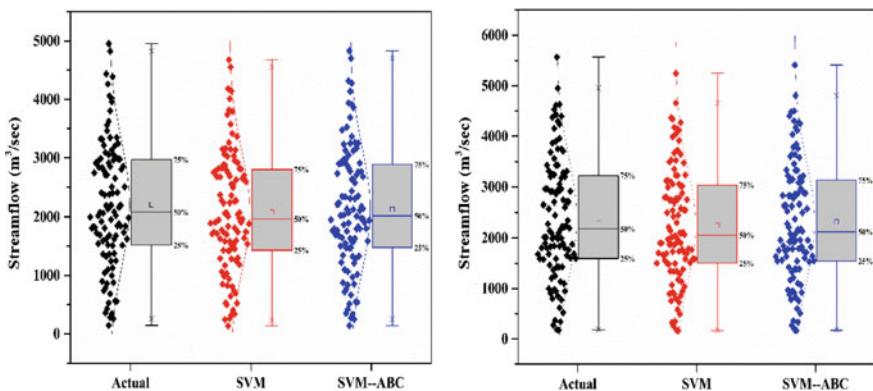


Fig. 6 Box plot of comparison for proposed method at Sundargarh and Salebhata

References

1. Sahoo A, Samantaray S, Ghose DK (2021) Prediction of flood in Barak River using hybrid machine learning approaches: a case study. *J Geol Soc India* 97(2):186–198
2. Sahoo A, Samantaray S, Ghose DK (2019) Stream flow forecasting in mahanadi river basin using artificial neural networks. *Procedia Comput Sci* 157:168–174
3. Samantaray S, Sahoo A, Ghose DK (2020) Assessment of sediment load concentration using SVM, SVM-FFA and PSR-SVM-FFA in Arid Watershed, India: a case study. *KSCE J Civ Eng* 24(6):1944–1957
4. Bai Y, Chen Z, Xie J, Li C (2016) Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *J Hydrol* 532:193–206
5. Du J, Liu Y, Yu Y, Yan W (2017) A prediction of precipitation data based on support vector machine and particle swarm optimization (PSO-SVM) algorithms. *Algorithms* 10(2):57
6. Sridharam S, Sahoo A, Samantaray S, Ghose DK (2020) Estimation of water table depth using wavelet-ANFIS: a case study. In: *Communication software and networks* (pp. 747–754). Springer, Singapore (2020)
7. Rath A, Samantaray S, Bhoi KS, Swain PC (2017) Flow forecasting of hirakud reservoir with ARIMA model. International conference on energy, communication, data analytics and soft computing (ICECDS). IEEE, pp 2952–2960
8. Samantaray S, Sahoo A, Ghose DK (2019) Assessment of runoff via precipitation using neural networks: watershed modelling for developing environment in arid region. *Pertanika J Sci Technol* 27(4):2245–2263
9. Samantaray S, Sahoo A (2020) Prediction of runoff using BPNN, FFBPNN, CFBPNN algorithm in arid watershed: a case study. *Int J Knowl Based Intell Eng Syst* 24(3):243–251
10. Samantaray S, Tripathy O, Sahoo A, Ghose DK (2020) Rainfall forecasting through ANN and SVM in Bolangir Watershed, India. In: *Smart intelligent computing and applications* (pp. 767–774). Springer, Singapore
11. Bafithile TM, Li Z (2019) Applicability of ϵ -support vector machine and artificial neural network for flood forecasting in humid, Semi-Humid and Semi-Arid Basins in China. *Water* 11(1):85
12. Natarajan N, Sudheer C (2020) Groundwater level forecasting using soft computing techniques. *Neural Comput Appl* 32(12):7691–7708
13. Guo J, Zhou J, Qin H, Zou Q, Li Q (2011) Monthly streamflow forecasting based on improved support vector machine model. *Expert Syst Appl* 38(10):13073–13081

14. Kisi O, Nia AM, Gosheh MG, Tajabadi MRJ, Ahmadi A (2012) Intermittent streamflow forecasting by using several data driven techniques. *Water Resour Manage* 26(2):457–474
15. Shabri A, Suhartono (2012) Streamflow forecasting using least-squares support vector machines. *Hydrol Sci J* 57(7):1275–1293
16. Mandal SK, Chan FT, Tiwari MK (2012) Leak detection of pipeline: an integrated approach of rough set theory and artificial bee colony trained SVM. *Expert Syst Appl* 39(3):3071–3080
17. Kuo RJ, Huang SL, Zulvia FE, Liao TW (2018) Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction. *Knowl Inf Syst* 55(1):253–274
18. Farfán JF, Cea L (2021) Coupling artificial neural networks with the artificial bee colony algorithm for global calibration of hydrological models. *Neural Comput Appl*, pp 1–16
19. Yıldız C, Açıkgöz H (2021) Forecasting diversion type hydropower plant generations using an artificial bee colony based extreme learning machine method. *Energy Sources, Part B: Economics, Planning, and Policy*, pp 1–19 (2021)
20. Samantaray S, Sahoo A (2020b) Assessment of sediment concentration through RBNN and SVM-FFA in Arid Watershed, India. In: Smart intelligent computing and applications (pp. 701–709). Springer, Singapore
21. Sahoo A, Barik A, Samantaray S, Ghose DK (2020) Prediction of sedimentation in a watershed using RNN and SVM. In: Communication software and networks (pp 701–708). Springer, Singapore
22. Samantaray S, Ghose DK (2020) Modelling runoff in an arid watershed through integrated support vector machine. *h2oj* 3(1):256–275

Optimizing Power Saving in Cloud Computing Environments Through Server Consolidation



Parthasarathi Pattnayak

Abstract In cloud computing, surroundings power saving is a multifaceted challenge, as raising the system reliability, which can directly decrease the carbon dioxide emission and the in-use cost (Parthasarathi and Pamela in IJLTE 017–025 [1]). This paper suggests ways for saving the power through server consolidation in a cloud computing environment by minimizing the unused number of resources. Server consolidation is a resourceful method that increases the utilization of system resource. Here, two types of energy-aware heuristics are analysed for consolidating the tasks. In this paper, an algorithm called bi-objective server consolidation, combining both heuristics, is suggested in order to increase the energy-efficient consolidation of the server. This method increases the efficiency of energy without reducing the performance of the server.

Keywords Server consolidation · Cloud computing · Virtualization VMs · Energy consumption · Bi-objective server consolidation

P. Pattnayak (✉)

School of Computer Applications, KIIT Deemed to be University, Bhubaneswar, Odisha, India

1 Introduction

A computing paradigm for cloud computing has been developed well through software applications, hardware capacities and network devices. The resources can range from several physical servers to an intact data centre, and these computing systems can be widely distributed. These resources can be used at different levels and can be combined with the competent management methods. As a result, research has switched a few days now to efficiently use resources to minimize energy consumption in data centres. Virtualization in cloud computing is the core technology. So, several techniques have been developed and suggested to enhance the utilization of resource like memory firmness, defining request inequity among VMs and server allocation. Objective of research is to reduce the cost in consumption of power for data centres. Energy consumption is directly based on CPU utilization. But if the utilization of CPU of a system is higher which is not equal to the efficiency of energy of that particular system then the objective of cost reduction is defeated. Therefore, this procedure has inspired the proposal by maximum levels of utilization for minimizing the power. As we understand, the main focus of workload consolidation problem is server or server consolidation. Our paper addresses this problem.

The rest of this paper is structured accordingly. Section 2 provides an overview of the related work involved. The model of Cloud system, server consolidation problem is described in Sects. 3 and 4. Section 5 describe the algorithm of Server consolidation, while Sects. 6 and 7 describe bi-objective procedure, the Bi-objective Server Consolidation Algorithm, and Sect. 8 explains results. Section 9 concludes the study.

2 Related Works

Green computing and cloud computing examples are directly related. Today, the most crucial research issues are the energy efficiency in cloud computing. Low-powered CPUs, solid state drives along with energy-efficient computer monitors are certain degree hardware technologies, which have reassured the power problem [2]. In the meantime, a substantial quantity of software moves towards server consolidation ([3–6]) and resource allocation and scheduling ([7, 8] in [8] move towards using usual problem of bin-packing with two major uniqueness in the task consolidation: (1) disk usage and (2) CPU usage. The task consolidation mechanisms deal with the energy minimization using unusual approaches, i.e., compression of memory and discriminatory request ([9, 10]). Substantial research has been carried out on software approaches, such as planning and allocation of resources ([11, 12]); ([11, 12]). In the consolidation of tasks, virtualized technologies remain crucial. Imbalance/fluctuation in load distribution is responsible for drainage. It has led to several studies relating to consolidation of tasks ([10–13]). The mechanisms for consolidation of tasks detailed in [13] deal, particularly in [13], with energy reduction using extraordinary methods.

3 Model of Cloud System

The cloud that defines the job consolidation problem and function with energy models has been described in this section. Through virtualization, we create a single physical machine (a set of VMs) for developing the return on investment by the cloud providers. The power consumption can be condensed by turning off nodes that are idle. It is necessarily implying that we require processing time and processor utilization as parameters for the determination of power consumption for a particular job. The utilization U_i of r_i resource at any specified time is defined as

$$U_i = \sum_{j=0}^{n-1} u_{i,j} \quad (1)$$

Here, $u_{i,j}$ is the resource usage of a task t_j , and the number of jobs running at prearranged time is n . At any given time, E_i is the power consumption of r_i resource and is defined as:

$$E_i = (p_{\max} - p_{\min}) \times U_i + p_{\min} \quad (2)$$

Here, in dynamic mode, p_{\min} represents minimum energy consumption at a low of 1%, and p_{\max} represents maximum energy consumption at a high of 100% utilization. As a result, E_R is the whole power consumption as the whole utilization U_R at any given time is defined as:

$$E_R = \sum_{i=0}^{m-1} E_i \quad (3)$$

Here, m is the number of resources. A stationary basis whichever specific time has been set to 10% of p_{\min} by mostly power used. Because the overall amount of time required turning a resource off and return is unknown, the resources those are idle in nature are opted out in this study. ([10–13]). We classify VMs at six different levels in Fig. 1. One is unused, while others represent different levels of CPU use. We have “n” cloud resources from “m” tasks to maximize the use of resources in order

$$E_t(V_i) = \begin{cases} \alpha \text{ watts / s, if } idle \\ \beta + \alpha \text{ watts / s, if } 0\% < CPU \text{ utilization} \leq 50\% \\ 2\beta + \alpha \text{ watts / s, if } 50\% < CPU \text{ utilization} \leq 70\% \\ 3\beta + \alpha \text{ watts / s, if } 70\% < CPU \text{ utilization} \leq 80\% \\ 4\beta + \alpha \text{ watts / s, if } 80\% < CPU \text{ utilization} \leq 90\% \\ 5\beta + \alpha \text{ watts / s, if } 90\% < CPU \text{ utilization} \leq 100\% \end{cases}$$

Fig. 1 CPU utilization in five levels [8]

to reduce energy use with a certain time restriction. The consolidation algorithm is assigned to the server (task).

4 Server Consolidation Problem

The server consolidation problem (task consolidation/workload consolidation) is the process of assigning a set $T = \{t_0, \dots, t_{n-1}\}$ of jobs n to a set $R = \{r_0, \dots, r_{m-1}\}$ of resources m with no infringing time restriction. Primary intention is minimizing the energy consumption and maximize utilization of resources. Resources usage and time constraints are directly linked by server jobs. In the problem of consolidation, resource allocation of a particular job should adequately supply the resource of that job.

5 The Algorithm of Server Consolidation

The task allocation is a non-deterministic polynomial time issue in the cloud. Hence, heuristics algorithm has been considered to be the most efficient technique for scheduling in cloud. They are efficient because they can optimally distribute and deliver the solution. As we know, unused resources consume energy; the algorithm that we have proposed help minimize the number of unused sources by assign a job instantly to a virtual machine which is presently unused. The energy-aware server consolidation heuristics (*ECTC* & *MaxUtil*) usually submitted as cost functions [9]. *ECTC* algorithm is implemented if none of the machine is idle. The decision of server consolidation makes by *MaxUtil* based upon the resource utilization. The two cost functions (*ECTC* & *MaxUtil*) are focused on the differences, whether measure the consumption of power implicitly or explicitly. The algorithm verified each resource for a given task and spot the most efficiency of energy resources for that job, which costing of efficiency of energy resources is depend on the *MaxUtil* or *ECTC*. The algorithm explains the server consolidation procedure.

Algorithm 1

Sever Consolidation Algorithm

```

input:  $t_j \in \tau = \{t_0, \dots, t_{n-1}\}$ ,  $\mathcal{R} = \{r_0, \dots, r_{m-1}\}$ 
output:  $r^* \in \mathcal{R}$ 
begin
     $r^* \leftarrow \phi$ 
    for all the  $r_i \in \mathcal{R}$  do
        Compute the cost function value  $f_{i,j}$  of  $t_j$  on  $r_i$ 
        if  $f_{i,j} > f_{*,j}$  then
             $r^* \leftarrow r_i$ 
             $f_{*,j} \leftarrow f_{i,j}$ 
        Assign  $t_j$  to  $r^*$ 

```

Table 1 provides four specifically selected task properties (as a working example) for indicating ECTC and MaxUtil's divergent behaviour. We specified the time of arrival, the processing time (Sect. 0), and the use of resources for each task (t_j) (U_j). In the example, p_{\min} is presumed to be 20 and p_{\max} 30. For example, these values are considered to be rough estimates on current resources, respectively, of 200 watts and 300 watts. In accordance with the respective properties presented in Table 1, the more energy-efficient resource (r_i) will be assigned to each task(t_j). Fig. 2 depicts share of the first three jobs, where job t_3 examples the different as of the result get from the relevant cost functions. In Fig. 2a, power consumption of the job, t_3 assigns by ECTC with the r_1 resource, while accessible rate of the resource utilization, t_3 assigns by MaxUtil with r_0 resource in Fig. 2b. When job t_4 should be assigned to a resource, the difference between the two function becomes more important. While example in Fig. 3a, t_4 assigns by ECTC with r_2 empty resource in Fig. 3b. As t_4 assigns by MaxUtil with r_1 in Fig. 3b. In this instance, ECTC appeared minimum "Energy-efficient" than MaxUtil.

Table 1 Job properties

Job (t_j)	Arrival Time (a_j)(s)	Processing Time (τ_0) (s)	Utilization (U_j) (%)
0	00	20	40
1	03	08	50
2	07	23	20
3	14	10	40
4	20	15	70

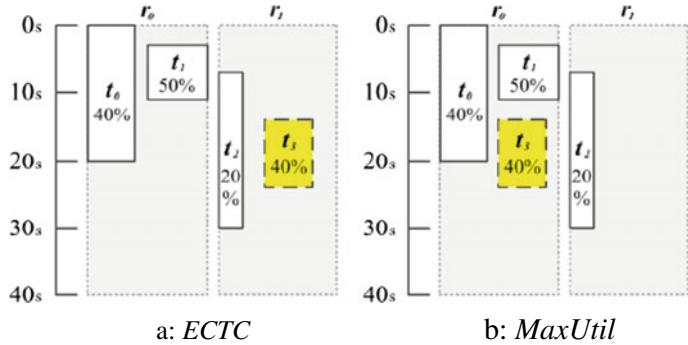


Fig. 2 Representation of the 1st three job

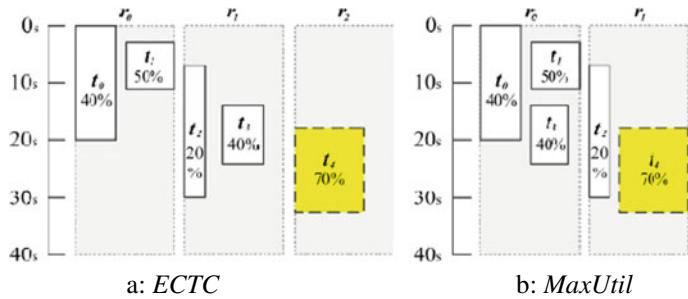


Fig. 3 Concluding representation for every task

6 Bi-objective Procedure

The benefits of combining two cost functions contribute to a reflection on the two-objective procedure. The algorithm gives extra “Energy-efficient” based on both aspects.

Algorithm 2

Bi-objective Procedure

```

Input:  $t_j \in \tau = \{t_0, \dots, t_{n-1}\}$ ,  $\mathfrak{R} = \{r_0, \dots, r_{m-1}\}$ 
Output:  $r^* \in \mathfrak{R}$ 
begin
     $r^*, optimum \leftarrow \phi$ 
    for all the  $r \in \mathfrak{R}$  do
         $x \leftarrow f_x$ 
         $y \leftarrow f_y$ 
         $result \leftarrow (x, y)$ 
        if  $result \neq optimum$  then
             $optimum \leftarrow result$ 
             $r^* \leftarrow r$ 

```

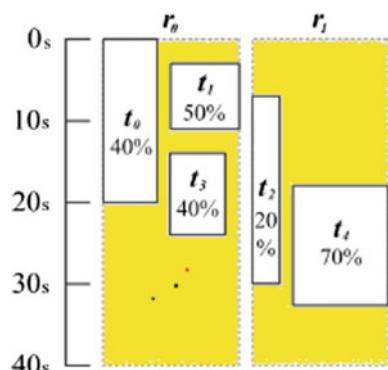
In terms of energy consumption and the use of resources, this two-target model is designed to combine the two cost functions so that they can only take advantage of their benefit. This means that the algorithm will provide the “energy efficient” resource on both aspects. In other words, the consumption (sole) of energy as the use of resources.

7 The Bi-objective Server Consolidation Algorithm

The bi-objective server consolidation algorithm uses two heuristics ECTC and MaxUtil to construct the proposed algorithm.

In Fig. 4, the four jobs allocated just a few resources. Another example of the most favourable situation provided by this algorithm minimizes energy consumption.

Fig. 4 Final results represent from bi-objective algorithm



The results were the same as in Fig. 3b as the MaxUtil for every prearranged job bi-objective server consolidation.

Algorithm 3

Bi-objective Server Consolidation Algorithm

```

input  $t_j \in T = \{t_0, \dots, t_{n-1}\}, R = \{r_0, \dots, r_{m-1}\}$ 
output  $r^* \in R$ 
begin
     $r^*, optimum, F \leftarrow \phi$  for all the  $r \in R$  do
         $x \leftarrow f_x, y \leftarrow f_y$ 
         $\delta \leftarrow y_{\max} - |x - y|$ 
         $result \leftarrow x, y)$ 
        if  $result > optimum$  then
             $optimum \leftarrow result$ 
             $r^* \leftarrow r, F \leftarrow \phi$ 
             $F \leftarrow F \cup result, r, \delta)$ 
        if  $((result > optimum) \wedge optimum > result))$  then
             $F \leftarrow F \cup result, r, \delta)$ 
        for all the  $f, r, \delta \in F$  do
            if  $((f_x \neq optimum_x) \wedge f_y \neq optimum_y))$  then
                if  $((\delta > \delta_{optimum})$  then
                     $optimum \leftarrow f$ 
                     $\delta_{optimum} \leftarrow \delta$ 
                     $r^* \leftarrow r$ 
                elseif  $((f_x + f_y) > optimum_x + optimum_y))$  then
                     $optimum \leftarrow f$ 
                     $\delta_{optimum} \leftarrow \delta$ 
                     $r^* \leftarrow r$ 
            
```

8 Results

Energy efficiency of MaxUtil, ECTC and algorithm for Bio-objective task consolidation (BTC) are evaluated. The following figures are used for this study: of the four selected topology, Fig. 5a, b showed six graphs showing the behaviour of energy efficiency and task consolidation.

They were as follows: (a) ten resources and twenty resources with four cores, and (b) five resources and ten resources with eight cores.

In the figures above, the solid (blue) line represents total system usage rate, while the slashed (red) line represents energy usage. At each task assignment, sampling

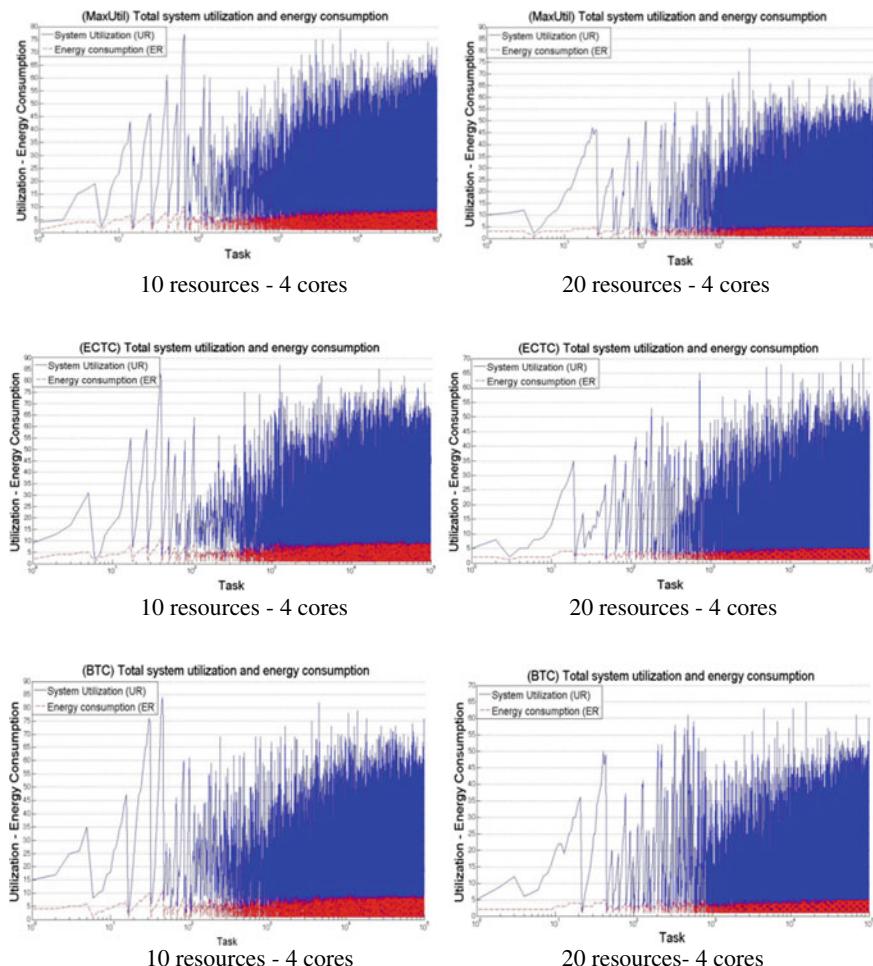


Fig. 5 **a** 10 resources, and 20 resources with 4 cores in MaxUtil, ECTC and BTC Algorithim, **b** 5 resources, and 10 resources with 8 Cores in MaxUtil, ECTC and BTC Algorithm

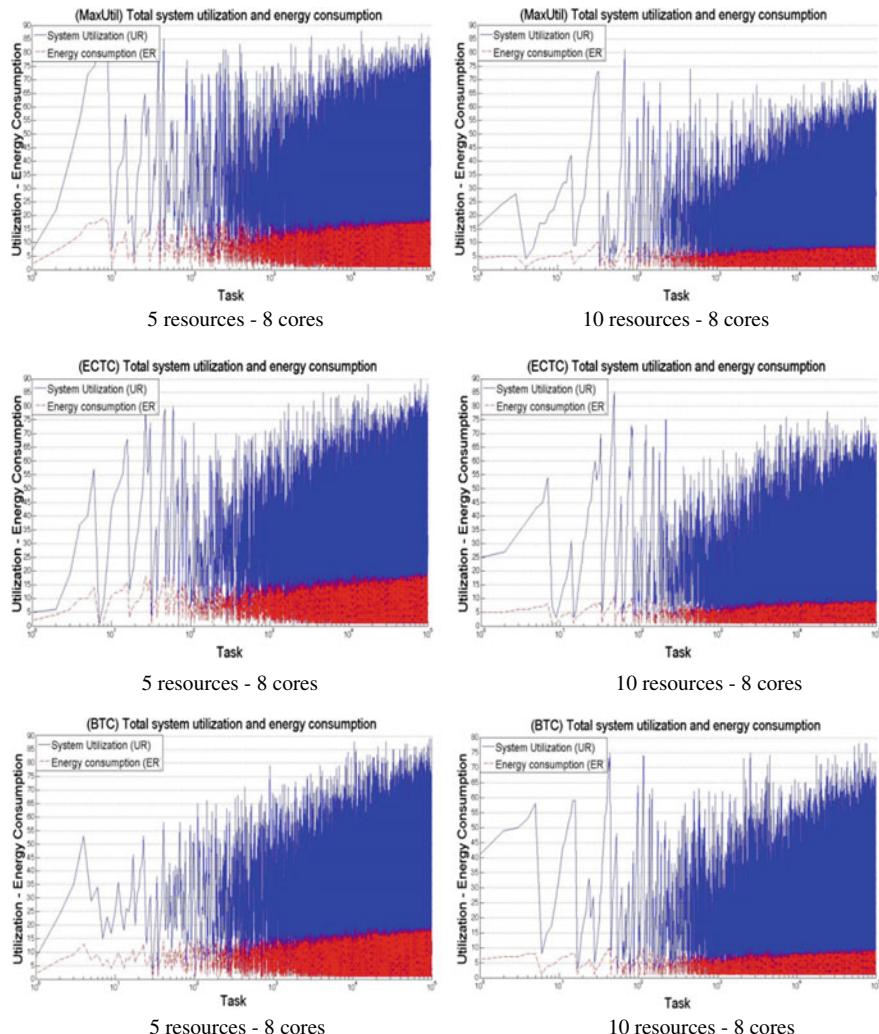


Fig. 5 (continued)

rates for the system status collection were performed. The y-axis shows the usage as a unit scale (normalized), whereas a logarithmic x-axis scale showed the task index. The corresponding figures are Fig. 5a, b, respectively, for MaxUtil, ECTC and BTC algorithm. No significant differences between the three heuristics that correspond to the problem of consolation can be noted from our results. Therefore, “efficient energy” can be considered equal to MaxUtil, ECTC and BTC algorithm.

9 Conclusion

The two heuristics were implemented separately to calculate the result of the bi-target server consolidation algorithm. The consolidation of the bi-objective algorithm provides the optimum heuristically energy-efficient solution. The mixture of the method presented to optimize models for energy-efficient use without performance decreases. Task consolidation has become an important approach to streamline resource use, in turn, improvement of energy efficiency, particularly in cloud computing systems. We analysed two existing energy-conscious task consolidation heuristics, which offer different energy savings options. We identified the respective disadvantage for both heuristic systems and proposed the bi-target task consolidation (BTC) algorithm as a (single) solution. The algorithm above combines the two heuristics to create the respective two-target search room. We observed three major aspects of each task in terms of the system's energy efficiency and scalability like overall consumption of energy, total usage of resources and also time period required for solving the problem optimally. Two heuristics are implemented separately. They are considered as the primary index of both scalability and energy efficiency for evaluating the functions of BTC algorithm.

References

1. Parthasarathi P, Pamela P (2017) Energy—efficient cloud computing with task consolidation. IJLTE 017-025. e-ISSN:2278-621X
2. El Kafhali S, Salah K (2017) Stochastic modelling and analysis of cloud computing data center. In: Proceedings of 20th conference on innovations in clouds, internet and networks. IEEE, pp 122–126
3. Patel N, Patel H (2017) Energy efficient strategy for placement of virtual machines selected from under loaded servers in compute Cloud. J King Saud Univ Comput Inf Sci. <https://doi.org/10.1016/j.jksuci.2017.11.003>
4. Lee YC, Zomaya AY (2010) Energy efficient utilization of resources in cloud computing systems. Published online: 19 Mar 2010
5. Yuan H, Bi J, Zhou M (2019) Spatial task scheduling for cost minimization in distributed green cloud data centers. IEEE Trans Autom Sci Eng 16(2):729–740
6. Zhang P, Zhou MC (2018) Dynamic cloud task scheduling based on a two-stage strategy. IEEE Trans Autom Sci Eng 15(2):772–783
7. Li Z, Yan C, Yu L, Yu X (2018) Energy-aware and multi-resource overload probability constraint-based virtual machine dynamic consolidation method. Future Gener Comput Syst 80:139–156
8. Wang H, Tianfield H (2018) Energy-aware dynamic virtual machine consolidation for cloud datacenters. IEEE Access 6:15259–15273
9. Hanini M, El Kafhali S, Salah K (2019) Dynamic VM allocation and traffic control to manage QoS and energy consumption in cloud computing environment. Int J Comput Appl Technol 60(4):307–316
10. Lee YC, Zomaya AY (2009) Minimizing energy consumption for precedence-constrained applications using dynamic voltage scaling. In: 9th IEEE/ACM international symposium on cluster computing and the grid (CCGrid), Shanghai, China, pp 92–99

11. Srikanthaiah S, Kansal A, Zhao F (2008) Energy aware consolidation for cloud computing. In: Proceeding USENIX HotPower08: workshop on power aware computing and systems in conjunction with OSDI, San Diego, CA, USA, Dec 2008
12. Torres J, Carrera D, Hogan K, Gavalda R, Beltran V, Poggi N (2008) Reducing wasted resources to help achieve green data centers. In: Proceedings of 4th workshop on high-performance, power-aware computing (HPPAC'08)
13. Tang Q, Gupta SK, Vassamopoulos G (2008) Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: a cyber-physical approach. IEEE Trans Parallel Distrib Syst 19(11):1458–1472

Siamese Network with Transfer Learning for Similar Query Retrieval in Online Health Community Forums



B. Athira and Sumam Mary Idicula

Abstract In health community forums, patients seek information, acquire knowledge and support each other. They find help by asking questions about their illness. Finding similar people with similar issues to obtain correct information is often a great challenge for them. Hence, this paper proposes a query similarity model to retrieve the similar query that is already discussed. The model is implemented with a Siamese neural network to assess the semantic similarity between the queries. The presence of several medical terms imposed great challenge to the model to generalize the query well. Hence, the model is fine-tuned with a transfer learning technique on a medical question–answer dataset. Several variants of the model are compared, and the best model could achieve an F1-score of 82.4%. The result showed that the proposed method is very effective for natural language understanding in medical domain.

Keywords Query similarity · Siamese network · Online health community · Transfer learning

1 Introduction

In today's digital world, people are turning to Internet for advice, particularly when they diagnosed with some life-threatening diseases. Earlier, patients were just acted as passive health information recipients. But now, focus of health care is shifted from clinician to patient-centered, where they seek information and even provide advice/suggestions. The World Health Organization has stated that patient involvement in care is not only desirable but also a social, economic and technical necessity [1]. Often, clinicians tend to focus on clinical impact of disease and may ignore

B. Athira (✉)

Department of CS & IT, Amrita School of Arts and Science (Amrita Vishwa Vidyapeetham), Kochi, Kerala, India

S. M. Idicula

Department of Computer Science and Engineering, Muthoot Institute of Technology and Science, Kochi, Kerala, India

patient's well-being and daily life. Online health communities (OHCs) are such platforms where patients form groups and discuss their various medical concerns. Even though doctors and other health professionals are the final decision-makers, advice/recommendations from similar patients also have a significant role [2]. OHCs contain multiple message boards or forums formed by patients [3]. OHCs contain different threads started by a patient with a query, and responses are posted to that thread. If the patient can locate the query that is already discussed, he/she can experience a fast and effective information search. The authors considered a large breast cancer forum Breastcancer.org as OHC for the current study. The survival rate of breast cancer is steadily increasing due to early diagnosis and better treatment. Still, it is a great challenge for these patients to handle the condition and maintain the quality of daily life [4]. Hence, seeking and providing support for similar patients is very essential for patients living with cancer. However, a user having a similar problem, being already discussed in the thread, may not want to go through all threads. As a result, a similar query retrieval system can assist patients in conducting a quick information search, allowing them to make a quick and informed decision.

To retrieve the most relevant query to a user's new query from among the similar posted queries, it is essential to measure the semantic relatedness between them. For instance, the studies in [5, 6] are such types of question–question matching task on online user forums. However, these approaches do not perform well in online health community forums owing to the presence of many medically related terms. One of the previous works [7] has tried to generate medical query pairs using augmentation rules. But that dataset follows similar lexical structures similar question pairs and no overlapping keywords for dissimilar pairs. Many recent studies [8–10] investigated the usage of re-trained BERT on different text corpora. But this is very time-consuming to use in every new domain. However, our investigation was to make use of fine-tuning in similar domain tasks.

A deep learning architecture, a Siamese network, is used in the current study to find the closest query. Siamese network uses a similar neural network architecture for extracting features of both queries and compares them using a distance measure. To assimilate medical knowledge contained in the query to the deep learning model, a transfer learning technique is used in the current study. In the case of non-similarity queries, it can be noted the moderator/experts in the forums to respond. Thus, the main objective of the study is to retrieve a similar query that was already discussed in the forum. Since query consisted of various medical-related terms, it is also an objective to imbibe medical knowledge into the model.

2 Related Studies

Semantic similarity between sentences has been extensively studied by researchers for a long time. Previously, the main focus of researchers was on linguistic characteristics such as word overlap together with traditional classifiers of machine learning such as SVM [11]. But subsequently, some word embedding mechanisms such as

word2vec [12], Doc2vec [13] and Sent2vec [14] have introduced a revolutionary approach to semantic similarity tasks. Following this trend, the most recent additions to these methods were BERT and ELMO [15, 16], two contextual word embeddings. As a part of the deep learning uprising, researchers now rely on deep neural networks to address a wide variety of NLP tasks. To name a few, CNN, GRU, RNN and Tree-LSTM were architectures of this kind used for textual similarity tasks [17, 18].

A study in [5] used two identical LSTMs called Siamese network to access the semantic relatedness and reported more promising results than the above neural networks. In a very recent similar study [19], authors proposed a Siamese text matching transformer model to predict the similarity between Chinese medical questions. They used transformer as encoder to obtain the query representation. Instead, the current study made use of contextual embedding techniques of BERT and BioBERT to embed the queries and used Siamese neural network to obtain the final representation. However, one of the essential criteria for deep learning is the availability of a vast amount of training data. Therefore, a large medical query pair training dataset is very critical to the performance of deep learning models in the medical domain. This challenge has recently been well addressed by the implementation of a transfer learning technique [20]. The methodology was implemented with a double fine-tuning approach, where the model was first trained with large medical question-answer dataset and then fine-tuned with a small query–query dataset. This research has inspired us to follow a double fine-tuning approach to imbibe the medical knowledge found in the query into the model.

3 Methodology

3.1 Dataset

The dataset used for the present study, as described earlier, was collected from a large OHC, Breastcancer.org [21]. The website is organized into numerous forums, each containing hundreds of threads. Each thread begins with a user query followed by responses from survivors/other patients in the form of posts. Nearly, 300 queries were collected from the threads and saved it in the database. A similar and dissimilar query pair was built for each query. If a thread from which the query was taken contains similar query, then that one was considered as similar query. If no such similar query is found, then rewrite the original query differently by changing the structure of the query sentence. Queries from different forums were considered for dissimilar query pairs. In this way, a dataset of 300 queries was created with similar and dissimilar pairs. A sample of this dataset is shown in Table 1. For the technique of transference learning, the dataset of WebMD [22] was considered. WebMD is an online medical information publisher that includes articles, videos and frequently asked questions

Table 1 A sample of query-pairs: similar as well as dissimilar

Sample query	Similar query	Dissimilar query
Do you know the differences in the different wigs in how they're made?	My question is for those that have been through it and are currently using or have used wigs what type do you suggest?	What do you think of immediate vs delayed reconstruction?
What is your doctor's reasoning for advising that you delay reconstruction?	What do you think of immediate vs delayed reconstruction?	I'm praying that they might be willing to consider other options for rad position but doesn't look likely. Any experience?
Hello I think I have thrush??? What do I do for it???	I have thrush coats in my tongue and inside of mouth. What could be done?	My question is for those that have been through it and are currently using or have used wigs what type do you suggest?
Wanted to see if anyone else got "permission" to alter the arm position during rads?	I'm praying that they might be willing to consider other options for rad position but doesn't look likely. Any experience?	I have thrush coats in my tongue and inside of mouth. What could be done?
What is the chance of success of post-radiation reconstruction?	Has anyone done reconstruction after radiation? Is it success?	What is your doctor's reasoning for delaying reconstruction?

(FAQ). WebMD's FAQ consists of 46,872 question-answer pairs, which were used for our model's pre-training.

3.2 *Query Similarity Model Using Siamese Network*

As stated in the related study section, transfer learning on the Siamese network is carried out to determine the semantic similarity between queries. The model is first tuned to the WebMD question-answer pair dataset and then fine-tuned to our query pair dataset. In the identical sub-networks, Siamese networks, two deep neural networks, LSTM and BiLSTM, were used and the results were compared.

Figure 1 shows the architecture of the proposed Siamese network with LSTM units. It consists of an embedding layer, two identical neural network layers, LSTMs subnet 1 and subnet 2, a prediction layer and dense layer. First, both the queries, user queries and archived queries (the question that has already been replied in threads) are fed to the embedding layer. The queries are transformed into sequence of word vectors by the embedding layer using BERT and BioBERT [23] techniques. The word vector sequence is fed to both subnets, and the absolute difference of their outputs is determined and given to the prediction layer for predicting similarity. The BRETT-base model was used to embed words into vectors of dimension 768. BioBERT, a

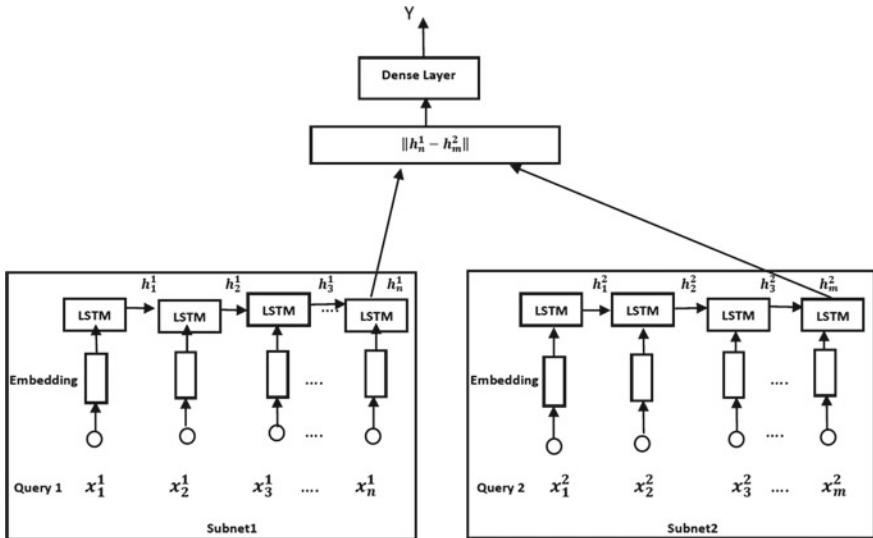


Fig. 1 Query similarity model with LSTM architecture

BRET model pre-trained on PubMed biomedical texts, was also used to evaluate the performance of the embedding technique.

Each LSTM unit includes several hidden layers of memory cells that can either recall or forget information. For each word x_i , the cell value h_i is computed using the following equation:

$$h_i = f(w_h h_{i-1} + w_x x_i) \quad (1)$$

where f is tanh activation function, w_h is the weight of the hidden layer, h_{i-1} is the previous state, w_x is the weight of current input, and x_i is the current input word vector.

The above equation is applied in all the hidden states, then the element wise absolute difference between vectors is computed by $\|h_n^1 - h_m^2\|$, and finally, the value is subsequently fed to dense layer where sigmoid function is used to get the similarity value, S . If $S \geq 0.5$, the label Y is predicted to be 1 (the queries are similar) and the label Y is predicted to be 0 (the queries are different), otherwise. The loss function for each query pair is binary cross entropy, which is defined by:

$$\text{Loss} = \hat{y} \log(S) + (1 - \hat{y}) \log(1 - S) \quad (2)$$

where \hat{y} is the true label and S is the output probability for label 1 and $(1 - S)$ is the output probability for label 0. Stochastic gradient descent method is used. By using

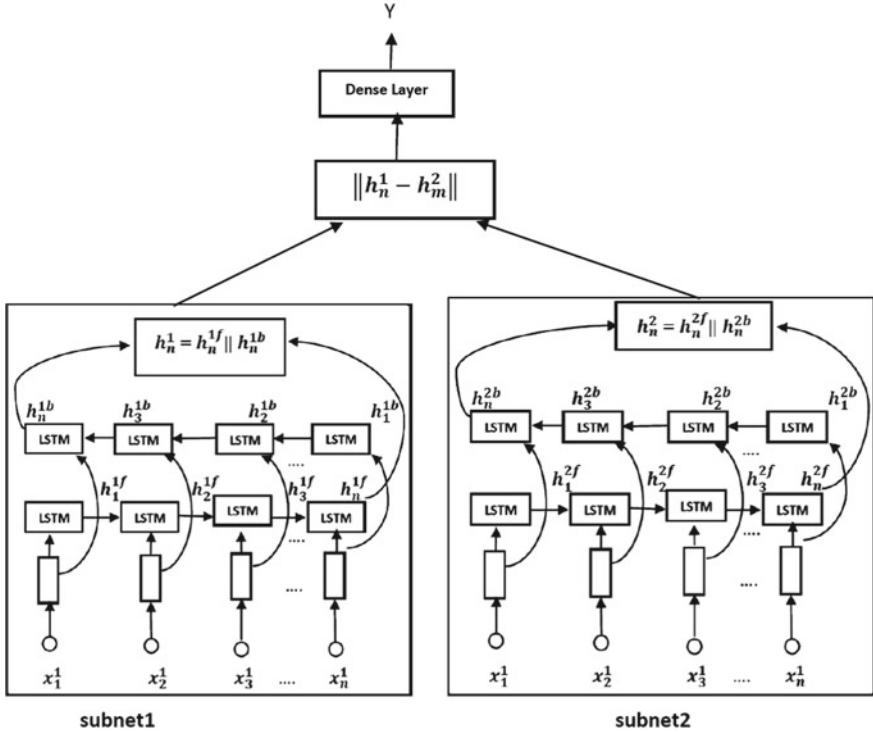


Fig. 2 Query similarity model with BiLSTM architecture

backpropagation, it is calculated. The code was written in Python using the Keras library.

Query similarity model using BiLSTM is shown in Fig. 2. Two separate forward and backward input states formed by two LSTMs constituted the BiLSTM architecture. Word sequences from left to right x_1 to x_n are fed to the forward LSTM, while word sequences from right to left x_n to x_1 are fed to the backward LSTM. As a result, the first LSTM is used to capture past data, while the second LSTM is used to extract future data. As shown in Fig. 2, the final output h_n^1 for first sub-network is obtained by concatenating two outputs from the forward LSTM h_n^{1f} and backward LSTM, h_n^{1b} by Eq. (3). The same process is occurred for second sub-network, and the final output h_m^2 is obtained by the Eq. (4). The final similarity score S and the label Y were determined in the same way as described in the LSTM network.

$$h_n^1 = h_n^{1f} \parallel h_n^{1b} \quad (3)$$

$$h_m^2 = h_m^{2f} \parallel h_m^{2b} \quad (4)$$

Transfer learning on these models is carried out using the WebMD question-answer pair dataset. The question-answer pair dataset is used to pre-train the models, with the goal of predicting the correct answer for a given query. Each question with its given answer is labeled as positive (1) and each query with a random other answer is labeled as negative (0). These labeled pairs are embedded with the technique of BERT/BioBERT and are fed to each subnets (LSTM/BiLSTM). The models were trained at 10 epochs to classify the data as either 1 or 0. After this pre-training, the models are fine-tuned to our query pair dataset.

4 Experiments and Results

4.1 Performance Evaluation

Different word embedding techniques BERT and Bio-BRET with LSTM and BiLSTM deep neural networks are used to test the model's performance. First, our query pair data collection is directly fed without any pre-training to the Siamese deep networks and the similarity is examined. The number of LSTM/BiLSTM units was 128 and the number of epochs was 10, with a learning rate of 0.01 and a drop out of 0.2. The model was trained with 80% of the query pair data and tested with 20% of the data. The results are tabulated in Table 2. Thus, there were 8 models to compare and the base model, model 4 (BiLSTM with Bio-BRET) yielded an F1-score of 65%.

Table 2 Valuation of query similarity

Techniques	Models	Precision	Recall	F1-score
Without pre-training (base model)	BC-QQP + BERT + LSTM(Model 1)	0.48	0.52	0.50
	BC-QQP + BERT + BiLSTM(Model 2)	0.53	0.50	0.51
	BC-QQP + BioBERT + LSTM(Model 3)	0.58	0.61	0.60
	BC-QQP + BioBERT + BiLSTM(Model 4)	0.63	0.66	0.65
With pre-training (WebMD)	BC-QQP + BERT + LSTM(Model 5)	0.80	0.81	0.805
	BC-QQP + BERT + BiLSTM(Model 6)	0.81	0.82	0.815
	BC-QQP + BioBERT + LSTM(Model 7)	0.83	0.81	0.82
	BC-QQP + BioBERT + BiLSTM(Model 8)	0.84	0.81	0.824

Table 3 Sample query pairs with actual and predicted labels

	Query 1	Query 2	Actual label	Labels from models	
				Model 4	Model 8
1	Does anyone else feel that on the mastectomy side things feels a bit different/bonyier than on there normal side?	My surgery side is very boniyer than the normal side. Has anyone get similar experience?	Similar	Similar	Similar
2	Wanted to see if anyone else got “permission” to alter the arm position during rads?	I’m praying that they might be willing to consider other options for rad position but doesn’t look likely. Any experience?	Similar	Dissimilar	Dissimilar
3	What is the chance of success of post-radiation reconstruction?	What is your doctor’s reasoning for delaying reconstruction?	Dissimilar	Similar	Dissimilar

To improve the performance of the model, WebMD query question-answer pair dataset was fed to the model for pre-training. As previously mentioned in the Methodology section, the models were pre-trained. Following this pre-training, the model was tuned with 80% of the query pair dataset and then tested with 20% of the dataset. As shown in Table 2 (the value in bold), now, the model is improved and a more promising result of F1-score of 82.4% was yielded from model 8.

We conducted an error analysis to provide a deeper insight into our best query similarity model, model 8. Table 3 shows some instances of query pairs with their actual labels and the labels from the query similarity models, model 4 and model 8. The reason for considering these models was the fact that BioBERT with BiLSTM architecture was better than other architectures. It was noted both the models could understand the query pairs 1 well but were unable to do so with query pairs 2. In pairs 2, though the question was about the position during rad, the terms ‘arm position’ and ‘rads position’ as well as the negation part (‘but doesn’t look likely’) could be the reason for giving wrong result from all the models. However, the model 8 provided the exact result for query pairs 3. The model was able to distinguish between the different query pairs 3, despite the fact that they included identical terms. Both queries were about reconstruction surgery. The first one was asking about success of reconstruction after radiation and the second one was about asking the reasons for delaying the reconstruction. These analyses provide us with a clear insight into the interpretability of our model and the need for additional training data.

5 Conclusion and Future Direction

The current research uses deep neural networks to show a method for detecting semantic similarities. For query similarity models in the medical domain, transfer learning on a Siamese network is very efficient. The BiLSTM architecture with BioBERT embedding outperformed all other models. The current study is a pioneer in the area of breast cancer that aims to help the patients a fast information search and thereby gain accurate information. The study's question pairs give useful insight into the different struggles that patients face during their treatment. This knowledge will help clinicians to understand the issues they have faced outside clinical matters.

Since the study was limited to 300 query pair dataset, we aim to include more number of queries in the future. Although the query similarity model performed well, there were still few instances, as shown in Table 3 in Performance Evaluation section, in which our model failed. This indicates that the model requires more tuning to handle queries that are closely related but conceptually distinct and to generalize the similarities. This implies that further tuning is needed for the model to manage certain instances. Future studies should also take care in these directions.

The present study can be extended to other health forums to obtain awareness in other areas of concerns.

References

1. Waterworth S, Luker KA (1990) Reluctant collaborators: do patients want to be involved in decisions concerning care? *J Adv Nurs* 15(8):971–976
2. Li M, Shi J, Chen Y (2019) Analyzing patient decision making in online health communities. In: 2019 IEEE international conference on healthcare informatics (ICHI) 2019 Jun 10. IEEE, pp 1–8
3. Fan H, Smith SP, Lederman R, Chang SLC (2010) Why people trust in online health communities: an integrated approach. In: Proceedings of the 21st Australasian conference on information systems, Texas (AIS'10)
4. Athira B, Idicula SM, Jones J (2020) Multi-label topic classification of patient generated content in a breast-cancer community forum. In: Proceedings of the 4th international conference on medical and health informatics, pp 266–274
5. Bogdanova D, dos Santos C, Barbosa L, Zadrozny B (2015) Detecting semantically equivalent questions in online user forums. In: Proceedings of the nineteenth conference on computational natural language learning, pp 123–131
6. Chali Y, Islam R (2018) Question-question similarity in online forums. In: Proceedings of the 10th annual meeting of the forum for information retrieval evaluation, pp 21–28
7. Li Y, Yao L, Du N, Gao J, Li Q, Meng C, Fan W (2018) Finding similar medical questions from question answering websites. arXiv preprint [arXiv:1810.05983](https://arxiv.org/abs/1810.05983)
8. Beltagy I, Cohan A, Lo K (2019) Scibert: Pretrained contextualized embeddings for scientific text. 1(3):8. arXiv preprint [arXiv:1903.10676](https://arxiv.org/abs/1903.10676)
9. Huang K, Altosaar J, Ranganath R (2019) Clinicalbert: modeling clinical notes and predicting hospital readmission. arXiv preprint [arXiv:1904.05342](https://arxiv.org/abs/1904.05342)
10. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J (2020) BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36(4):1234–1240

11. Godbole A, Dalmia A, Sahu SK (2018) Siamese neural networks with random forest for detecting duplicate question pairs. arXiv preprint [arXiv:1801.07288](https://arxiv.org/abs/1801.07288)
12. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. In: Proceedings of ICLR workshops track. arxiv.org/abs/1301.3781
13. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
14. Pagliardini M, Gupta P, Jaggi M (2017) Unsupervised learning of sentence embeddings using compositional n-gram features. arXiv preprint [arXiv:1703.02507](https://arxiv.org/abs/1703.02507)
15. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
16. Singh AK, Shashi M (2019) Vectorization of text documents for identifying unifiable news articles. Int J Adv Comput Sci Appl 10(7):305–310
17. Kiros R, Zhu Y, Salakhutdinov RR, Zemel R, Urtasun R, Torralba A, Fidler S (2015) Skip-thought vectors. Adv Neural Inf Process Syst 28:3294–3302
18. Tai KS, Socher R, Manning CD (2015) Improved semantic representations from tree-structured long short-term memory networks. In: Proceedings of ACL. arXiv preprint [arXiv:1503.00075](https://arxiv.org/abs/1503.00075)
19. Wang K, Yang, Xu G, He X (2019) Medical question retrieval based on siamese neural network and transfer learning method. In: International conference on database systems for advanced applications. Springer, Cham, pp 49–64
20. McCreery CH, Katariya N, Kannan A, Chablani M, Amatriain, X (2020) Effective transfer learning for identifying similar questions: matching user questions to COVID-19 FAQs. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp 3458–3465
21. Jones J, Pradhan M, Hosseini M, Kulanthaivel A, Hosseini M (2018) Novel approach to cluster patient-generated data into actionable topics: case study of a web-based breast cancer forum. JMIR Med Inf 6(4):e45
22. Lasse Regin Nielsen (2017) Medical question answer data. <https://github.com/LasseRegin/medical-question-answer-data>
23. Lee J et al (2019) BioBERT: pre-trained biomedical language representation model for biomedical text mining. arXiv preprint [arXiv:1901.08746](https://arxiv.org/abs/1901.08746)

Emergency Bot in Healthcare Using Industry 4.0



Sahil Sobhani, Sejal Balasaheb Shirsale, Sajal Saxena, Vartika Paharia, and Somya Goyal

Abstract The digitalization of healthcare is underway. With populations aging, chronic diseases rising and medical expenses skyrocketing, the healthcare universe is in dire need of the improvements that digitalization will bring in terms of saving costs, improved diagnostics and more effective care. Moreover, the global lack of specialists, clinical chaperons and trained professionals and experts requests for improved effectiveness and the requirement for technology to help bridge the demand-supply crevice in services. With the combination of artificial intelligence (AI), Internet of things (IoT) and ever improving imaging techniques, great leaps forward in symptomatic precision and more finely targeted therapies are also being made. Data from the healthcare sector of the digital universe is growing at a 48% annual rate, and the challenge is to decipher that surge of information into significant data. That is how AI, with its capacity to link it to data sources within IoT, can make a huge difference. Our implementation of IoT in the field of healthcare is development of an app that makes it easier for first responders of medical emergencies. Since emergencies will occur, preplanning is fundamental. At the onset of an emergency, a lot of decisions ought to be made in a brief period of time. It offers assistance the vigilante with that by providing them with resources that will help them respond to the situation in a better way.

Keywords Artificial Intelligence · IOT · Healthcare 4.0

1 Introduction

For the past couple of years, we have seen a tremendous amount of change in the process of production thanks to the digitization of manufacturing. The shift in paradigm is so huge that it is being considered another step in the stairs of industry

S. Sobhani · S. B. Shirsale · S. Saxena · V. Paharia · S. Goyal (✉)
Manipal University Jaipur, Jaipur 303007, Rajasthan, India

growth and is being labelled as Industry 4.0 to represent another generation of production and manufacturing industry. Starting from the initial stages of industrial revolution to the heavy mass production and using developments in electricity in the second, the current stage of industrial revolution will build up on the advancements of the third generation with the utilization of technology, and AI and take it to another level with automated systems being the gifts of data science and machine learning.

It was with the help of cutting-edge technology that is available with humans today that we could fight a global pandemic like COVID-19. With fast and effective communication, scientific communities all over the world could share with each other the causes and possible ways to prevent getting infected with COVID. It was because of this global coordination, that possible cures and vaccines could be developed in record time with efficacy never seen before. It also took the popularity of telemedicine to a new level, with most patients preferring to consult the doctor online rather than visiting hospitals/clinics. One of the most important applications of Industry 4.0 has been the emergence of the Internet of things (IoT). Healthcare 4.0 has also sought to address one major problem: emergency services. Often, a lot of lives are lost because patients are unable to get timely treatment during the ‘golden hour.’ Lack of resources is often blamed here, as ambulances and/or doctors are short of the WHO standards in many parts of the world. Also, the first respondents sometimes lack appropriate knowledge/experience while attending to a patient in critical need of health services.

Our work is to address the needs of first responders and try to build a more efficient work environment for them. We initially built a basic prototype of how our bot would function, the various technology it works around and other specific details and cases. We also conducted a pretty intensive literature review so that we can thoroughly understand the issues faced by first responders to medical emergencies.

In the United States of America, we are aware of the implementation of a telephone-based service called 911. It is a first responders’ service which the majority of people utilize to deal with a multitude of emergencies be it medical, criminal or any other emergency we could imagine. But like any other innovation in public care, the implementation is still not 100% secure and perfect and can improve on a variety of aspects. Our idea tends to replace the human 911 attendee with a computer bot.

Industry and Healthcare 4.0 is a great way to enhance the quality of healthcare and public safety in our country [1]. This is the focus of our project, and we are really excited to work on it and explore the world of Industry 4.0. We are hopeful that our model will prove to be helpful for the advancement of healthcare and industry 4.0.

2 Related Works

This section gives the glimpse of related work carried out in the field of research in public healthcare tabulated as Table 1.

Table 1 Literature review

S. No.	Study	Contribution	Technique used	Strength	Weakness
1	Yang et al. [2]	Evaluated factors around first response to fire emergency	Conducted survey of occupants and first responders	Reports additional factors affecting implementation	Instant actions of the first responders are not properly guided
2	Nolan et al. [3]	Emphasized the usage of computer forensics for first responders	Created a trusted set of tools for collection of data	Targets a large variety of audience	Strategies are not established properly before use
3	Greene et al. [4]	Evaluated primary objectives of users and its effect on public health	Multiphase research project and survey with subject matter experts	Entire detail of the project including the methodologic al approach is present	Sample size of the research was relatively small
4	Mohanta et al. [5]	Considered the impact of IoT and 5G in Healthcare	Various technology models around IoT	Compares pros and cons of IoT based Healthcare systems	Threats posed by data loss, and piracy
5	Steves et al. [6]	Addressed problems faced by first responders	Surveys a very large dataset of first responders	Found out unique issues and covers a large number of scenarios	The research only revolves around 9-1-1 and does not consider other aspects of first response
6	Dawkins et al. [7]	Evaluated the impact of human factors research in public healthcare	Conducts human factors research on 133 first responders	Addresses new methods of communication	Sample size of the research was relatively small
7	Dawkins et al. [8]	Evaluated the implementation n of technology in various incidents	Conducted research around issues in public safety. technology space	Conducts Multi-phased research and covered a vast array of topics	These technologies must also be developed at price points that the department can afford

(continued)

Table 1 (continued)

S. No.	Study	Contribution	Technique used	Strength	Weakness
8	Greene et al. [6]	Conducted qualitative analysis on issues posed by geography	Conducted in-depth interviews with a diverse group of first responders	Mentions the challenges faced by first responders and suggest solutions	Lack of proper protective equipment and access to technology remain an issue
9	Li et al. [9]	Reviewed factors affecting first responders of a fire emergency	Conducted survey of occupants and first responders	Reports additional factors affecting implementation	Instant actions of the first responders are not properly guided
10	Hathaliya et al. [10]	Evaluated privacy and security issues in Healthcare 4.0	Survey and peer review of security technologies	Mentions pros and cons of various technologies in a structured way	Lack of strong data protection policies pose a threat of data theft
11	Bessen et al. [11]	Evaluated public response around the usage of the	Survey conducted in New Hampshire, USA with first responders	Addresses misconceptions around the usage of the drugs	Sample size of the survey was very small
12	Gupta et al. [12]	Built a smart contract-based telesurgery system 'AaYusH'	Conducts telesurgery using AI-based Tactile Internet	Outperforms the traditional telesurgery system	The system may not work as efficiently with existing technology

3 Proposed Methodology

This study aims to create an emergency bot, which can guide first responders to deal with an emergency situation. This bot is made using a C program which uses concepts like switch case, if-else statements, functions and pointers. The software, when called, will ask the user the kind of emergency response required. The emergency situations have been categorized as: medical, police and fire.

The user shall select from any of the three categories. Next, the software will use a GPS tracking sensor which will send the user's location to the emergency service system. This will be done using NodeMCU ESP8266 board and UBLOX NEO 6 M GPS module. We will configure the setup using Arduino IDE, and then, the model will send our location (longitude and latitude) to the nearest hospital/police station/fire station.

If the user chooses to call the police station or the fire brigade, then the bot will automatically request such service from the nearest available police station/fire station too, sending them the user's location as mentioned above.

If the user chooses to call the police station or the fire brigade, then the bot will automatically request such service from the nearest available police station/fire station too, sending them the user's location as mentioned above.

If the user chooses to call paramedics, they will have to describe the emergency. The options will be (see Fig. 1): heart attack (or its symptoms) vehicular accident other medical emergency like (i) seizures (ii) pregnancy first response (iii) fainting and (iv) other. The user can also use the bot to know the symptoms of heart attack and tips for giving CPR to the patient.

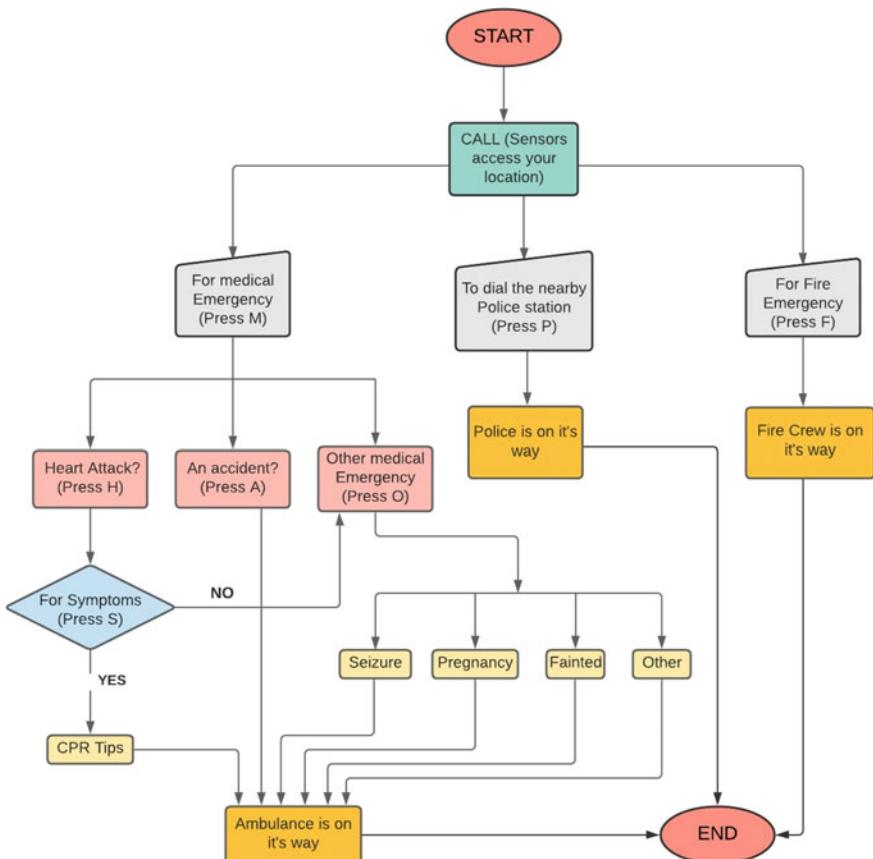


Fig. 1 Proposed model

4 Results and Discussions

Emergency situations demand immediate and quick response. Such a response is often impossible due to remote connectivity of certain areas and lack of appropriate technological know-how. This study proposes the use of an emergency response bot for emergency situations. The bot as in Fig. 2 can be used as a one-stop place to go to incase of any emergency. The bot will take the user's location and send it to the concerned authorities. If the user wants to call for a fire brigade, the following will be the output (Fig. 3).

If the user wants to call for police, the following will be the output as in Fig. 4.

If the user wants to call for a medical emergency, the output is shown in Fig. 5.

The user gets options to choose the kind of emergency. If the user feels the patient is showing heart attack symptoms, the bot can show them the symptoms as in Fig. 6.

If the patient is showing these symptoms, the user can give CPR to the patient until help arrives. The bot will guide the user with CPR steps shown in Fig. 7.

If there was a vehicular accident and the user needs to call an ambulance for those injured in it, the bot will do so while also guiding the user with steps to attend to such a person while help arrives. Other types of emergency situations have also

```
First of all, do not panic,We are here for you
Press M for medical emergencies
Press P to dial the Nearby Police Station
Press F for Fire Emergency
```

Fig. 2 First screen of the bot

```
First of all, do not panic,We are here for you
Press M for medical emergencies
Press P to dial the Nearby Police Station
Press F for Fire Emergency
f
Calling Firemen
Stay Calm!█
sajalsaxena@Sajals-MacBook-Pro Downloads % █
```

Fig. 3 Bot helping the user

```
First of all, do not panic,We are here for you
Press M for medical emergencies
Press P to dial the Nearby Police Station
Press F for Fire Emergency
p
Calling Police
Stay Calm!█
sajalsaxena@Sajals-MacBook-Pro Downloads % █
```

Fig. 4 Calling police via bot

```

First of all, do not panic,We are here for you

Press M for medical emergencies
Press P to dial the Nearby Police Station
Press F for Fire Emergency
m

Calling Ambulance, Stay Calm!
Please Describe the medical emergency :
Press H for Heart Attack (if u want to know heart attack symptoms, press S)
Press A for a Vehicular Accident
Press o for other medical emergency ■

```

Fig. 5 Calling medical emergency via bot

```

First of all, do not panic,We are here for you

Press M for medical emergencies
Press P to dial the Nearby Police Station
Press F for Fire Emergency
m

Calling Ambulance, Stay Calm!
Please Describe the medical emergency :
Press H for Heart Attack (if u want to know heart attack symptoms, press S)
Press A for a Vehicular Accident
Press o for other medical emergency s

->Crushing chest pain
->Arms may feel heavy
->Shortness of breath
->Coughing
Is the patient experiencing any of these symptoms? Press 1 if Yes ■

```

Fig. 6 Bot to diagnose the kind of emergency

```

First of all, do not panic,We are here for you

Press M for medical emergencies
Press P to dial the Nearby Police Station
Press F for Fire Emergency
m

Calling Ambulance, Stay Calm!
Please Describe the medical emergency :
Press H for Heart Attack (if u want to know heart attack symptoms, press S)
Press A for a Vehicular Accident
Press o for other medical emergency s

->Crushing chest pain
->Arms may feel heavy
->Shortness of breath
->Coughing
Is the patient experiencing any of these symptoms? Press 1 if Yes 1

The Person have had a heart attack. Stay calm and wait for ambulance
Meanwhile, you can give the person CPR if he is unconscious.
Press 1 for CPR steps 1
->Place the person on their back and open their airway
->Check for breathing,perform CPR if they stop breathing,
->Perform 30 chest compressions,
    Place one of your hands on top of the other and clasp them together.
    With the heel of the hands and straight elbows,
    push hard and fast in the center of the chest, slightly below the nipples.
->Perform two rescue breaths
    Making sure their mouth is clear, tilt their head back slightly and lift their chin.
    Pinch their nose shut, place your mouth fully
    over theirs, and blow to make their chest rise.
->Repeat the above steps■

```

Fig. 7 Automated healthcare using smart bot

been mentioned and the bot can be called for. These include seizures, pregnancy and fainting/losing consciousness.

5 Conclusion

This study works on the principles of IoT and AI and tries to reap the benefits of modern technology in the field of healthcare. We all understand that digitalization is an ongoing process presently, and its utilization in healthcare is of utmost importance. We designed a model which helps reduce hassle and panic for first responders of medical and other emergencies. We proposed a bot which utilizes a GPS module/sensor which picks up instantaneous real-time location of the first responders and directs help towards them from the nearest source possible, be it paramedic, the fire brigade or the police station. In future, we propose to extend our model will prove to be helpful for the advancement of healthcare and Industry 4.0 utilizing the machine learning techniques [13–16].

References

1. <https://www.siemens-healthineers.com/insights/digitalizing-healthcare>
2. Li N, Yang Z, Ghahramani A, Becerik-Gerber B, Soibelman L (2014) Situational awareness for supporting building fire emergency response: Information needs, information sources, and implementation requirements. *Fire Safety J* 63:17–28. ISSN 0379-7112
3. Nolan R, O'sullivan C, Branson J, Waits C (2005) First responders guide to computer forensics. Carnegie-Mellon Univ Pittsburgh PA Software Engineering INST
4. Greene KK, Dawkins S, Choong Y, Theofanos MF, Prettyman S, Furman S, Steves M (2019) Characterizing first responders' communication technology needs: towards a standardized usability evaluation methodology. In: Mattson P, Marshall J (eds) Homeland security and public safety: research, applications and standards (West Conshohocken, PA: ASTM International, 2019), pp 23–48
5. Mohanta B, Das P, Patnaik S (2019) Healthcare 5.0: a paradigm shift in digital healthcare system using artificial intelligence, IOT and 5G communication. In: 2019 International conference on applied machine learning (ICAML), Bhubaneswar, India, 2019, pp 191–196. <https://doi.org/10.1109/ICAML48257.2019.00044>
6. Greene K, Dawkins ST, Theofanos MF, Steves MP, Furman SM, Choong YY, Prettyman SS (2019) Voices of first responders—examining public safety communication from the rural perspective
7. Dawkins S, Greene K, Steves M, Theofanos M, Choong YY, Furman S, Prettyman SS (2018) Public safety communication user needs: voices of first responders. In: Proceedings of the human factors and ergonomics society annual meeting, vol 62, no 1. Sage CA: Los Angeles, CA: SAGE Publications, pp 92–96
8. Dawkins ST, Greene K, Prettyman SS (2020) Voices of first responders—nationwide public safety communication survey findings: mobile devices, applications, and futuristic technology, Phase 2. vol 2
9. Li N, Yang Z, Ghahramani A, Becerik-Gerber B, Soibelman L (2014) Situational awareness for supporting building fire emergency response: information needs, information sources, and implementation requirements. *Fire Saf J* 63:17–28

10. Hathaliya JJ, Tanwar S (2020) An exhaustive survey on security and privacy issues in healthcare 4.0. *Comput Commun* 153:311–335
11. Bessen S, Metcalf SA, Saunders EC, Moore SK, Meier A, McLeman B, Walsh O, Marsch LA (2019) Barriers to naloxone use and acceptance among opioid users, first responders, and emergency department providers in New Hampshire, USA. *Int J Drug Policy* 74:144–151.k
12. Gupta R, Shukla A, Tanwar S (2020) Aayush: a smart contract-based telesurgery system for healthcare 4.0. In: 2020 IEEE international conference on communications workshops (ICC Workshops). IEEE, pp 1–6
13. Goyal S, Bhatia PK (2020) Feature selection technique for effective software effort estimation using multi-layer perceptrons. In: Proceedings of ICETIT 2019. lecture notes in electrical engineering, vol 605. Springer, Cham. pp 183–194. https://doi.org/10.1007/978-3-030-30577-2_15
14. Goyal S, Bhatia PK (2019) A non-linear technique for effective software effort estimation using multi-layer perceptrons. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), Faridabad, India, 2019, pp. 1–4. <https://doi.org/10.1109/PDGC50313.2020.9315754>
15. Goyal S (2020) Heterogeneous stacked ensemble classifier for software defect prediction. In: 2020 Sixth international conference on parallel, distributed and grid computing (PDGC), Waknaghat, Solan, India, pp 126–130. <https://doi.org/10.1109/PDGC50313.2020.9315754>
16. Goyal S, Bhatia PK, Comparison of machine learning techniques for software quality prediction. *Int J Kowl Syst Sci (IJKSS)* 11(2):21–40. IGI Global. <https://doi.org/10.4018/IJKSS.202040102>

Intelligent Traffic Management System Using Industry 4.0



Aman Panwar, Saagar Bafna, Arjun Raghav, and Somya Goyal

Abstract Traffic is the most drastic effect of this ever-increasing urbanization, the inclusion of cutting-edge technologies in traffic management systems will be effective in reducing traffic-related issues, and thus the commute time. In mid-2020, urbanization stood at around 56%. This rapidly increasing traffic not only affects us directly but has severe implications on our mother nature by unbalancing the biological life cycle and thus has indirect effects on us too. According to a research, urbanites lead more connected lives than their rural counterparts, so a modern traffic management system is the need of this era. Our proposed system uses cutting-edge technologies like Internet of things and has a genuine approach to minimize traffic. This system uses algorithms to manage all kinds of traffic issues prevailing at the time, in a very precise manner. The system will help in overcoming the flaws made by previous traffic management systems. In this system, traffic density is taken as input from the cameras which is then abstracted via digital image processing technique (called as DIP which is a step ahead the analog one) and sensors data, resultantly giving the output as a signal management. Algorithms can be applied for predicting the density of traffic for assisting and planning to subsequently diminish the traffic congestion. The system has a feature to prioritize the emergency vehicles like fire brigade, police vehicles and ambulances by the use of RFID tags in those vehicles. In case of conditions like firebreak out or explosion, sensors specific to fire and smoke gets deployed for detection of such conditions.

Keywords Smart vehicle · IoT · RFID · Traffic management

1 Introduction

In the recent years, there has been a big boom in the automobile sector which has made the traffic conditions complicated and even chaotic. In reference to the urban mobility report (UMR) of Victoria Transport Policy Institute dated December 18, 2014 [1],

A. Panwar · S. Bafna · A. Raghav · S. Goyal (✉)
Manipal University Jaipur, Jaipur 303007, Rajasthan, India

a large chunk of money and time gets wasted due to improper traffic system. UMR predicted that there will be a significant increase in the congestion cost from \$121 billion (in 2011) to \$199 billion (in 2020). Majority of road accidents are due to malfunctioning of traffic management systems and mere negligence of driver. An improvement in the operations of traffic management system will ensure the safety of public vehicles and also improve the efficiency of transportation system. Older traffic control management systems utilize static signaling at point of intersections and does not prioritize ambulances, fire brigades or emergency vehicles, which may lead to human loss, property loss or increased utilization of fuel, causing pollution and congestion. Intelligent Traffic Management System (as in Fig. 1) uses robust communication and processing technologies and complex algorithms for managing traffic effectively during emergencies. Traffic management system not only aims to provide safe and secure transportation services, decrease the number of accidents but also to provide adequate driving skills, maintaining road infrastructure and implement traffic laws.

This is a sensor-oriented traffic protocol where automatization comes in to play. The main objective of this system is to provide accurate data to the management team by developing a digital system to make processes regarding Traffic Management easier. The specific objectives are: 1. To provide direct access to drivers through Web application system, 2. To view the current status of all vehicles using Targo numbers, tax or car information, and 3. To decrease speed of car so as to avoid brutal accidents.

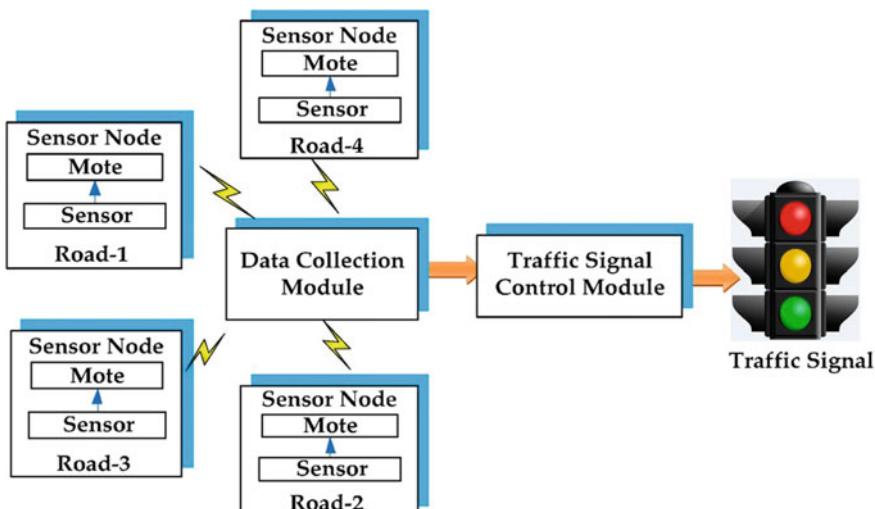


Fig. 1 Sensor-based traffic management

2 Related Works

This section brings the literature review of the state-of-the-art in the field of smart traffic management. The related works are tabulated as Table 1.

Table 1 Table captions should be placed above the tables

S. No.	Study	Contribution
1	Bazzan et al. [2]	Proposed a technique called traffic weighted multi-maps (TWMS)
2	Hamdani et al. [3]	Proposed a system of automation of traffic called autonomous traffic management (ATM)
3	Osman et al. [4]	Proposed a research which includes an intelligent and automated traffic control system which uses conventional computer vision techniques
4	Sumi et al. [5]	Proposed a system which is inspired by the concepts of Internet of things (IoT) and Vehicular ad hoc network (VANET)
5	Salama et al. [6]	Proposed a research in which the integrated intelligent system is used for managing and controlling traffic lights by applying a distributed and long-range photoelectric sensors
6	Bauza and Gozálvez [7]	Proposed cooperative traffic congestion detection (CoTEC),
7	Araujo et al. [8]	Proposed cooperative vehicular traffic congestion identification and minimization (CARTIM)
8	De Souza et al. [9]	A faster and fully distributed TMS is used to increase the efficacy and efficiency of Traffic control systems
9	The Iris/ARTES [10]	Air traffic services (ATS) and aeronautical operational control (AOC) used to complement the aeronautical communications infrastructure
10	Mitsakis [11]	DTA which make trajectory of individual trip starting from origin to destination
11	Han [12]	Dynamic traffic assignment can analyze network performance in extremities like evacuation
12	Allan et al. [13]	Proposed a TMS under VANETs. The first one is V2V, i.e., vehicle-to-vehicle communications
13	Gorey et al. [14]	Improving efficiency of transport systems through development of ITS
14	Sunehra et al. [15]	An automatic street light control system is implemented based on wireless sensor networks (WSN)
15	Sheetal Koul [16]	Automatic detection for speed limit violation This technique can be used to calculate the speed at which a motorist is traveling

3 Proposed Methodology

In our proposed system, the road light glows continuously for a moment and then switches off.

When a vehicle approaches, the lights nearby and a distance ahead will glow for the time being, whereas the lights preceding it dims (Fig. 2).

In beginning, LDR sensor senses the intensity of light in the surrounding environment and consequently transfers the data to Arduino. Arduino will convert the received data into different discrete values, starting from 0 to 1023 (where 0 is the maximum possible darkness, and 1023 depicts the maximum possible brightness) and adjusts the output voltage accordingly 3v/7v (dim/high) depending upon the received value. The output will be 3v in the darkness, i.e., during the night time. Thus, the dim LEDs glows, i.e., with half of max. brightness, and during the daytime, the output would be 0v; thus, the LEDs would be entirely switched off. Initially, when there is no vehicle passing nearby, the IR obstacle detection sensor would be high. IR transmitter continuously transmits the IR radiation. When an obstacle or a vehicle passes from the front of the sensors, emitted rays get blocked by the obstacle and cannot be received by the receiver. Microcontroller senses it as a motion, thus the corresponding LEDs turns from **dim** to **high**. As the obstacle moves forward, blocks the next IR sensor, the corresponding LEDs will turn from **dim** to **high** and the LEDs preceding it will turn from **high** to **dim**. This process continues during the entire course of the vehicle.

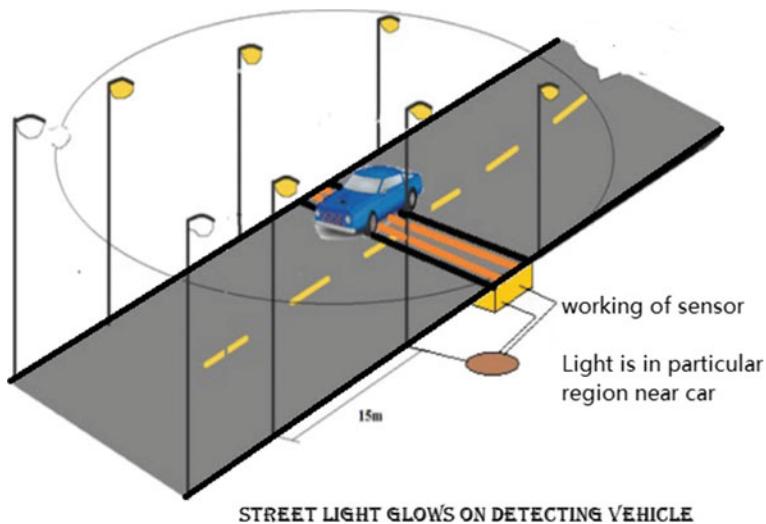


Fig. 2 Demonstration of how street lights nearby to the vehicle glows on detecting

3.1 Communication in Traffic Management System

Two types of communications are there in TMS. One of it is vehicle-to-vehicle type of communications, i.e., V2V. This is used generally when the vehicles need to communicate each other and does not require any other infrastructure. The other is vehicle-to-infrastructure type of communications, i.e., V2I. This is used when a vehicle needs to access content from the Web or send some information or instructions to a central core. Different technologies of communication have been developed to cater to the need of the fellow users (Fig. 3).

This Fig. 4 depicts the overall architecture of a TMS, comprising of vehicles that collects data related to traffic and its issues through OBU. Now, this data is shared to the vehicles nearby using V2I communication. The data is sent to an RSU or a central core like a TMC, i.e., traffic management controller by an access network. Vehicles with RSUs and the sensors available on the road can collect data pertaining to traffic or related issues and send it to the cloud. The central network connects the access

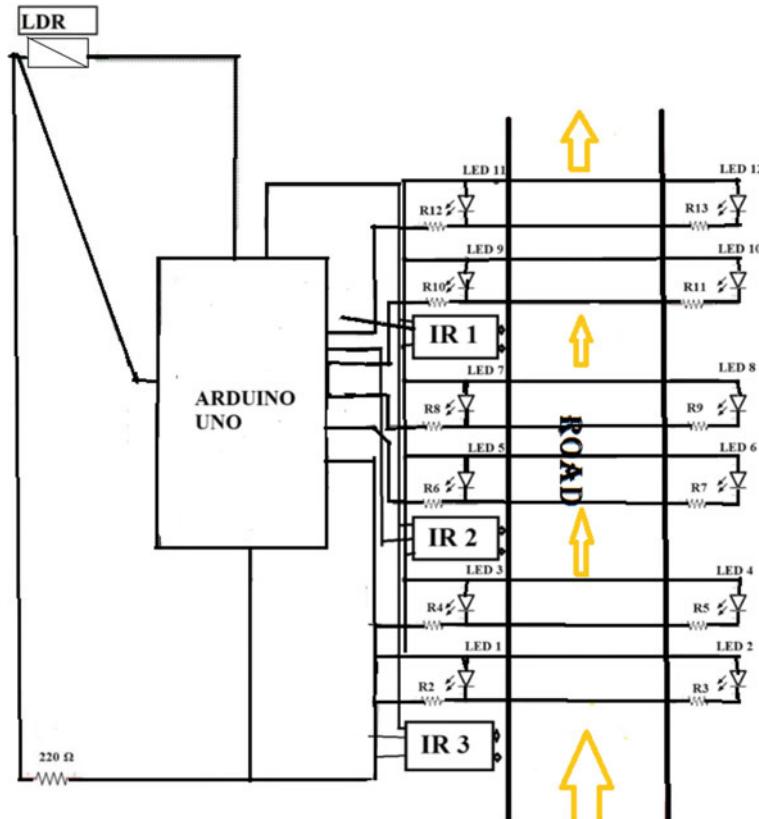


Fig. 3 Working of intelligent street light system using Arduino, LEDs and LDR

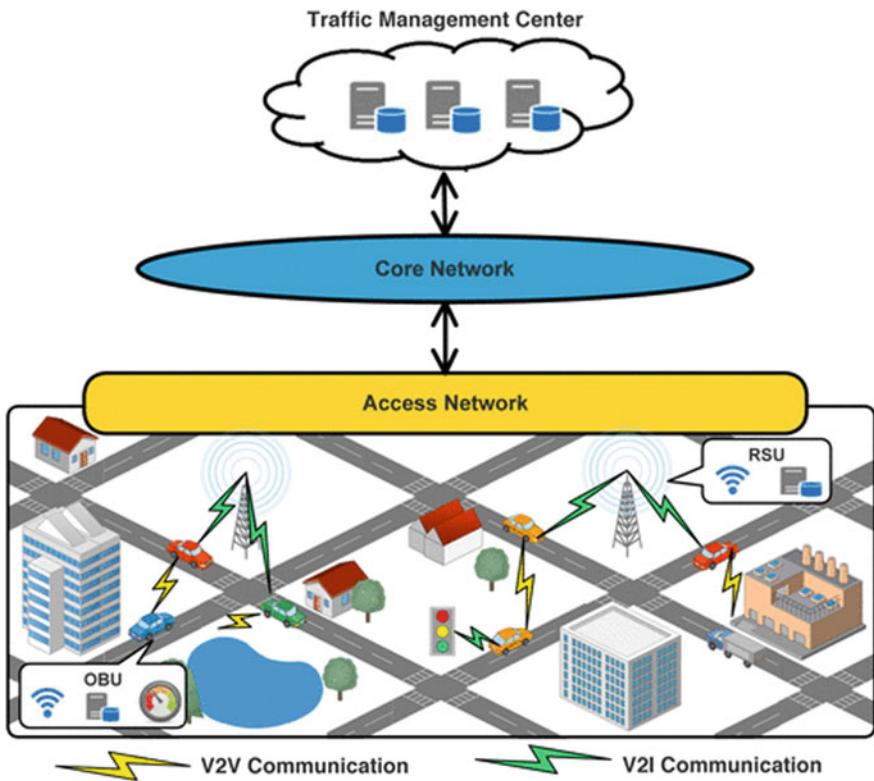


Fig. 4 Overall working of TMS system

network to the cloud, providing a variety of functions, like authentication, switching, aggregation and routing. Also, a variety of sources can be deployed to provide data through this core network and thus improving the efficiency of the services which are delivered by the TMS.

By exploiting these services, TMSs would substantially improve the traveling, reduce traffic, save the commute time and reduce our incident-response time particularly in counter insurgency operations. The TMS comprises of three main phases: (1) information gathering, this is solely responsible for collecting data pertaining to traffic from various random sources; (2) processing of information, its main objective is collecting the data, processing it to further identify any hazard, if there; (3) service delivery, its major task is to provide services to control specific hazards pertaining to traffic and thus improve the overall efficiency of the traffic.

Here, the three main phases of TMS (in Fig. 5) have been described the *information gathering* phase collects data relating to traffic and share the data to the *information processing* phase, where it is exploited for identifying hazards pertaining to traffic. At the end comes the *service delivery phase* which provides counter services depending on the hazards.

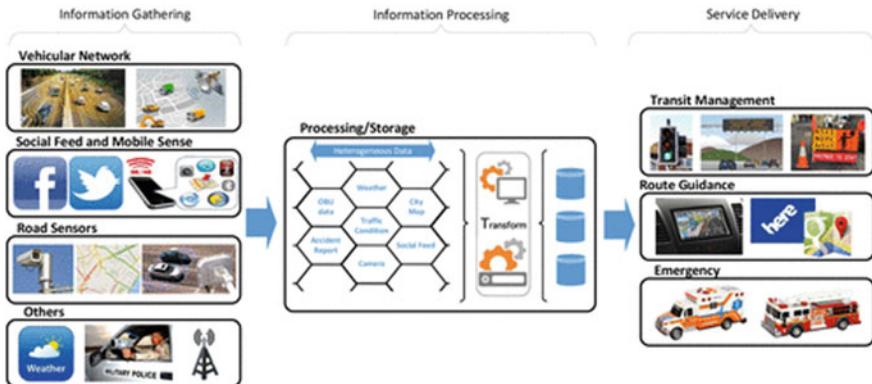


Fig. 5 Three main phases of TMS

4 Results and Discussion

A mechanism can be developed for adjusting speed and enabling a smoother traffic flow and thus can help in reducing the fuel consumption, which in turn will decrease the greenhouse emissions. The main objective is to minimize the traffic accelerations, breakdowns and idle engine for the proper consumption of the fuel. TMS uses data which is shared by the traffic lights so that in case of any accidental event or emergency situations, the vehicles could be made aware of such events so that they have ample time to take appropriate measures.

So, the system focuses on identifying any extremity and simulating the information to other vehicles. For this, we can use a data dissemination protocol.

It can also be used for addressing a broadcast storm and partition of a network. Although, our proposed mechanism cannot be used to provide the ground report and thus cannot give a clear picture of the whole scenario, which may affect other areas as it can lead to congestion in the nearby routes too, but they are worthy enough for controlling congestion caused by traffic accidents effectively.

The proposed mechanism works in interlinking with one another. However, this bulky mechanism can be used to identify the condition of the traffic which may afflict a strain in the network. Though we can detect congestion, but we do not have any mechanism to minimize this congestion.

5 Conclusion

Improving efficiency of the transportation system is still the most strenuous task due to the captious of the transportation infrastructure which are under surveillance by such systems. The article provides a detailed study of the TMSs, focusing on the major challenges and improvising the current existing systems and providing

efficient and eco-friendly ways to improve the efficacy of TMS. The article showcases a detailed structure of state-of-the-art and high-end technology utilized in TMS; the main phases of TMS, i.e., gathering, processing and delivery of the information. At the end, we aim on increasing efficiency of TMS by blending it with state-of-the-art technology [17–20] to provide safe and secure communication for all; furthermore, we can identify and discuss some more issues prevailing and potential efforts to rectify them.

References

1. De Souza AM, Yokoyama RS, da Fonseca NLS (2015) Garuda: a new geographical accident. (CIT/IUCC/DASC/PICOM), Liverpool, 26–28 Oct 2015. IEEE, New York, pp 596–602
2. Bazzan (2014) *Knowl Eng Rev* 29(3):375–403
3. Hamdani S, Benamar N (2018) In: Conference: TAMPAS at Tangier, Morocco on ATM
4. Osman T (2017) In: IEEE 7th annual computing and communication workshop and conference (CCWC)
5. Lucy Sumi, Ranga V (2017) *Int J Eng Trans B Appl Sens Enabled IOT Smart Cities* 2:23–31
6. Salama AS (2010) ICRTMS Conf Comput Technol Dev (ICCTD)
7. Bauza and Goz'alev (2017) *Int J Distrib Sens Netw* 13(4)
8. Araujo (2017) *Int J Distrib Sens Netw*
9. de Souza AM, Celso ARL (2017) *Int J Distrib Sens Netw*
10. The Iris ARTES: Institute for Communication and Navigation, DLR, News Archive ATS, ESA (European Space Agency), 22 January 2010
11. Mitsakis E, Maria Salanova J, Giannopoulos G, Procedia a social and behaviour sciences vol 20, pp 427–436. Elsevier
12. Han S, Fang S, Wang X, Chen X, Cai, *Integr Transp Syst ASCE*, pp 883–95
13. DeSouza AM, Yokama (2019) *J Internet Serv Appl Article number 15*
14. Solodkij S, A. Gorev, OV Popova and DT. Ospanov: formation of priority movement corridors of urban passenger transport, research gate.net
15. Sunehra D, Rajasri S (2017) Automatic street light control system using wireless networks In: 2017 IEEE International conference on power, control and instrumentation engineering (ICPCSI)
16. Lanke N, Koul S, Automatic speed limit violation, smart traffic management system. *Int J Comput Appl* 75 (7):0975–8887
17. Goyal S, Bhatia PK (2020) Feature selection technique for effective software effort estimation using multi-layer perceptrons. In: Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering, vol 605. pp 183–194. Springer, Cham. https://doi.org/10.1007/978-3-030-30577-2_15
18. Goyal S, Bhatia PK (2019) A non-linear technique for effective software effort estimation using multi-layer perceptrons. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), Faridabad, India, 2019, pp 1–4. <https://doi.org/10.1109/COMITCon.2019.8862256>
19. Goyal S (2020) Heterogeneous stacked ensemble classifier for software defect prediction. In: 2020 Sixth international conference on parallel, distributed and grid computing (PDGC), Waknaghat, Solan, India, pp 126–130. <https://doi.org/10.1109/PDGC50313.2020.9315754>
20. Goyal S, Bhatia PK, Comparison of machine learning techniques for software quality prediction. *Int J Knowl Syst Sci (IJKSS)* 11(2):21–40. IGI Global. <https://doi.org/10.4018/IJKSS.2020040102>

ELPR: Energy Efficient Load Balanced PRoPHET Routing for Delay Tolerant Networks



Amit Malik, Amita Rani, and Milan Kumar

Abstract Delay Tolerant Networks (DTNs) are networks that have frequent disconnection, large and variable data delay, asymmetric data rates, high error rates and no end-to-end path availability between the source and destination. Most of the resources of nodes in these types of networks are consumed during network discovery, node scanning response and data transmission. So, designing an efficient routing protocol is quite challenging. In this paper, we propose energy efficient load balanced PRoPHET routing strategy (ELPR) in which message forwarding depends upon the remaining energy and available buffer space on a node. ELPR increases the lifetime of a network by balancing the load of highly encountered nodes, which otherwise dies too soon due to overburden, leaving the network sparse. The simulation results obtained has shown more than 60% decrease in number of dead nodes as compared to the basic PRoPHET protocol, improving the performance of various other network parameters such as delivery probability, number of message drop and remaining energy.

Keywords Delay tolerant networks · PRoPHET · DTN routing · Energy efficient routing · ELPR

1 Introduction

DTNs [1] are a network of regional networks that have characteristics like inconsistent delays, patchy connectivity, uneven data rates and excessive error rates. Due to these characteristics DTNs are designed to operate in applications where complete

A. Malik (✉) · A. Rani · M. Kumar

Department of Computer Science and Engineering, DCRUST, Murthal, Sonipat, India
e-mail: amit.schcse@dcrustm.org

A. Rani
e-mail: amitamalik.cse@dcrustm.org

M. Kumar
e-mail: milan.kumar@india.nec.com

path cannot be guaranteed always. A DTN node stores the messages, carries them till the connectivity is established with next node, and then forwards with the help of Bundle Protocol [2], which is different approach from usual store-and-forward strategy used by prominent networks such as TCP/IP. In DTNs a node has to wait for the contact opportunities with intermediate node or the destination node to transfer the data. During this time period node keeps the data bundles into its buffer and carries it. But, when the nodes participating in the network do not have enough energy or enough buffer space to hold the messages, the network reliability may decrease or the network may also fail.

In DTNs most of the nodes are mobile in nature, therefore, it is very important for the mobile nodes to use their resources efficiently. Since the connections are intermittent in DTNs, a lot of energy is consumed in discovering neighbor nodes so as to forward data packets. Saving node's energy is very important for a longer life time of the network. Also, buffer space should be used efficiently to increase the network productivity because the chances of a message getting dropped are higher when the data is sent to the nodes with less buffer space. In this case, the nodes either reject the new inbound message or drop the older message to make space for new message.

In DTN like networks, it is difficult to predict the successful delivery of data packets, since the direct delivery of messages from source to destination is very rare. As we discuss in the next section, various strategies have been proposed to increase the packet delivery in DTNs. PRoPHET routing [3], which uses the delivery predictability calculation method to select the intermediate node, is one such well-known routing protocols for DTNs. Delivery predictability indicates the chances of the node to meet the other nodes in the near future. The calculation of delivery predictability is done on the basis of the history of encounters between nodes. However, the major limitation of PRoPHET [3] is that only those nodes which have high predictability values are repeatedly selected. It leads to exhaustion of their resources due to uneven load distribution that further results in poor network performance or node failures. In this paper, we propose an energy-aware PRoPHET routing protocol for DTNs, which balances the load in the network and takes data forwarding decision on the basis of remaining energy and buffer space of neighboring nodes.

In the following section, we discuss some of the existing routing protocols for DTN. Section III discusses the motivation behind the proposed work. In section IV, detailed description of the proposed Energy Efficient Load Balanced Probability PRoPHET routing Protocol using History of Encounters and Transitivity (ELPR) has been given. Section V provides simulation results and analysis and section VI concludes the paper.

2 Related Work

DTN routing protocols can be classified into three major categories [4] which include: (1) Naive Replication, (2) Utility forwarding and 3) Hybrid Approach. In naive

replication, multiple copies of same data bundles are generated without any prior knowledge about contactable nodes. The main motive of this scheme is to increase delivery probability. Protocols such as Epidemic [5], Spray and Wait [6] and Spray and Focus [7] belong to this category. In utility forwarding, the forwarding decision is taken based on utility value of nodes. The utility value is calculated on the basis of parameters such as, buffer capacity, delivery probability, etc. No replication of messages is made in this scheme, i.e., only single copy is forwarded. Protocols such as First Contact [8] and Seek and Focus [9] fall in this category. Hybrid approach, on the other hand, combines the features of both naive replication and utility forwarding approaches. The main purpose of this approach is to make use of the advantages of both previously described routing approaches. Protocols such as PRoPHET [3], RAPID [10] and Bubble Rap [11] can be considered under this category.

Apart from these conventional DTN routing protocols, many energy-aware routing schemes have also been proposed by different authors. In [12], the authors proposed n-epidemic routing protocol. It is an energy-aware routing protocol in which message is forwarded by a node to their neighbors only when they are above a certain threshold value in terms of number of neighbors available. By imposing this threshold value, n-epidemic routing scheme saves energy and delivery probability is also increased.

Gao et al. in [13] proposed a strategy that betters the routing process of Spray-n-Wait protocol with reduced power consumption. The criterion used for calculations includes power left and pace of the nodes and on the basis of that it is decided that how many replicas of messages will be sent to the intermediate nodes.

In [14], authors proposed an energy-aware scheme by modifying the spray phase of Spray and Wait protocol. The vibrancy values are shared when two nodes come in contact with each other. Depending upon the vibrancy and the residual energy at source node, the decision of the number of copies that should be sent to neighbor node is taken.

An energy-aware routing scheme proposed in [15] selects an intermediate node on the basis of utility function. This utility function determines whether the intermediate node has enough resources to receive and handle the message. This way, a node having low energy is prevented to deliver the message.

Yao et al. introduced an energy-aware history dependent prediction for infrastructure less DTNs [16]. Utility function is calculated which is used to tell the next best node to share the message. The factors which are derived to calculate the utility function are (1) the perpendicular factor—the distance between the sender and the next hop is calculated because nodes with small distances consumes less energy, (2) the transmission factor—is used to calculate the number of copies. These copies are forwarded by the sender to the next hop. (3) A sparse constant factor—by which generation of message is controlled.

The In [17], authors proposed energy-aware scheme for sparse Opportunistic Networks [18]. In these types of networks, large numbers of isolated nodes are present. The nodes waste their energy even if no nodes are in their contact to forward or receive the message. For such networks authors proposed energy-aware scheme which is based on asynchronous sleep approach. In this scheme, the nodes which are in idle state are registered into ‘low power consumption’ state. These nodes awaken

at appropriate intervals to see whether other nodes are in their communication range or not.

In the next section, we propose the ELPR approach for PRoPHET [3] routing protocol to achieve longevity of the nodes in terms of life through load balancing, which ultimately leads to better network performance on many parameters.

3 Motivation and Design Parameters

We begin, in this section, by explaining the approach of basic PRoPHET routing algorithm and the motivation for improvement in the basic approach, before describing the design parameters.

PRoPHET—A greedy approach. Basically PRoPHET [3] routing is a greedy approach that uses the concept of delivery predictabilities (DPs). The DPs are the chances of a node to meet with other node in the network and are calculated on the basis of history of encounters of a node with other network nodes. It also considers the transitivity time and the DP value decreases gradually as the time of last encounter between two nodes increases. Hence, the value of DP varies. Every node in the network maintains such DPs in a data structure.

Design Motivation. It is due to the greedy behavior of PRoPHET routing protocol that the intermediate nodes that have higher DPs are considered for data forwarding, which in turn takes these nodes toward dead state very soon, leading to very high packet drop rate. Also, network nodes consume energy in performing various tasks such as network scanning, message routing, task scheduling. These tasks must be performed in an orderly and efficient manner for better utilization of the available energy. Tasks which are unnecessary, or may not be able to fulfill the purpose of their occurrence must be avoided. Here, we want to prevent those routing tasks from occurring which may not achieve the goal of reaching the destination successfully. The event, in which a message which is generated by a source node and successfully received by the destination node through ‘ i ’ number of intermediate nodes, is termed as ‘successful delivery’. It is termed as ‘unsuccessful delivery’ if the message fails to ultimately reach the destination but has been forwarded by the intermediate nodes.

In this work, we propose that the network load should be properly balanced and, therefore, a node must avoid forwarding the messages to the nodes which have lesser resources available and are highly loaded. The message forwarded to such nodes may not lead to its successful delivery to the destination. We consider the following parameters to formulate the design of our strategy.

Remaining Energy. Remaining energy is the amount of energy available from the power source attached to the node, at a time t , when it comes in contact with another node. We have considered the fact that a node must not choose to forward a message to the node with lesser energy, as it may not lead to successful delivery, thus, wasting the resources.

Available Buffer Space. Available buffer space is the total amount of space which is un-occupied in the buffer of a node, at time ‘t’, when it comes in contact with some other node. Each node stores its own messages in the buffer, as well as, the messages it received from other network node. If the buffer is full to its maximum capacity, then either the old messages are dropped from the buffer or the new messages are rejected. We propose to avoid the message forwarding to a node which may have lesser buffer space available and may not be able to carry the message for successful delivery.

Next, we present the Energy Efficient Load Balanced Probability Routing Protocol using History of Encounters and Transitivity (ELPR) algorithm, which combines the above two parameters of remaining energy and available buffer space for load balancing.

4 Energy Efficient Load Balanced Probability Routing Protocol Using History of Encounters and Transitivity (ELPR)

This section discusses the proposed ELPR protocol which is based on computing the residual energy and available buffer space of a node in order to make routing decisions. The protocol tries to balance the traffic load among available set of intermediate nodes. The protocol employs the following steps:

Probable Relay Node Selection. Energy and space utilization are two most important aspects which are considered in designing any efficient routing technique. Here, we combine these two aspects and check the remaining energy and available buffer space, when two nodes comes in contact, to decide whether to continue further with message forwarding or discontinue it at this point. If the resources of the receiving node are lesser, then the forwarding is discontinued, as it may lead to unsuccessful delivery. This saves the node’s energy which, ultimately, improves the network performance, as elaborated in the subsequent sections.

Initially, when a node wants to send a message, it first waits for communication opportunities. When some intermediate node ‘I’ come into contact with the sender node, it first calculate its ratio EL_I of percentage of available buffer space to percentage of remaining energy using (1),

$$EL_I = ((I_{OB} \div I_{TB}) \times 100) / (I_{RE} \div I_{TE}) \times 100 \quad (1)$$

Here, I_{OB} is the occupied buffer space, I_{TB} is the total buffer space, I_{RE} is the remaining energy and I_{TE} is the total energy of node ‘I’. This value is shared by the node with the sender node. The sender node ‘S’ also calculates its own ratio EL_S of percentage of available buffer space to percentage of remaining energy using (2),

$$EL_S = ((S_{OB} \div S_{TB}) \times 100) / (S_{RE} \div S_{TE}) \times 100 \quad (2)$$

Now, the sender node compares its own ratio with the value shared by neighbor nodes using (3) as,

$$\text{EL}_S \leq \text{EL}_I \quad (3)$$

All the neighbor nodes which satisfy (3) are stored in a Table T1 and are the probable relay nodes.

Delivery Predictability and Message Forwarding. Now, the DP for each probable node stored in Table T1 is updated using (4) as,

$$DP_{(X,Y)} = DP_{(X,Y)old} + (1 - DP_{(X,Y)old}) \times DP_0 \quad (4)$$

Here, $DP_{(X,Y)}$ is the delivery predictability of the two encountering nodes ‘X’ and ‘Y’, $DP_{(X,Y)old}$ is the previous value of their delivery predictability and $DP_0 \in [0,1]$ is the delivery predictability value assigned to each node initially. Moreover, as the time passes, the value of DP decreases and this change is calculated using (5) in which, ‘ α ’ is the aging constant having value between [0,1],

$$DP_{(X,Y)} = DP_{(X,Y)old} + \alpha^t \quad (5)$$

The DP of a node may also be updated using transitivity rule, given by (6), according to which if node ‘X’ encounters node ‘Y’ frequently and node ‘Y’ encounters another node ‘W’ frequently, then, any node ‘L’ which encounters node ‘X’ may forward the messages to ‘X’ which has node ‘W’ as the destination node, because the chances of a message reaching ‘W’ via ‘X’ and ‘Y’ faster are greater as compared to ‘L’ encountering ‘W’ or ‘Y’

$$DP_{(X,W)} = DP_{(X,W)old} + (1 - DP_{(X,W)old}) DP_{(X,Y)} \times DP_{(Y,Z)} \times \beta \quad (6)$$

Here, ‘ β ’ is the scaling constant for transitivity having value between [0, 1]. The value of DP_0 , β and α is taken as 0.75, 0.25 and 0.998, as suggested by [3].

After updating, the DP of nodes stored in T1 is compared with the DP of sender node one by one. The nodes for which the DP is greater than the DP of sender are stored in Table T2 using (7).

$$T2 = DP(\text{neighbour}) > DP(\text{sender}) \quad (7)$$

Finally the message is transferred to all the neighbor nodes which are in Table T2.

ELPR Protocol Algorithm

In ELPR protocol works when neighbor nodes come in contact with sender node, they information about their remaining battery level and buffer space. Sender then calculates the percentage of remaining energy and percentage of remaining buffer

space of all the neighbor nodes with respect to itself. After that it compares remaining energy level and available buffer space by using (1). The neighbor nodes that satisfy (1) are then stored in a Table T1 held by sender. After that node sender compare its DP with all the nodes which are in Table T1. If the delivery predictability of neighbors' is greater, then it sends the data to the neighbor nodes. The procedure is followed until the message is reached to the destination or the time-to-live of the message has been expired. The pseudo-code for above procedure is given below:

```

1.  START
2.  if (sender_buffer is not empty)
3.      for (all encountering node)
4.          if (node_encountered == destination)
5.              if(No message copy at destination)
6.                  transfer messages;
7.              endif;
8.          end if;
9.          calculate ELI using (1);
10.         calculate ELS using (2);
11.         if (ELI > ELS)
12.             add node_encountered to table T1;
13.         end if;
14.         go to step 3;
15.     for (all nodes in T1)
16.         if (DPsender < DPneighbour)
17.             forward message copy;
18.         end if;
19.     go to step 2;
20. END
```

In next section we explain the simulation parameters and result analysis.

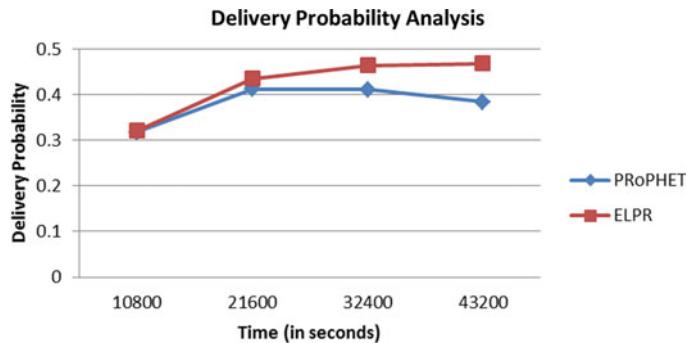
5 Simulation Results and Analysis

We use ‘The ONE’ simulator [19] to evaluate and compare the results of PRoPHET routing protocol with ELPR. MessageStatsReport and EnergyLevelReport available in the simulator provide the results related to the delivery probability, overhead ratio, energy and buffer usage. Values of various parameters used during the simulation are shown in Table 1.

Most of the parameters are set to their default values as suggested in [19]. However, the value for initial energy and buffer capacity varies, depending upon the type of node, which replicates the real world scenario. Next, we discuss the performance comparison of ELPR with PRoPHET for various network parameters.

Table 1 Parameters used for simulation

Parameter	Value	Parameter	Value
Simulation Area	4500*3400 m	Message Size	500 Kb – 1 Mb
Number of node groups	6	Message Generation Interval	25-35 s
Total numbers of nodes	250	Transmission Energy	0.08 J
Communication Interface	Bluetooth	Scan Energy	0.06 J
Transmission Range	10 m	Scan Response Energy	0.08 J
Speed of Transmission	2Mbps	Base Energy	0.08 J
Buffer Capacity	10 MB, 15 MB	Node's Movement Model	MapBased

**Fig. 1** Delivery Probability vs. Time when TTL is 150 min

Delivery Probability. The delivery probability of the proposed ELPR and existing PRoPHET routing is examined when the time-to-live of the message is 150 min. The results are analyzed from 10,800 s to 43,200 s (12 h). As in Fig. 1, for both of the protocols the delivery probability increased initially and then decreased because as the time passes, most of the nodes are dead and not able to take part in the data transmission. But still, ELPR gives higher delivery probability because message is forwarded only to nodes that have enough buffer space and energy. At the end of the simulation (after 12 h) the delivery probability by PROPHET is 0.38 and by proposed ELPR is 0.48 approx.

Remaining Energy. The average remaining energy is calculated at the end of the simulation (after 12 h). The average remaining energy when we use ELPR is 78.55 J while for PRoPHET it is only 3.94 J. The results are better in case of ELPR because the routing decision is taken by considering the energy of nodes. Nodes having less energy have no need to spend their energy in term of node scanning and response to other nodes. The approx. increase in the lifetime of the nodes is observed to be 21.5 min in 12 h. Figure 2 shows the comparison of average remaining energy between the ELPR and PRoPHET.

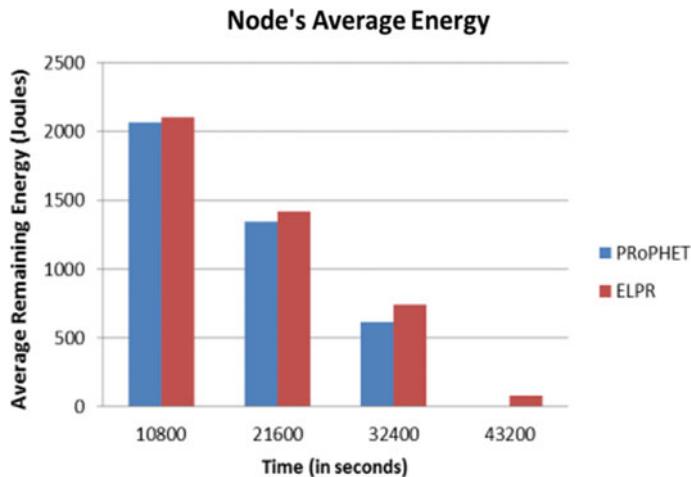


Fig. 2 Average Remaining Energy vs. Time when TTL is 150 min

Number of Dead Nodes. We also investigate the number of dead nodes at different time intervals for both ELPR and PRoPHET routing protocol. At 40,800 s of simulation ELPR has no dead nodes while PRoPHET has 82. At the end of simulation (after 12 h) PRoPHET has 233 numbers of dead nodes while ELPR has only 80 nodes out of 250 nodes. Figure 3 shows the comparison between the numbers of dead nodes.

Message dropped percentage. As simulation time increases, percentage of packet dropped increased for both PRoPHET and ELPR protocol. At the end of the simulation, the percentage of messages dropped by ELPR is 40 and by PRoPHET it is

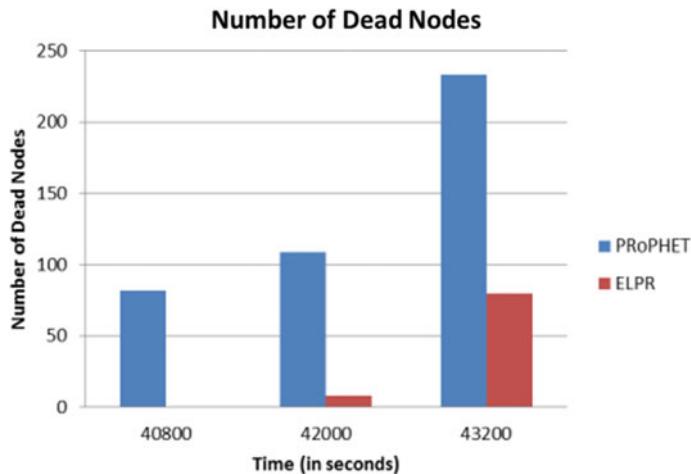


Fig. 3 Numbers of Dead Nodes versus Time when TTL is 150 min

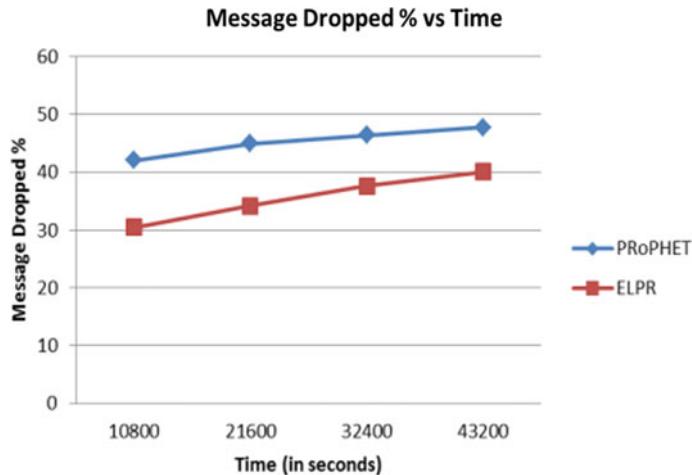


Fig. 4 Message Dropped % vs. Time when TTL is 150 min

48. By using ELPR message dropped percentage is reduced because messages are transferred to the nodes which have enough buffer space. The chances of holding message into buffer is increased when the nodes have enough buffer space as it has no need to drop the older message in order to hold the new one. Figure 4 shows the comparison of message dropped percentage.

6 Conclusion

This paper proposed an efficient PRoPHET-based routing protocol for DTNs named as ELPR. In ELPR the messages are pushed further, in effort to reach destination node, depending upon the node's remaining power and remaining buffer space. Simulator-based results show that ELPR performs better than PRoPHET in field of delivery probability, energy consumption, dead nodes at the end, and percentage of message lost. Subsequently, efforts can be made in this field by using real mobility traces. Practical experimentation need to be done in this area to make the network more realistic.

References

1. Fall, K., & Farrell, S. (2008). DTN: an architectural retrospective. *IEEE Journal on Selected areas in communications*, 26(5).
2. Scott, K. L., & Burleigh, S. (2007). Bundle protocol specification.

3. Lindgren A, Doria A, Schelén O (2003) Probabilistic routing in intermittently connected networks. ACM SIGMOBILE mobile computing and communications review 7(3):19–20
4. Gonçalves Filho, J., Patel, A., Batista, B. L. A., & Celestino, J. (2016). A systematic technical survey of DTN and VDTN routing protocols. Computer Standards & Interfaces, 48, 139-159
5. Vahdat, A., & Becker, D. (2000). Epidemic routing for partially connected ad hoc networks.
6. Spyropoulos, T., Psounis, K., & Raghavendra, C. S. (2005, August). Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking (pp. 252–259). ACM.
7. Spyropoulos, T., Psounis, K., & Raghavendra, C. S. (2007, March). Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility. In Pervasive Computing and Communications Workshops, 2007. PerCom Workshops' 07. Fifth Annual IEEE International Conference on (pp. 79–85). IEEE.
8. Jain, S., Fall, K., & Patra, R. (2004). Routing in a delay tolerant network (Vol. 34, No. 4, pp. 145–158). ACM.
9. Spyropoulos, T., Psounis, K., & Raghavendra, C. S. (2004, October). Single-copy routing in intermittently connected mobile networks. In Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on (pp. 235–244). IEEE.
10. Balasubramanian, A., Levine, B., & Venkataramani, A. (2007, August). DTN routing as a resource allocation problem. In ACM SIGCOMM Computer Communication Review (Vol. 37, No. 4, pp. 373–384). ACM.
11. Hui P, Crowcroft J, Yoneki E (2011) Bubble rap: Social-based forwarding in delay-tolerant networks. IEEE Trans Mob Comput 10(11):1576–1589
12. Lu, X., & Hui, P. (2010, July). An energy-efficient n-epidemic routing protocol for delay tolerant networks. In Networking, Architecture and Storage (NAS), 2010 IEEE Fifth International Conference on (pp. 341–347). IEEE.
13. Gao, S., Zhang, L., & Zhang, H. (2010, October). Energy-aware spray and wait routing in mobile opportunistic sensor networks. In Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on (pp. 1058–1063). IEEE.
14. Patel VG, Oza TK, Gohil DM (2013) Vibrant energy aware spray and wait routing in delay tolerant network. Journal of Telematics and Informatics 1(1):43–47
15. Wang, G., Tao, J., Zhang, H., & Pan, D. (2017). A improved Prophet routing based on congestion level of nodes in DTN. 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 1666–1669.
16. Hui, M., & Tao, L. (2018). Application of Prophet Routing Based on Throughput in DTN [J]. Computer Technology and Science, 28(255 (07)), 193–197.
17. Mao Y, Zhou C, Ling Y, Lloret J (2019) An optimized probabilistic delay tolerant network (DTN) routing protocol based on scheduling mechanism for internet of things (IoT). Sensors 19(2):243
18. Henmi, K., & Koyama, A. (2020). Hybrid type DTN routing protocol considering storage capacity. International Conference on Emerging Internetworking, Data & Web Technologies, 491–502.
19. Madni MAA, Iranmanesh S, Raad R (2020) DTN and Non-DTN routing protocols for inter-cubesat communications: A comprehensive survey. Electronics 9(3):482
20. Keränen, A., Ott, J., & Kärkkäinen, T. (2009, March). The ONE simulator for DTN protocol evaluation. In Proceedings of the 2nd international conference on simulation tools and techniques (p. 55). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

LSTM-Based IDS System for Security of IoT



Ajay Kumar and Amita Rani

Abstract Deep LSTM IDS is a powerful tool to secure the network. It makes use of different deep learning techniques to train the network. The proposed work deals with various phases such as dataset selection, preprocessing, training, and testing. The current IDS security has made it possible to counter security threats in the IoT environment. Development in IDS presents an opportunity to achieve significant improvements in this security domain. The LSTM layers have been used to train IDS, and the dropout layers are used for overfitting. The IDS is very effective and self-learning in nature. The critical analysis of IDS performance is done concerning different parameters such as accuracy, precision, recall, and F1-score using via confusion matrix. The effectiveness of IDS is found to be 96.41% for a total of 9 attacks. Dataset taken into account is IOTID20.

Keywords IDS · IoT · LSTM · RNN · DNN

A. Kumar (✉) · A. Rani

Deenbandhu Chhotu Ram University of Science and Technology, 50th K.M. Stone, N.H. 1,
Murthal 131039, Haryana, India

A. Rani
e-mail: Amitamalik.cse@dcrustm.org

1 Introduction

1.1 Internet of Things (IoT)

The Internet of Things [1] is for devices that focus on communication and computation capabilities. IoT is a connection of physical objects. These objects may be vehicles, buildings, and electronics. It is a network that allows devices to interact. The network plays a significant role in storing and exchanging information. In other words, IoT is a system of internet-connected and connected objects. These objects are capable of storing and transmitting data on a wireless network and over to the Internet [2]. IoT is developed on top of services that permit coordination with different. Such developments in IoT have ushered in more machine-to-machine transmission with the support of the Internet without any requirement of manual intervention [3–5]. With fast-going development, the IoT mechanisms are utilized in a broad area. IoT is becoming an ideal target for cyber attacks due to its distributed nature. Its open nature is also a threat to its security. Attackers target IoT nodes as they are storing sensitive data, and that is why there is a need to curb such security risks [6].

1.2 Intrusion Detection System (IDS)

An intrusion detection system [7] is the software capable of viewing the network in case of any malicious operation and violation of given policies. Detection of intrusion plays a significant part in ensuring the data protection of information. The primary mechanism used is capable of checking different types of attack over the network in an accurate manner. A network-dependent system has been found to check network-connected operation such as volume of traffic, remote IP addresses, ports used, and utilized protocols [8].

An IDS is an application that is monitoring network traffic to trace suspicious activity after that issues an alert if any such activity is detected. It checks the system for dangerous operation as well as any breach of policy. The intrusion detection system is composed of various components [9–13].

Intrusion detection systems [14] operate by considering signs of known attacks/deviations during usual operations. Such deviations/anomalies pushed up the stack and checked at the protocol and application layer.

1.3 Deep Neural Network (DNN)

A DNN is a network of artificial neurons that functions like the human brain, which has several layers such as input, hidden, and output layers [15]. DNN detects

valid mathematical manipulation to produce output as per input. There is a linear relationship as well as a non-linear relationship between inputs and outputs.

Structured deep learning can be utilized for hierarchical learning for IDS [16] detection. DNN is dependent on artificial neural networks. Research has utilized deep thinking and quick learning for viable AI.

1.4 Recurrent Neural Network (RNN)

RNN is a type of neural networks where interconnectivity among nodes forms a directed graph and creates a temporal sequence allowing it to show temporal dynamic behavior [17]. RNNs utilize their memory to process variable-length sequences of various inputs derived from feedforward neural networks. It makes them implement operations like unsegmented as well as interconnected consideration [18].

RNN has been utilized to take into account two broad categories of networks supporting the usual structure. Here, one is having a finite impulse, and another one is an infinite impulse. Such categories of networks have exhibited runtime actions that are not permanent [19].

1.5 Long Short-term Memory (LSTM)

LSTM is an RNN that is frequently utilized in the area of deep learning. LSTM has feedback interconnection that has its memory [20]. It can use complete sequences of information like audio and video without the use of a matrix. The LSTM [21] networks are considered suitable for performing classification as it performs processing and makes predictions based on time series information. There might be lags of duration that are not known in significant events during time series. LSTM is found capable of getting taught order dependence in case of the sequence prediction problem. LSTM can be used in case of complicated issue domains like translation by machine and NPL [22].

2 Proposed Work

The proposed LSTM model has performed training on seventy percent of the dataset and testing on thirty percent of the dataset.

2.1 Process Flow of Deep LSTM IDS

Deep LSTM IDS is a powerful tool to secure the network. It makes use of different deep learning techniques to train our network. There are various steps with different techniques involved in making an IDS. The proposed work deals with various phases.

- **Dataset selection**
- **Pre-processing**
- **Training**
- **Testing.**

2.1.1 Dataset

The dataset that has been used in research is IoTID20 [22]. Table 1 is presenting three cases where binary classification is made considering normal and anomaly. The category section considers five categories normal, DoS, Mirai, MITM, and scan, and subcategory presents different categories derived from normal categories in Table 2.

Table 1 Classification of IDS dataset

Binary	Category	Subcategory
Normal, anomaly	Normal DoS, Mirai, MITM, Scan	Normal, Synflooding, Brute force, http flooding, UDP flooding, ARP spoofing, Host port, OS

Table 2 Subcategories of dataset

Subcategory distribution	
type	Instances
Normal	40,073
DoS	59,391
MiraiAck flooding	55,124
Mirai brute force	121,181
Mirai http flooding	55,818
Mirai UDP flooding	183,554
MITM	35,377
Scan host port	22,192
Scan port OS	53,073

Table 3 Removed features

Total features	Feature name
12	Active_Max, Bwd_IAT_Max, Bwd_Seg_Size_Avg, Fwd_IAT_Max, Fwd_Seg_Size_Avg, Idle_Max, PSH_Flag_Cnt, Pkt_Size_Avg, Subflow_Bwd_Byts, Subflow_Bwd_Pkts, Subflow_Fwd_Byts, Subflow_Fwd_Pkts

2.1.2 Pre-processing

In the previous phase, after selecting the dataset, there is a need to clean the data, also known as preprocessing. The preprocessing of dataset is done before training. During preprocessing, useless attributes that are not playing any significant role in decision-making are eliminated, and it also takes care of the missing value. The Shapiro–Wilk algorithm is utilized to eliminate such attributes considering their significance. Data preprocessing is a process that is performed on raw data before training it. This algorithm is utilized to eliminate attributes. The Shapiro–Wilk algorithm checks the uniformity of distribution of the occurrences to characteristics and removes characteristics with less than a 0.5 ranking score. Thus, it helps in eliminating attributes utilized in the dataset based on their ranking score.

We have removed a total of 12 features from the dataset that has less than a 0.5 feature ranking value shown in Table 3. More than 70% of features have more than 0.5 value, and they are included in training and testing.

The different anomaly are classified and supposed to be predicted by the trained model. The anomaly may be DoS, Mirai, MITM, and scan.

2.1.3 Neural Network-based IDS

It deals with the setup of the neural network. Neural network models are performing supervised learning operation. It builds knowledge from IDS datasets. It considers the results that are previously known for testing purposes. Networks are then learning to tune themselves to detect the correct result themselves. These neural networks can be utilized to solve several issues. These issues might be interconnected to network security like intrusion detection, sales forecasting, and customer research. Other issues are the validation of data along with the management of risk.

LSTM and Its Training Mechanism

The trained network ‘net’ is stored in the system for further testing. The LSTM is implemented with two LSTM layers to produce a trained network. The proposed model uses two LSTM layer along with a dropout layer during a training operation. To perform training, 70% dataset is considered for training and the remaining 30% to perform the test. Based on feature, LSTM-dependent neural is trained. The factors that are influencing the time of training are batch size. The hidden layers and dropout layers are playing a significant role in increasing accuracy. After getting a dataset of

IDS, the selection of characteristics is performed to train the dataset. The training and testing ratio is set, then LSTM1 layer with 12 hidden layers and LSTM2 layer with five hidden layers. Dropout layers are used to resolve the issue of overfitting, then a fully connected layer for bias vector and softmax layer for activation are applied. Classification operation is performed to perform decision-making to predict intrusion.

LSTM layer

In the research, the LSTM mechanism is used to train the network, which uses deep learning and uses feedback interconnection. In the proposed model, LSTM networks are used to perform the IDS attack classification with the support of hidden layers, dropout layers, fully connected, and classification layers. LSTM mechanism is used to perform processing and make predictions based on the given IDS dataset.

The trained model classifies different anomalies. Two LSTM are used to model the IDS; they have been joined consecutively. Each has different hidden layers, i.e., 12 and 5, respectively. These hidden layers have increased the accuracy but cause the issue of overfitting. Overfitting arises during the neural network model training. If the training is continued, then the model will adopt idiosyncrasies.

These layers have to learn long-term dependencies. It learns from different time steps lying in time series along with sequence contents. It performs additive interactions. The layers are supposed to enhance gradient flow in long sequences during training.

Dropout layer

Overfitting can be traced during plotting by checking the validation loss. With the continued training, then the model becomes idiosyncrasies. After sometime, training becomes less suitable for data that is new to it. This data can be of different samples from the population. The model is overfitting when it is too well-adapted to training as well as validation data. The dropout layer is one of the ways to resolve the issue of overfitting. The dropout layer can handle such a problem. The model is overfitting when training loss is constant or it is decreasing. Techniques known as regularizes are used to minimize the influence of overfitting. Dropout is considered one of them.

Fully connected layer

A fully connected layer is multiplying the input using a weight matrix. It gives a biased vector afterward. In the proposed work, fully connected layer (output size) function is to return a fully connected layer and specify the output size property.

Softmax

The activation layer is applied to the last layer in the case of a neural net. It is utilized in place of ReLU, sigmoid, tanh activation functions. The softmax layer is needed because it converts the last layer's output to the neural network. Softmax is usually executed with the support of the neural network layer just before the output layer. This layer should have the exact counting of nodes that are in the output layer.

Classification layer

The classification has been considered the most active research and application area of neural networks. Classification is a significant characteristic to detach massive datasets in classes to generate a rule.

Gradient threshold

It is described as a comma-separated pair. It consists of gradient threshold as well as a positive scalar. If the gradient is more than the value of gradient threshold, then the gradient is decreased as per the gradient threshold method. The proposed work has utilized two as a gradient threshold value.

Learning Rate

This is termed as a configurable hyperparameter. It is controlling the intensity of modification of the model considering estimated error. Every time, the weights of the model are modified. The learning rate is utilized in the training of neural networks where small positive values are present. These positive values lie from 0 to 1. The learning rate is controlling the adaption issues of the model. The learning rate in the case of the proposed work is 0.001.

2.1.4 Testing Phase

The testing phase is the phase where the accuracy of the trained model is checked. The sample dataset is taken to test the accuracy of the model. The network model trained by the previous dataset is processed with various datasets to perform prediction. The testing face presents the reliability of the model. In the testing phase, trained network which is considered is supervised. Then, the dataset for testing is taken. The data considered for testing is then processed with a trained network to find accuracy, F1-score, and precision considering new test values.

For our analysis, let us train model on first 70% of data and test it on remaining 30%.

Algorithm for proposed model

1. Get dataset of IDS
2. Select characteristics in order to train dataset
3. Set ratio of training to 70% and testing to 30%.
4. Apply LSTM1 layer and LSTM2 layer with dropout layers
5. Apply fully connected layer, softmax layer and perform classification
6. Perform decision making to predict whether label is normal or type of attack.
7. Get content from csv file for testing and consider attributes of records and process textTextData
8. Get trained network net and XNew1 as test Data and apply classifier on to perform prediction on bases of net and XNew1
9. Find accuracy, precision, F1-score using confusion matrix (Fig. 1).

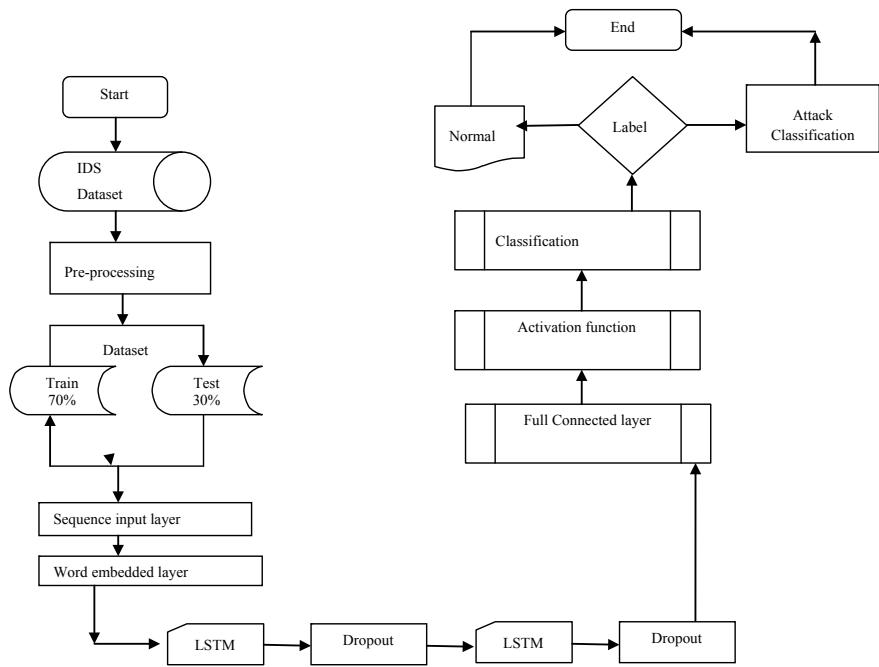


Fig. 1 Proposed LSTM-based IDS detection model

3 Results and Analysis

In the proposed work, MATLAB (2020 a) is used as a simulation tool. MATLAB is used in different sectors of education such as mathematics, academies. It is primarily helpful in universities for research purposes. MATLAB allows the execution of computationally tasks quicker as compared to other languages. It has several toolboxes such as Image Toolbox, Simulink, Simscape. MATLAB is allowing in performing the tasks rapidly.

On the other hand, MATLAB offers to work within complete matrices quickly and easily. In the proposed work, the network is trained using the IDS dataset. Two LSTM layers are used in the proposed simulation environment. The hidden layers, batch size parameters are also defined simulation. The following Fig. 2 is showing a simulation environment in MATLAB to implement the proposed LSTM model.

3.1 Deployment Model

To simulate, the deployment model that considers the training of a dataset of 4,28,500 record is used. The model has been used to perform accurate classification in IDS.

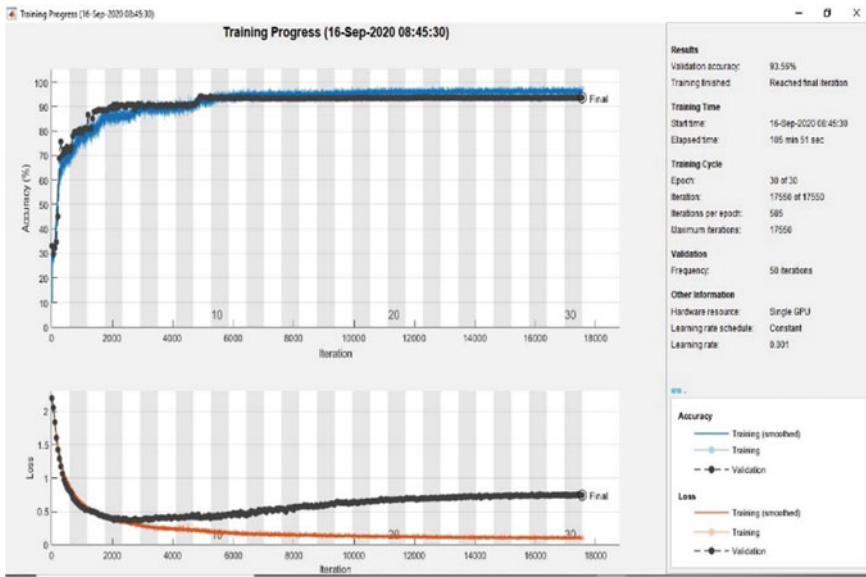


Fig. 2 Training in proposed LSTM model

Before training network model, there is a need for dataset preprocessing. During the initial dataset preprocessing, Shapiro–Wilks is used to eliminate the attribute with less significance. Then, useless attributes that have a single value in all records are eliminated. Then, data is classified as 70% for training and 30% for testing. In the proposed deployed model, two LSTM layers are used. Twelve hidden layers are used in the first LSTM layer, and five hidden layers are used in the second LSTM layer. However, hidden layer has a quality of retaining the previous value, but it also raises a problem overfitting that network will not be robust and will increase the loss function, and it will make network to be useless when there is a need for a new dataset or new values of data; thus, there is need of improvement in the network. Thus, dropout layer with 20% is used that would drop 20% of neurons from the network. This dropout would help remove the problem of overfitting the network as it will make the network more robust toward a new dataset that can be used in the proposed system. A fully connected layer is used to find the label correctly. Such a layer does not drop any neurons. The softmax layer is applied; afterward, an activation layer is used in place of the sigmoid function. It is used where there is a classification of more than two classes takes place. Classification layers are used to perform different predictions of different attacks in IDS. After the training of the network model, testing is performed on the dataset. Thirty percent of the dataset is used for the testing of IDS. Afterward, a confusion matrix is produced considering predicted and actual value to get true positive, false positive, true negative, false negative. The accuracy, precision, recall, F1-score are obtained to get overall accuracy.

3.2 Performance Parameters

The confusion matrix is produced presenting.

3.3 Simulation

3.3.1 Simulation of Proposed LSTM

The training is one of the important steps in IDS. Gradually, the progress of the model is increasing. The progress simulation has been the following figure.

Confusion matrix

After training of the dataset, testing module is run and then the confusion matrix is generated and is exhibited in Fig. 3. The following confusion matrix has shown 9 attributes DoS-Synflooding, MImT Arp spoofing, MiraiAck flooding, Mirai http flooding, Mirai host brute force, Mirai UDP flooding, normal, scan host port, scan port OS. The true classes are presented on the y-axis, and predicted classes are presented on the x-axis. With the help of this matrix, we can calculate different parameters to the efficiency of the system; also, we can compare them.

True Class	Predicted Class								
	DoS-Synflooding	MITM ARP Spoofing	Mirai-Ackflooding	Mirai-HTTP Flooding	Mirai-Hostbruteforceg	Mirai-UDP Flooding	Normal	Scan Hostport	Scan Port OS
DoS-Synflooding	10950				7945				
MITM ARP Spoofing		11220							
Mirai-Ackflooding	45		17193		83	7			8
Mirai-HTTP Flooding				12992	217	4299			
Mirai-Hostbruteforceg	1				38266				
Mirai-UDP Flooding					134	57441			
Normal	102				5		12430		
Scan Hostport								6998	
Scan Port OS									16710

Fig. 3 Confusion matrix of proposed model

Table 4 Accuracy chart of proposed model

Class	N (truth)	N (classified)	Accuracy (%)	Precision	Recall	F1-score
1	4246	5674	97.59	0.75	1.0	0.86
2	3370	3370	100	1.0	1.0	1.0
3	5177	5207	99.95	0.99	1.0	1.0
4	3926	5258	97.75	0.75	1.0	0.85
5	8127	11,490	94.17	0.70	0.99	0.82
6	23,454	17,289	89.58	1.0	0.74	0.85
7	3744	3756	99.98	1.0	1.0	1.0
8	2101	2101	100	1.0	1.0	1.0
9	5018	5018	100	1.0	1.0	1.0

Result

Considering the above, confusion matrix chart presenting accuracy, precision, recall value, and F1-score is generated. The accuracy chart in the case of existing work is presented in Table 4.

3.3.2 Comparison in Case of Previous and Proposed Work

The comparison of the proposed work and traditional LSTM model with a single layer is presented in Fig. 4. The following chart is presenting that the proposed work has more accuracy, precision value, recall value, and F1-scores value as compared to previous LSTM. Considering average value of accuracy, precision, recall, and F1-score from previous and proposed work in from table above, Table 5 has been produced to compare both models.

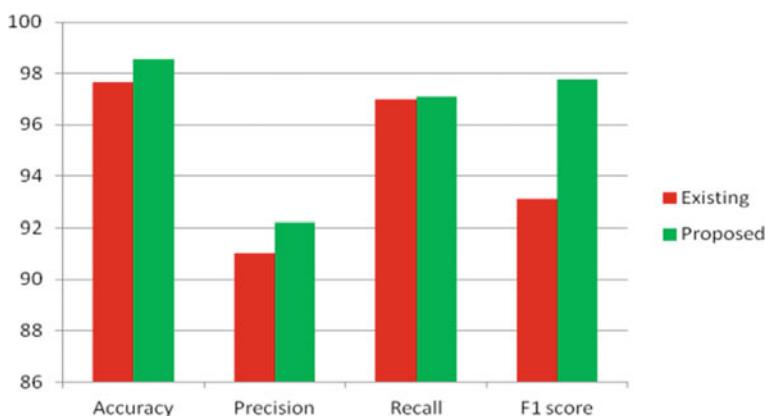
**Fig. 4** Comparison of proposed and previous LSTM model

Table 5 Comparison of previous model with proposed work

	Existing	Proposed	Difference	(%)
Accuracy	97.66	98.54	0.88	0.90
Precision	91	92.22	1.22	1.34
Recall	97	97.11	0.11	0.11
F1-score	93.11	97.77	4.66	5.00

Following chart is presenting comparison of accuracy, precision, recall, and F1-score in case of previous and proposed DEEP LSTM models. It has been observed that there is slight change in recall value, but there is significant difference in F1-score.

4 Conclusions

Through the comparison made between previous and proposed work, the finding is presenting that accuracy has significantly increased. Accuracy is intuitive measurement of performance, while precision is presenting the proportion ratio of right predicted positive results. Recall is representing sensitivity and is considered as a proportion of right predicted positive findings to all findings in real class true. F1-score is adjusted average of precision as well as recall. Finding of proposed model concludes that the accuracy in case of proposed work is 0.9% more as compared to previous work. However, precision, recall are 1.34%, 0.11% more, respectively, in case of proposed work, while F1-score is 5% more in case of proposed work.

5 Scope of Research

The research will play a significant role in predicting IDS with high accuracy. The proposed work, i.e., LSTM-based IDS system for security of IoT has provided a scalable and flexible approach that deals with various phases such as dataset selection, preprocessing, training, and testing. The probability of error would be less while calculating overall accuracy as the proposed model has made use of a huge dataset for training and utilized the LSTM model with multiple layers. Future research can have the benefit of rapid training from this research. The present research has eliminated the useless attribute that is not affecting the overall accuracy of the training model. Thus, the proposed approach has also proposed a fast-training mechanism that could train the model for further prediction by ignoring the characteristics that have less impact on results. Future research could make use of the same model to provide better IDS detection solution.

References

1. Li P, Zhang Y (2019) A novel intrusion detection method for Internet of Things. In: 2019 Chinese control and decision conference (CCDC). IEEE, pp 4761–4765
2. da Cruz MAA, Rodrigues JJP, Al-Muhtadi J, Korotaev VV, de Albuquerque VHC (2018) A reference model for Internet of Things middleware. *IEEE Internet Things J* 5(2):871–883. <https://doi.org/10.1109/IOT.2018.2796561>
3. Mohanta BK, Jena D, Satapathy U, Patnaik S (2020) Survey on IoT security: challenges and solution using machine learning, artificial intelligence and blockchain technology. *Internet of Things* 100227
4. Kozilek H, Burger A, Platenius-Mohr M, Rückert J, Mendoza F, Braun R (2020) Automated industrial IoT-device integration using the OpenPnP reference architecture. *Softw Pract Experience* 50(3):246–274
5. Kaur H, Kumar R (2020) A survey on Internet of Things (IoT): layer-specific, domain-specific and industry-defined architectures. In: Advances in computational intelligence and communication technology, Springer, Singapore, pp 265–275
6. Liu H, Han D, Li D (2020) Fabric-IoT: a blockchain-based access control system in IoT. *IEEE Access* 8:18207–18218
7. Li W, Yi P, Wu Y, Pan L, Li J (2014) A new intrusion detection system based on KNN classification algorithm in wireless sensor network. *J Electr Comput Eng* 2014
8. Buczak AL, Guven E (2015) A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Commun Surv Tutorials* 18(2):1153–1176
9. Javaid A, Niyaz Q, Sun W, Alam M (2016) A deep learning approach for network intrusion detection system. In: Proceedings of the 9th EAI international conference on bio-inspired information and communications technologies (formerly BIONETICS), pp 21–26
10. Bhattacharjee PS, Fujail AKM, Begum SA (2017) Intrusion detection system for NSL-KDD data set using vectorised fitness function in genetic algorithm. *Adv Comput Sci Technol* 10(2):235–246
11. Khan MA, Karim M, Kim Y (2019) A scalable and hybrid intrusion detection system based on the convolutional-LSTM network. *Symmetry* 11(4):583
12. Ashfaq RAR, Wang X-Z, Huang JZ, Abbas H, He Y-L (2017) Fuzziness based semi-supervised learning approach for intrusion detection system. *Inf Sci* 378:484–497
13. Revathi S, Malathi A (2013) A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection. *Int J Eng Res Technol (IJERT)* 2(12):1848–1853
14. Paulauskas N, Auskalnis J (2017) Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset. In: 2017 Open conference of electrical, electronic and information sciences (eStream). IEEE, pp 1–5
15. Tang TA, Mhamdi L, McLernon D, Syed ARZ, Ghogho M (2016) Deep learning approach for network intrusion detection in software defined networking. In: 2016 International conference on wireless networks and mobile communications (WINCOM). IEEE, pp 258–263
16. Sheikhan M, Jadidi Z, Farrokhi A (2012) Intrusion detection with support of reduced-size RNN dependent on feature grouping. *Neural Comput Appl* 21(6):1185–1190
17. Dong B, Wang X (2016) Comparison deep learning method to traditional methods using for network intrusion detection. In: 2016 8th IEEE international conference on communication software and networks (ICCSN). IEEE, pp 581–585
18. Khan AH, Li S, Chen D, Liao L (2020) Tracking control of redundant mobile manipulator: an rnn based metaheuristic approach. *Neurocomputing* 400:272–284
19. Wang F, Xuan Z, Zhen Z, Li K, Wang T, Shi M (2020) A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Convers Manage* 212:112766
20. Althubiti SA, Jones EM, Roy K (2018) LSTM for anomaly-based network intrusion detection. In: 2018 28th International telecommunication networks and applications conference (ITNAC). IEEE, pp 1–3

21. Ming T, dos Santos C, Xiang B, Zhou B (2015) LSTM-Based deep learning models for non-factoid answer selection. arXiv preprint [arXiv:1511.04108](https://arxiv.org/abs/1511.04108)
22. Ullah I, Mahmoud QH (2020) A scheme for generating a dataset for anomalous activity detection in IoT networks. In: Canadian conference on artificial intelligence. Springer, Cham, pp 508–520

Securing Smart Homes Using Face Recognition



Mehul Sinha, Raj Chaurasiya, Arav Pandey, Yuvraj Singh,
and Somya Goyal

Abstract In today's era of automation and smart devices, there is a crucial need to modify our current security system. With the help of technologies like IoT and machine learning, we can not only improve the current door lock security system but also seamlessly integrate it in our daily lives. The machine learning algorithm uses highly sophisticated algorithm in recognizing faces. Only the person whose face is matched can be able to unlock the door. So, limitation of managing keys will be resolved. This technology opens door to a vast number of possibilities where high security is required. Most importantly, this technology is highly affordable so that everyone can have access to maximum security at minimum cost.

Keywords IoT · Smart home · Home automation · Image processing

1 Introduction

In our world that is twenty-first century, smart devices and all are needed to be modified for day to day use, and this modifying will make electronic devices and gadgets smart, and also, in our world, it is not the era where we can trust our old security which mainly includes our old conventional locks for the houses as home security has become very important in our current times. One of the major drawbacks of door lock is that key used in door lock can be made by duplicating key, and the key can also be stolen. Also, traditional keys are easy to lose creating a lot of hassle for the user. So, to overcome these problems, smart door via face recognition is introduced in our society to improve safety of houses and the people living in the houses.

Automated entrances and exits door's system are widely used in every public places such as big malls, businesses office, railway stations, airports to abolish the need of actions done manually by people. Present day sensors-based automatic door lock system's technologies encompass ultrasonic/radio, infrared, and other sensors having wireless methods.

M. Sinha · R. Chaurasiya · A. Pandey · Y. Singh · S. Goyal (✉)
Manipal University Jaipur, Jaipur 303007, Rajasthan, India

We automate the entrance door to next level by integrating face recognition technology to furthermore increase convenience and security of these automate doors. This smart face recognition technology can be integrated for both industrial and residential usages. This system will contain many layers of security and will only grant access to people recognized by the system and will also check their intentions of going in by using various sensors. This will enable to integrate a new generation of smart home and industrial securities [1].

2 Related Works

This section gives the review of literature work in the field of ‘smart door via face recognition’. From literature, there is huge contribution in the field which is tabulated as Table 1.

3 Proposed Methodology

The methodology to perform the proposed work is explained in this section. The prototype is built by combining the part of machine learning and IoT together [11–14]. Components of smart door via face recognition are—camera, fingerprint sensor, Raspberry Pi module, servo door motor, motion sensor. All these sensors are connected to each other and communicate using Internet which is what we call IoT. The whole system will work in the following way (see Fig. 1):

1. When a person arrives at the door, the motion sensor will alert the system and the camera will turn on, capturing various images of the person standing there. The algorithm through the persons movement trajectory will determine whether the person wants to go in.
2. If the person wants to go in, then the system will send various captured images to the database. Then, the algorithm with the help of machine learning will match the face of the person with the profiles in its database. The process is done by the machine using dots and points of importance as stated above.
3. If the system is able to match the face with the database, it will simply send a notification to the administrator of the identified person’s profile and grant them access, while the machine also uses the captured image of that person to further train itself and improve the profile of that person in its database
4. In case of an unidentified person, the system will notify the administrator with the captured person’s images and ask if they want them to grant access. If the administrator grants them access, the algorithm will automatically create a new profile of that person. This will help the system to recognize the same person in the future. The administrator can also instruct it to save the profile for future use.

Table 1 Literature work

S. No.	Study	Contribution	Technique used	Strength	Weakness
1	Hasan et al. [2]	Image recognition in a lock	OpenCV	Faces and other objects identified easily	Less efficiently in low light
2	Pawar et al. [3]	Control of security to owners mobile	Raspberry Pi, IoT, ultrasonic sensor, LBP	Alert the owner through SMS	May stop working during power failure
3	Deshmukh et al. [4]	Automation at local level like home, offices, and campus. It also promotes self-sufficiency	SMIP and IMAP. LBPH and Pi camera	Easy to use and simple	Tough to setup
4	Suhani et al. [5]	Wireless control system and accessibility to a home environment	Zigbee module and ARM1176JZF-S	Webpages and applications in Android	System can be hacked by hackers
5	Kim et al. [6]	Contribution objectives to make it safe. It avoids stepper motors and various wiring process	Technique there has been an efficient use of solenoids	Strength has lots of IoT-based systems Microsoft FACE API	Methodology and systems are complex
6	Mohmmad et al. [7]	This model is unique with its one-of-a-kind combination of functionalities	Computer vision of this system is very safe and simple to others	Microsoft webcam audio speaker Raspberry Pi 3 Model Bt	The process is slow
7	Nasir et al. [8]	Works well in multiface recognition can be controlled from mobiles, laptops, etc.	Security can be controlled from various devices	Can work on Android and IOS devices	Applications can sometimes crash and not work properly
8	Yang et al. [9]	The device will function with an application and will have an interactive UI	It allows more research work in the field of AI and ML	Safe protocols that cannot be breached	No improvement in results
9	Purnama et al. [10]	Automation of office, home, government organizations, or for banks	Face recognition will be done by Raspberry Pi 3	Highly secure	Not capable to detect face from group of persons

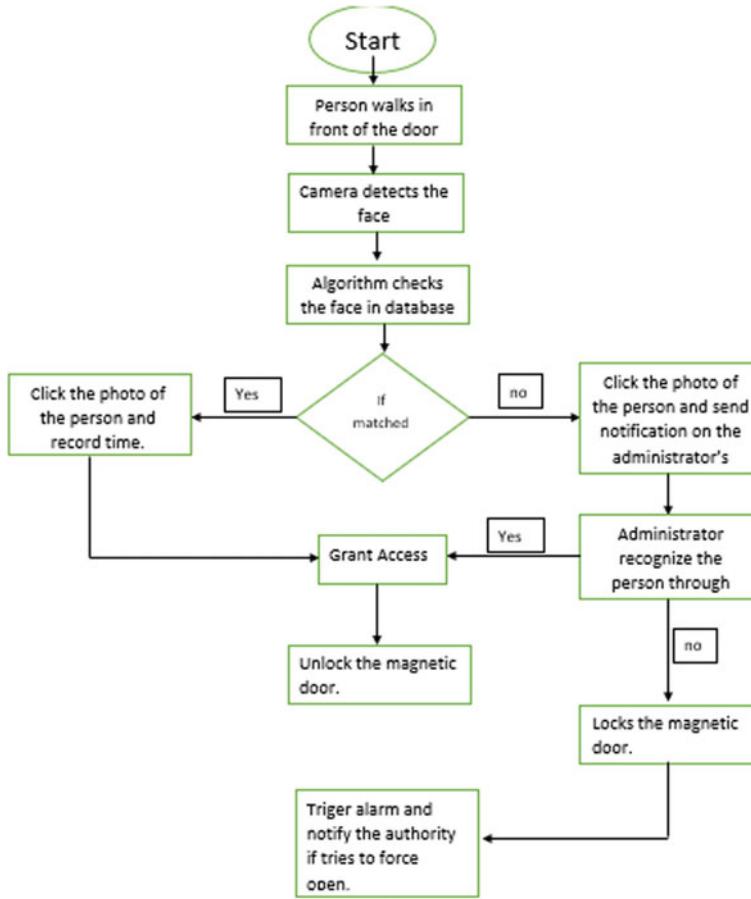


Fig. 1 Proposed model of smart home system

5. The machine learning algorithm needs to adjust with the small changes in facial features in a human. This includes new haircuts, facial hair, makeups, wearing eye glasses. Therefore, it needs to keep updating its database with new images. This is a way to further train the algorithm. Hence, every time it recognizes a face, it will save those images and train itself.
6. Once the person is granted access by the system, a message will be sent by the Raspberry Pi to the servo motor. The servo motor will produce desired amount of torque for rotating the door handle.
7. In case of failure of face recognition, by any reason, then for backup, a fingerprint sensor will also be installed by the door side. This will enable the person to enter even in the absence of the administrator.
8. If an unauthorized or unrecognized person tries to force open the door, the motion sensor combined with the machine learning will recognize that someone

is trying to forcefully enter and will simultaneously ring the alarm and inform administrator.

4 Results and Discussions

We finished building a face recognition system that uses machine learning algorithms to build a smart door. Face recognition and all the devices connected through IoT work as expected. There is high accuracy in recognizing faces, and it successfully communicates through other devices using IoT. We all are adequately satisfied to build it. Our system is better because our systems have been enhanced, and the response time is very quick making this project practical. The webcam captures high-resolution photographs and quickly runs it against the database. It is a hassle-free process which is very user-friendly, and it takes a little bit time to recognize.

To assess the performance of proposed algorithm and model, we chose four people at random and tested our algorithm by making them wear different accessories like—glasses and face mask. We also observed their right-view face and left-view face. Table 2 shows the result of how our algorithm performed under various conditions. Each person went through the test five times.

Table 3 shows the accuracy and error rate of the algorithm. For this test, we chose 10 people and made them go through the machine 100 times in a controlled environment. The results are shown as Fig. 2.

The project was completed and is working as presented in this proposal. The system allows the authorized user to monitor and control the access to the door remotely. If the intruder is detected after the image processing is done in the IoT, the image is captured, and this intruder's image is sent to the authorized person. The admin can effortlessly create new profiles allowing ease of handling of the security system. Overall, it is a very secured security system with promising results, and the low cost of the system makes it easily affordable. The only drawback we noticed is that the webcam sometimes struggles in low-light conditions. To counter this, we have also given an alternative entry through a biometric scanner. This makes our system highly secure and at the same time fail proof.

5 Conclusions

In this work, we have implemented smart door unlock system by using face recognition system. The smart face security system is able to detect the face accurately and recognizes the face and then checks the database. Raspberry Pi controller controls the door access. This system can be used in places where there is a need of high security. The system used in our work is of moderate cost, so an average-salaried person can also use this system. The project discusses in intricate detail about a smart home protection system which not only is useful for a small-scale integrated system.

Table 2 Performance of algorithm recorded

Person ID	Time to detect face	Time to detect face with glasses	Time to detect face with mask	Time to detect right-view face	Time to detect left-view face
P1	1.35	1.61	3.82	2.08	2.82
	1.40	1.58	4.01	2.75	3.16
	1.32	1.42	3.97	2.76	2.60
	1.37	1.46	3.66	3.22	3.20
	1.48	1.39	4.28	2.45	3.42
P2	1.33	1.55	3.92	2.76	3.32
	1.29	1.48	3.77	2.63	2.96
	1.36	1.62	4.28	3.02	3.12
	1.50	1.38	4.39	3.12	3.01
	1.31	1.40	3.98	2.93	3.22
P3	1.43	1.62	3.72	3.12	3.26
	1.38	1.51	3.88	3.22	3.32
	1.36	1.39	4.12	2.89	3.36
	1.32	1.47	4.02	3.01	2.88
	1.34	1.44	3.96	2.82	3.52
P4	1.60	1.58	4.28	2.92	3.61
	1.38	1.57	4.38	3.10	2.92
	1.42	1.49	4.13	2.88	3.01
	1.36	1.61	3.97	3.16	3.55
	1.39	1.62	3.90	2.90	3.28

Table 3 Accuracy and error rate

Person ID	Access granted	Access denied	Accuracy (%)	Error (%)
P1	100	0	100	0
P2	100	0	100	0
P3	99	1	99	1
P4	100	0	100	0
P5	99	1	99	1
P6	98	2	98	2
P7	100	0	100	0
P8	100	0	100	0
P9	99	1	99	1
P10	100	0	100	0
Average			99.5	0.5

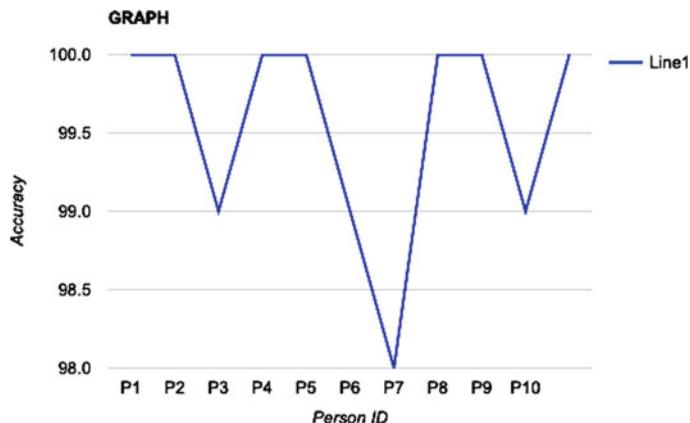


Fig. 2 Accuracy of algorithm with ten testing instances

The IoT-based device will automatically remove the need of manually locking and unlocking the door. A cheap and effective home security system is the need of the hour in the current scenario. Most of the home security systems involve the use of very expensive, heavy scale, and overly emphasized components and supplements. This device also provides an extra security layer to the residents to stay inside the house.

References

1. D Deshmukh A, GNakrani, M, L Bhuyar D, B Shinde U (2019) February. Face Recognition using OpenCv based On IoT for smart door. In Proceedings of international conference on sustainable computing in science, technology and management (SUSCOM), Amity University Rajasthan, Jaipur-India
2. Hassan A, Fahamida Z, Sen P, Hossain SA, Afanul (2021) Smart lock using image recognition. Global J Comput Sci Technol [S.I.J]. ISSN 0975-4172
3. Pawar S, Kithani V, Ahuja S, Sahu S (2018) Smart home security using IoT and face recognition. In: 2018 Fourth international conference on computing communication control and automation (ICCUBEAE), Pune, India, pp 1–6. <https://doi.org/10.1109/ICCUBEAE.2018.8697695>
4. D Deshmukh A, G Nakrani M, Mahender, L Bhuyar D, B Shinde U (2019) Face Recognition using OpenCv Based on IoT for smart door. In: Proceedings of international conference on sustainable computing in science, technology and management (t-), Amity University Rajasthan, Jaipur, India, 26–28 Feb 2019
5. Sahani M, Nanda C, Sahu AK, Pattnaik B (2015) Web-based online embedded door access control and home security system based on face recognition. In: 2015 International conference on circuits, power and computing technologies [ICCPCT-2015], Nagercoil, pp 1–6. <https://doi.org/10.1109/ICCPCT.2015.7159473>
6. Kim Y, Yoo J-H, Choi K (2011) A motion and similarity-based fake detection method for biometric face recognition systems. IEEE Trans Consum Electron 57(2):756–762
7. Sajjad M et al (2020) Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities. Future Gener Comput Syst 108:995–1007

8. Sajjad M, Nasir M, Muhammad K, Khan S, Jan Z, Sangaiah AK, Elhoseny M, Baik SW (2020) Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities. Future Gener Comput Syst 108:995–1007
9. Yang JC, Lai CL, Sheu HT, Chen JJ (2013) An intelligent automated door control system based on a smart camera. Sensors 13(5):5923–5936
10. Rizki RP, Hamidi EAZ, Kamelia L, Sururie R (2020) Image processing technique for smart home security based on the principal component analysis (PCA) Methods. In: 2020 6th International conference on wireless and telematics (ICWT), pp 1–4
11. Goyal S, Bhatia PK (2020) Feature selection technique for effective software effort estimation using multi-layer perceptrons. In: Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering, vol 605. Springer, Cham. pp 183–194. https://doi.org/10.1007/978-3-030-30577-2_15
12. Goyal S, Bhatia PK (2019) A non-linear technique for effective software effort estimation using multi-layer perceptrons. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), Faridabad, India, pp 1–4. <https://doi.org/10.1109/COMITCon.2019.8862256>
13. Goyal S (2020) Heterogeneous stacked ensemble classifier for software defect prediction. In: 2020 Sixth international conference on parallel, distributed and grid computing (PDGC), Waknaghat, Solan, India, 2020, pp 126–130. <https://doi.org/10.1109/PDGC50313.2020.9315754>.
14. Goyal S, Bhatia PK (2020) Comparison of machine learning techniques for software quality prediction. Int J Knowl Syst Sci (IJKSS) 11(2):21–40. IGI Global. <https://doi.org/10.4018/IJKSS.2020040102>

Smart Quiz for Brain Stormers



Aditya Kumar, Rishabh Gupta, Naivedhya Sharma, and Somya Goyal

Abstract In this pitiable state of world-wide pandemic, everybody needs sources of entertainment or leisure activities provided going out is not an option anymore. During lockdowns, some people look for various sources of entertainment, whereas some people utilize their time by learning a new skill or increasing their knowledge. This game provides the best of both worlds, by acting as a source of entertainment as well as by acting as a source to enhance one's knowledge. The game's report found determination in the users to get better results with each round. It keeps users engaged, and rather than doing something unproductive in their spare time, they increase their knowledge in a fun way.

Keywords Quiz online · Brainstorming · Internet gaming · Smart application

1 Introduction

Television quiz programs like KBC and Who Wants to Be A Millionaire are becoming increasingly popular. These kind of game shows have large viewership as they offer cash in form of reward. Everyone wants to be a part of these shows or at least wants to feel involved in it in one way or another. Television quiz shows' strength lies in the cash prize, the reach of the television and the host. Many people watch the show just because of the host.

These shows have some limitations as well. It only involves one person at a time. So, the rest of the people are just watching the game and sitting ideally. Also, the participant can face embarrassment in front of a large viewership in case of an incorrect answer. Sometimes, strength of the show becomes its weakness as the host must be paid heavily in case to ensure the large viewership.

There is a need for a program which is interactive and allows everyone to participate. It must also provide the excitement of live participation in competitive contests

A. Kumar · R. Gupta · N. Sharma · S. Goyal (✉)
Manipal University Jaipur, Jaipur 303007, Rajasthan, India

while sitting in the comfort of homes. Television does not provide a two-way interface, it is one of its biggest limitations, and we need to overcome that.

The world has gone through a pandemic, and online trend, in every field, has gone through an exponential change. The demand for online games, online shopping and OTT subscriptions has increased significantly. The number of online gamers has increased significantly thanks to the ample free time available in the lockdown. Several people started participating in online gaming for good reason. Many young gamers started playing these esports tournaments and earned money by playing games.

During this time, every major esports portal reported a surge of 80–100% of users. While the number of people registering on esports portals was always rising, the pandemic just provided a boost to this rising trend. The maximum growth in esports companies was observed in the months after March. More users signed up for these companies in March as compared to January and February. This trend continued in April and May. May observed the maximum daily signups.

Since there was a ton of spare time available, people got more into the online platforms. Some people spent their time on games and some on increasing their knowledge. A quiz game provides the best of both worlds, serving as a source of entertainment and increasing knowledge.

This quiz game will allow users to choose a topic of their interest and then test their knowledge in that field. Users will either win or learn in this game without any fear of embarrassment or any greed. This game will also give users, who have plenty of spare time and who want to increase their knowledge, a leisure activity.

This quiz game is a multiple-choice questions-based game in which users are asked questions about a topic of their choosing. So, a user can either test his/her knowledge or challenge themselves by choosing a topic they find difficult.

This game will keep the users engaged. Rather than doing any unproductive activity to spend their leisure time, users can play this game and increase their knowledge. This game is perfect for users who have ample spare time and are interested in increasing their knowledge in a fun way. This game makes education fun!

2 Related Works

This section gives the review of literature work in the field of ‘smart gaming’. From literature, there has been a huge contribution in this field and many papers have been written suggesting ideas to make the project better. Listed below are some research papers used to make this study. Douligeris et al. [1] proposed a technique of using quiz games in classes to make the learning more efficient and effective. Quiz games can be used to make classes fun and interesting and can simultaneously review the concepts taught in previous sessions. They observed that CSCL trends are evolving constantly and rapidly. They suggested using various programming languages to make these quizzes.

Welbers et al. [2] proposed an idea of a gaming application for helping students learn about their subjects and course and give personalized feedback rather than generic feedback. They also included a session limit to stop users from playing the game too much as it could hamper deep learning.

Howell et al. [3] proposed idea of MCQ quizzes for fast assessment of the tests conducted. MCQ-based e-quizzes can give immediate results and can save a lot of time.

Kevin et al. [4] proposed the idea of using quiz games to improve the mood of people. People suffering from anxiety and stress can play the game to elevate their mood. The game can be developed on various topics, and the person can choose a topic of his/her own interest.

Aljezawi et al. [5] proposed the idea of using quiz games to check students' performance, knowledge and retention using a gaming application. It was found that gaming is more acceptable and is more liked by the students. It is clearly a better method of learning.

Krouská et al. [6] proposed a personalized quiz game, which used brain-based learning and Marzano Taxonomy as basic principles for improving user's higher order cognitive skills. The quiz content took in consideration the student's knowledge level, emotional state and the learning goal set. This approach revealed that this method had a positive effect on a student's performance and has outperformed the traditional systems of e-assessment.

Forrest et al. [7] proposed a game which included a number of cells, selected by users that extended from start to finish points, and each cell was compared with a character that gave an answer to a particular question/clue, provided as a category. One/plural participants moved to the finishing point by selecting cells whose characters, if in order, gave a correct answer. The participant was awarded a score on the basis of characters that formed a correct answer.

Hwang et al. [8] proposed the system 'Joyce', in which participants competed against other real-life participants or computer systems or played through the Internet. The system is adapted the format of multiple-choice questions (MCQs). It provided an environment that stimulated students' motivation to grasp knowledge. Participants were highly motivated to operate the system.

3 Proposed Methodology

This section describes the proposed working model for smart quiz (shown as in Fig. 1). The game algorithm has been designed to ask user questions related to the topic of his\her choice. There are a total of three topics available to the user as choice. The topics are: sports, movies and general knowledge. Thereafter, the user is asked five multiple-choice questions related to the topic of his\her choice. For each question, there are four options. Out of these four options, only one option is correct. There is no negative marking for the wrong answer. For each correct answer, one mark is awarded.

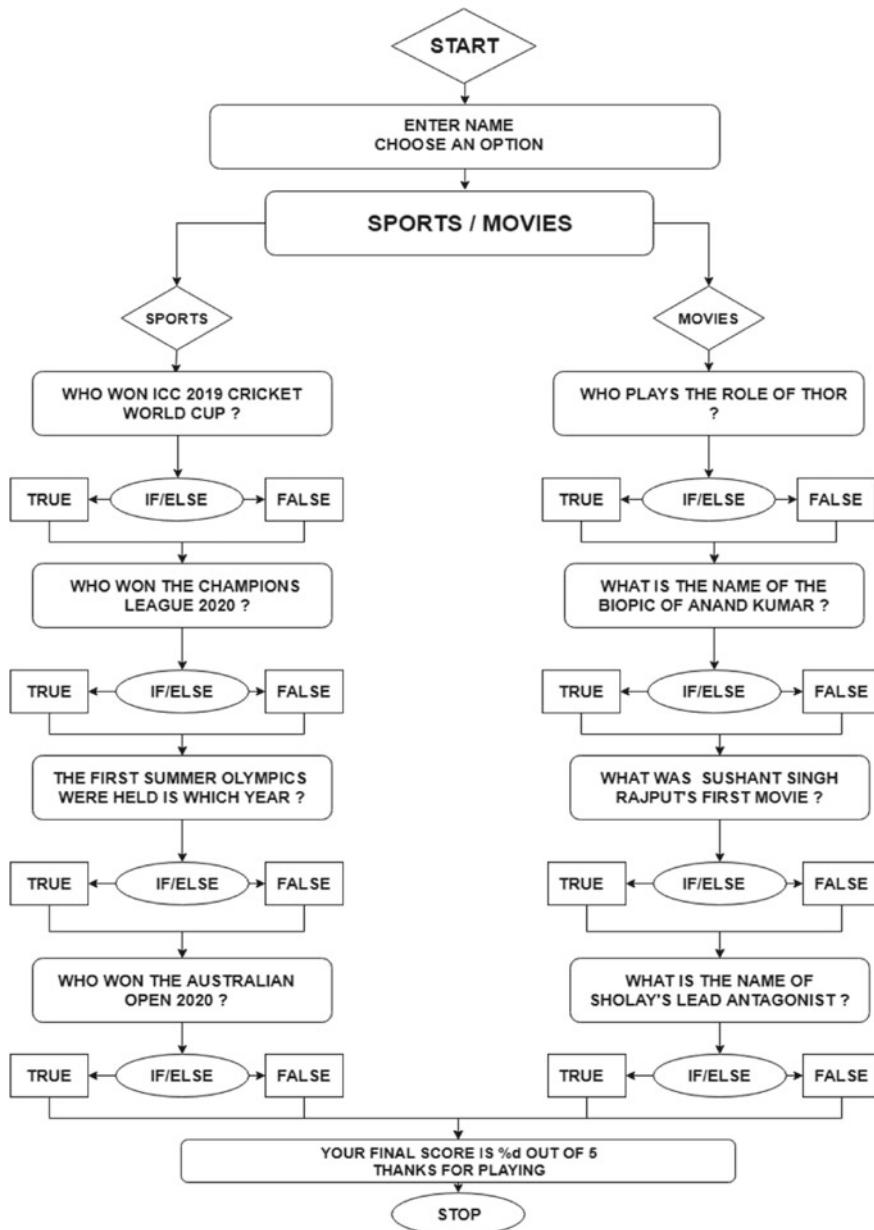


Fig. 1 Algorithm proposed for smart gaming

As soon as the player or the user enters the game, he\she will have to start the game following the instructions given. Thereafter, the user will enter his\her name. There will be a given set of instructions about how to play the game. The topic choice will be made available to the user after that. The user will play the quiz related to the topic of his\her choice. At last, the total score secured by the user will be displayed (out of five). This game has been designed using the basics of Python language. The game has been designed to make the game interactive and fun.

Furthermore, the algorithm of the game has been designed in such a way that it will keep the user engaged. Participants will receive their marks as the feedback and will know areas of their strengths and weaknesses.

The code of this game is easy to read and understand. The code is quite simple and can be modified easily to add some other features or add more topics or questions in the game. This game's algorithm can also be used to develop quizzes on a variety of topics for the effective and deep learning of the users.

Using this game, users can self-analyze themselves and check their knowledge in the topic of their interest or can challenge themselves in the topic they find difficult. This game will increase knowledge of the users and will also serve as a source of entertainment.

4 Results and Discussions

The results obtained are explained in this section as follows. The users played the game in one of the three sections provided in the game. The score obtained by the user was printed on the screen. The users who answered all the questions switched to the other quizzes to challenge themselves even further. Meanwhile, the users who did not perform well in the chosen quiz retried the quiz to get the result they desired. They also got to learn new things and increased their knowledge.

Also, the users were engaged in the game for a while, tried out all three components of the quiz and tried to advance and get perfection in every single one of them while learning new things along the way. Users seemed determined to beat their past scores and set new records. Users felt a feeling of satisfaction when they did better than the former rounds.

This game is different and better than other projects in a way that it offers users a choice and does not ask random questions. Though, this game also has a section for the users who want to check their knowledge in every field. But, this game also provides users a choice to choose a topic of their own interest and play accordingly. We have recorded the degree of players' participation for further study using machine learning and predict the choice of players as shown in Fig. 2.

Further, we performed an analytical study regarding the impact of 'not being involved in physical activity' on the mental health. We also analyzed the benefit offered by our game to the players in this pandemic era (see Fig. 3).

Our findings may be preliminary and in the basic stage, but we are very optimistic and positive about continuing this line of research in the future. The research suggests

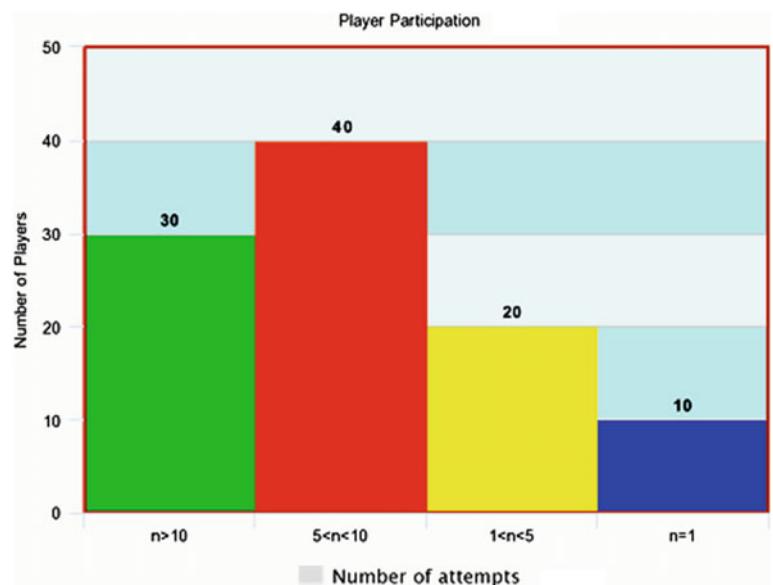


Fig. 2 Analysis of player's participation

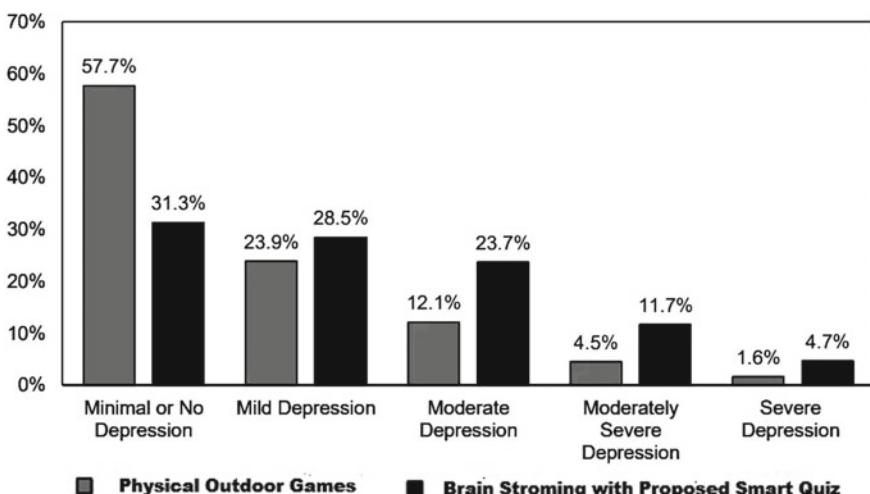


Fig. 3 Impact of player's participation on mental health

that careful use of the game mechanisms can have a long-lasting impact on sustaining and encouraging play via a gamified application. Despite various limitations of this study, we are positive and optimistic about our research and our findings and hope to continue this line of work to encourage the gamifying educational settings.

5 Conclusions

As the pandemic started last year, everyone was quarantined in their house. There were not many things to do when you are locked in your houses like no party, no sports, no learning properly. This is the reason which motivated me to make an online quiz game which will help people to learn knowledgeable stuff from playing quiz game.

There were a lot of problems in the starting about what type of quiz game we should make, on what topic we should do and for which type of people. We came up with a solution that we make a quiz game on general knowledge which can be played by anyone. The game contains questions on various fields which help people to spend time and increase their knowledge.

The concept of making the project came from the reality shows such as KBC and How To Become A Millionaire which are broadcasted on television, there are many other television shows, and there are many quiz games also which people play, so we made this game by using C programming language. In the future, the quiz is proposed to be automated using machine learning [9–12].

References

1. Douligeris C, Seralidou E, Gkotsopoulos P (2018) Let's learn with Kahoot!. In: 2018 IEEE global engineering education conference (EDUCON), Tenerife, pp 677–685. 1109/EDUCON.2018.8363296
2. Welbers K et al (2019) Gamification as a tool for engaging student learning: a field experiment with a gamified app. *E-Learning Digital Media* 16(2):92–109
3. Howell DD, Tseng DCY, Colorado-Resa JT (2017) Fast assessments with digital tools using multiple-choice questions, *College Teach* 65(3):145–147.1080/87567555.2017.1291489
4. Koban K, Breuer J, Rieger D, Mohseni MR, Noack S, Bente G, Ohler P (2019) Playing for the thrill and skill. Quiz games as means for mood and competence repair. *Media Psychol* 22(5):743–768. <https://doi.org/10.1080/15213269.2018.1515637>
5. Aljezawi ME, Albashtawy M (2015) Quiz game teaching format versus didactic lectures. *British J Nursing* 24(2):86–92
6. Krouská A, Troussas C, Sgouropoulou C (2020) A personalized brain-based quiz game for improving students' cognitive functions. In: International conference on brain function assessment in learning. Springer, Cham, pp 102–106
7. Forrest AR, Pruzan AJ, Belyea SA, Adams MR, Garrett PG (2001) Forrest Pruzan creative LLC, 2001. Interactive quiz game system and method. U.S. Patent 6,322,074
8. Hwang JP, Wu TT, Huang YM, Huang YM (2012) Development and evaluation of peer feedback in the English quiz game design in social network. In: 2012 IEEE 12th international conference on advanced learning technologies. IEEE, pp 235–239
9. Goyal S, Bhatia PK (2020) Feature selection technique for effective software effort estimation using multi-layer perceptrons. In: Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering, vol 605. Springer, Cham, pp 183–194. https://doi.org/10.1007/978-3-030-30577-2_15
10. Goyal S, Bhatia PK (2019) A non-linear technique for effective software effort estimation using multi-layer perceptrons. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), Faridabad, India, pp 1–4, <https://doi.org/10.1109/COMITCon.2019.8862256>

11. Goyal S (2020) Heterogeneous stacked ensemble classifier for software defect prediction. In: 2020 Sixth international conference on parallel, distributed and grid computing (PDGC), Waknaghat, Solan, India, pp 126–130. <https://doi.org/10.1109/PDGC50313.2020.9315754>
12. Goyal S, Bhatia PK (2020) Comparison of machine learning techniques for software quality prediction. *Int J Knowl Syst Sci (IJKSS)* 11(2):21–40. IGI Global. <https://doi.org/10.4018/IJKSS.2020040102>

Crash Detection-Based Fleet Tracking System Using VANETs



Parveen, RishiPal Singh, and Sushil Kumar

Abstract The rate of crashes on road is very high in India. Statistically, one fatality occurs in every fourth minute due to crash. The most prominent factors behind this are the high-speed driving, feeling sleepy, eating while driving, making phone calls, mixing the drinking and driving and ignorance to the traffic rules. To reduce the chances of crashes, some parameters namely the accurate location, the route information, the speed of vehicle and the alcohol degree consumed are to be monitored actively and constantly. An intelligent system is deployed to record these measurements, comprising of alcometer, smoke detector, camera, GPS tracker, automatic braking system, IR eye blink sensor and pressure sensor. On these parameters, the vehicles moving on the road can be classified as risky (crash-prone) or non-risky (non-crash-prone) using machine learning algorithms. Now, the vehicle which comes under the class of ‘risky’ one in real time can be saved by sending safety alarms. The proposed work is to allocate an ambulance to such a vehicle in real time in worst-case situation. Primarily, the driver and his concerned persons are to be intimated about the scenario on real-time basis. Multiple machine learning algorithms are implemented to find the most accurate model of detection of the crash and to avoid the predicted crash.

Keywords Vehicle tracking · Internet of vehicles (IoV) · Internet of Things (IoT) · Vehicular ad hoc networks (VANET) · Fleet tracking · GPS

Parveen (✉) · R. Singh

Department of Computer Science and Engineering, Guru Jambheshwar University of Science & Technology, Hisar, Haryana, India

S. Kumar

Jawaharlal Nehru University, New Delhi, India

Table 1 Year-wise record of crashes in India

	Road accidents	Percentage change	Persons killed	Percentage change	Persons injured	Percentage change
2014	489,401		139,661		493,464	
2015	501,432	2.46	114,123	4.62	500,278	1.37
2016	480,651	-4.14	150,775	3.16	494,623	-1.12
2017	464,909	-3.28	147,912	-1.90	470,965	-4.77
2018	467,043	0.46	151,417	2.36	469,408	-0.32

1 Introduction

Nowadays, smart city is improving the living standards of people making the life more comfortable and safer [1, 2]. As a matter of fact, one of the major arenas of our life is the ambiance in which we live. It includes the mechanism of transportation that a person uses to commute from home to office or office to home or from any place to any another place. While commuting from one location to another location, it is essential to ensure the safety of entire fleet system so that the lives boarding can be ensured safe. The fleet system is exposed to life-threatening risks in this changing scenario of road transportation. The reasons are increased number of vehicles causing heavy congestion and negligence to road traffic rules. Road accidents are one of the major sources of loss of life and property damage every year.

1.1 State of Crashes in India

The National Crime Records Bureau (NCRB) 2016 report highlights the high rate of death due to road accidents [3]. The reports suggest that we are lacking in the fleet tracking system somewhere. In India, every minute, an accident happens!—as per the Ministry of Road Transport and Highways [4]. The rate of road accident is very high in India. Table 1 depicts the 5-year data of crashes occurred in India.

It is really unfortunate that being so developed country, still we are lacking in fleet tracking system. The idea of fleet tracking systems is about an intelligent fleet mechanism based on the real-time gathered information to assist the drivers to avoid congested routes and increase safety imposing the traffic rules strictly.

1.2 Wireless Sensor Networks (WSNs) and Internet of Vehicles (IoV)

Wireless sensor networks (WSNs) and Internet of vehicles (IoV) have proven a boon for the fleet tracking system. As from last few years, researchers have contributed

so many new facets for more effective and safer traffic systems developed within the smart city projects. The target of fleet tracking system is to reduce the rate of road accidents, and in case, if it occurs, then the emergency services should be delivered in time to the accident spot without any delay. Various factors behind the road accidents and delay in the delivery of emergency services are: traffic congestion, mixing the driving and drinking, negligence to traffic rules and distracted driver or influenced driver. In our modern societies, the living standards have been raised with luxuries, but at the same time, we are risking our lives. The luxurious lifestyle has resulted in immense traffic on our roads with so many drivers who practice over speeding, rash driving, mix drinking while driving or neglecting the traffic rules.

1.3 Motivation Behind the Study

In India, government is taking serious steps to reduce the rate of road accidents via enforcing the traffic rules strictly all over the country. It is a pitiable situation when our safety is being endangered by the road accidents and timely non-availability of emergency services. Fleet transportation system has not yet so developed intelligently that it can resolve the congestion problem and can reduce the rate of road accidents. But, our collaboration of researchers and academicians has connected the vehicular ad hoc networks (VANETs), cloud and vehicles in a way that the driver can plan and dynamically decide the route to avoid traffic congestion. Fleet tracking systems are imparted with intelligence by means of vehicular ad hoc networks (VANETs), sensors, GPS and machine learning (ML) techniques along with the cloud-based computational powers.

1.4 Contribution

We are proposing to equip our vehicles and emergency services with smart devices for threefold benefits. So, one of the major contributions of this work is to make the vehicles adaptive to real-time traffic data to avoid the routes which are heavily congested; second is to embed the sensors and cameras to monitor the driver's state of mind constantly while driving to avoid the road accidents; and third is to inform the emergency services in case of accident happens. The proposed algorithm employs ML proficiencies so that vehicles can avoid road accidents, and in case of accident, the emergency services can be made available on time to the victim.

1.5 Organization of the Study

This chapter is organized as follows. Section 2 describes the fundamentals and current state-of-the-art of fleet tracking system, while Sect. 3 discusses the prototype model, the methodology being used and the features collected and selected on which making the prediction depends. Section 4 discusses the results with a comparative analysis of machine learning algorithms. Section 5 concludes the chapter with remarks on the future scope of the work.

2 Fundamentals of Fleet Tracking System and Related Work

This section is to bring the basic introduction of the fleet tracking system. It also discusses the work which has been contributed by various researchers in the field of fleet tracking system.

In today's time, VANETs-based applications are developed to track moving vehicles to ensure safety on the road. VANETs [5] provide important interface for fleet tracking system as shown in Fig. 1.

Jorge Zaldivar et al. [6] introduced an app which can view the on road moving vehicle via some onboard diagnostics (OBD-II) interface, with a facility to spot the accidents. Another app [8] based on Android was developed by Athavan [7] to locate and notify mishaps. Cai et al. [9] deployed few context-aware applications.

All the systems, we discussed above, were built with the target to respond to the accident after it has been occurred. Such system can support the life-care services after the mishappening caused. Now, we shall discuss the systems of what we are really concerned! The prevention systems!

The most exciting one is a mechanism developed using IR sensors to detect the drowsiness of driver to prevent accident that has been proposed to detect drowsy driving using eye blink sensor that is an infrared sensor (IR sensor). The concept is to monitor and control the rate of blinking the eyes using IR sensor. [10]. The status

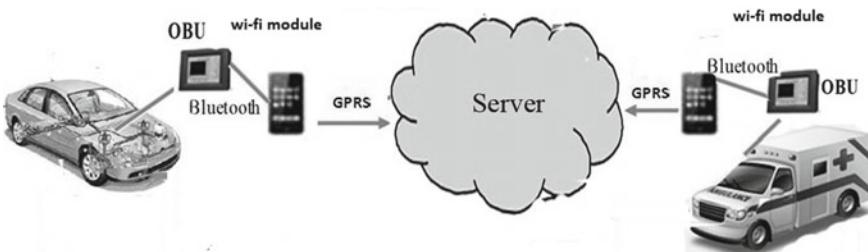


Fig. 1 Fundamental units of fleet tracking system

of IR receiver corresponds to the position of eye. Hence, it is monitored that whether the eye is in the closing position or in the opening position. A logic is programmed to set a threshold in the time and frequency to indicate the alarm. The system is developed to prevent accidents due to unconsciousness or drowsiness, through eye blink frequency [11].

Another important research work to be discussed in this section is how to manage and re-route the moving vehicles in real time via algorithm development [12–14].

Now, the time is to discuss one of the fundamental requirements for the fleet tracking system which has kept researchers attracting is how to make the delivery of real-time messages in more efficient and cost-efficient ways quickly [15, 16]. It has been reported that the route planning for fleet tracking system to avoid congestion and prevent accidents can be coordinated easily as the RSUs in VANET can improve the data collection and distribution in effective ways [17].

Now, the aspect to be discussed is associated with the luxury and quality of life that has been attained in this modern era. It is to deal with the delivery of multimedia data by a vehicular network that can be a bottleneck causing the transmission delays [18, 19]. After the discussion about the current state-of-the-art of the fleet tracking system, next, we shall move to the prototype model of the fleet tracking system so that the practical implementation can be demonstrated in a better way.

3 Prototype Methodology

This section discusses a prototype model explaining the working methodology of the system. It consists of an On-Board Unit (OBU), a server and roadside units (RSUs) as shown in Fig. 2. The vehicles moving on the road are equipped with OBU. The RSUs are installed at the road intersections. The real-time data which is being constantly collected via sensors fabricated in OBUs is stored and processed at the database server. VANET-based application is implemented for communication among all the units of the system. As the vehicle is moving on the road, sensors installed in OBU kept inside the vehicle continuously record the location, speed, heartbeat of the driver, the eye movement of driver, the degree of smoke and liquor consumed. This data is uploaded to the database server where machine learning algorithms are implemented to predict the degree of accident-proneness of the moving vehicle. As per the results, the vehicle is sent an alert and also emergency services are contacted in lieu of prevention of accident. In case, just a collision occurs instantly, then there is a facility of detecting the accident automatically and reporting to the emergency services.

The sensors embedded within the OBU to collect real-time dataset and to connect the OBU to the server or to the emergency services and to alert the driver are listed in Table 2. As per the mentioned table, Sensor nos. 1, 2, 3, 4, 5 and 10 are to sense the data. Module nos. 6, 7, 8 and 9 are to connect the OBU with the external entities [20].

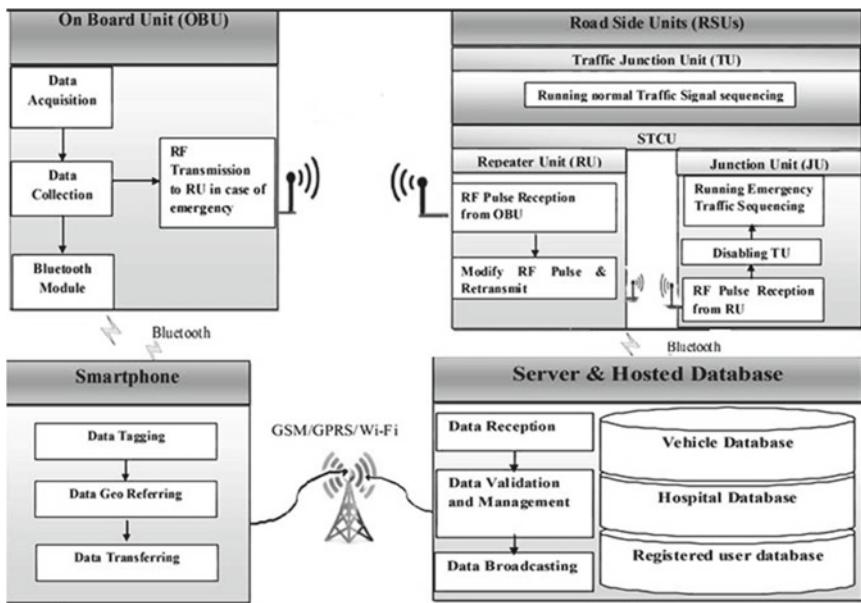


Fig. 2 Prototype of fleet tracking system

Table 2 Sensors on OBU and the data collected

S. No.	Sensor/module boarded	Parameter recorded
1	Ambient light sensor	Daytime/night time
2	Speedometer	Instantaneous speed
3	Alcometer	Degree of alcohol
4	Smoke detector	Degree of smoke
5	Pressure sensor	Pressure
6	Wi-Fi module	To connect
7	Bluetooth module	To connect
8	GPS module	Live location
9	GPRS module	To communicate
10	Camera	To monitor driver continuously

Table 2 gives the dataset attributes being collected by the OBUs in this specific prototype model. The set of attributes can be adjusted as per the requirement of situation. This prototype considers few major factors [21–25] which cause the road accidents including the over-speed driving, influenced driver (due to alcohol consumption or smoking), lack of lights, drowsiness of driver (camera to observe facial expression and eye movement).

Next is to study the analysis of data to make prediction before the accident occurs so that any mishap can be prevented before time. It is the most crucial part of entire fleet tracking system. It is about making the predictions about the accidents which may or may not occur. It classifies the moving vehicle as accident-prone or safe one. This classification is done with the machine learning algorithms under the supervised classification algorithms [25–27].

It is pertinent to mention here that this particular candidate prototype also uses deep learning for image processing. The image identification and feature selection are done with deep learning algorithm namely Haar algorithm [28]. Each picture taken by camera module passes through four core phases for feature extraction and detection of drowsy driver: (1) Haar-based feature selection, (2) creating an integral image, (3) training of the classifier using AdaBoost algorithm and (4) cascading classifiers to detect drowsy driver. Image processing for the live input recorded from the camera module is based on Viola-Jones object detection framework [29]. In this prototype, it is taken as basis for processing the captured live images for the reason that it is robust in nature [30]. Haar features are the most common features of every object of a specific class. Adaptive boost (AdaBoost) is ensemble ML algorithm that clubs the performance of multiple ML algorithms to increase the efficiency of the model [31]. AdaBoost is not so quick to detect the object in real time [32]. The solution is to arrange the strong classifiers in cascading fashion and to train these classifiers on only particular features which pass through the preceding one [33]. Open Source Computer Vision Library (OpenCV) is used for this prototype for computer vision and machine learning library functions [34]. The eye position and eye movement are assessed to detect the drowsiness of driver and further to imbibe the results in the overall prediction model. The upcoming section brings light to the analysis of performance of multiple machine learning algorithms [35] used in this prototype and discusses the result.

4 Results and Discussions

Now, we are at the point to discuss the performance of various ML algorithms which have been used in the above discussed prototype in a comparative manner. The aim of this particular section is to bring the clarity to the audience about the performance and accuracy of the prototype model. Let us discuss the criteria to compare the performance of executed ML algorithms. We consider the most popular criteria: ROC curve.

4.1 Receiver Operating Characteristic Curve (ROC Curve)

It is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: true positive rate and false positive rate (Fig. 3).

The following Fig. 4 shows a typical ROC curve [35].

For Classification of moving vehicle as accident prone or safe, we have implemented two most popular classification algorithms namely, Artificial Neural Network (ANN), and Support Vector Machine (SVM).

The implementation has been carried out in MATLAB with standard algorithms, and performances are recorded as in Fig. 4.

The accuracy of overall model is pretty good. ANN shows the best results with the data collected by the On-Board Unit (OBU) installed in the vehicles. The nonlinear

Fig. 3 TP versus FP rate at different classification thresholds

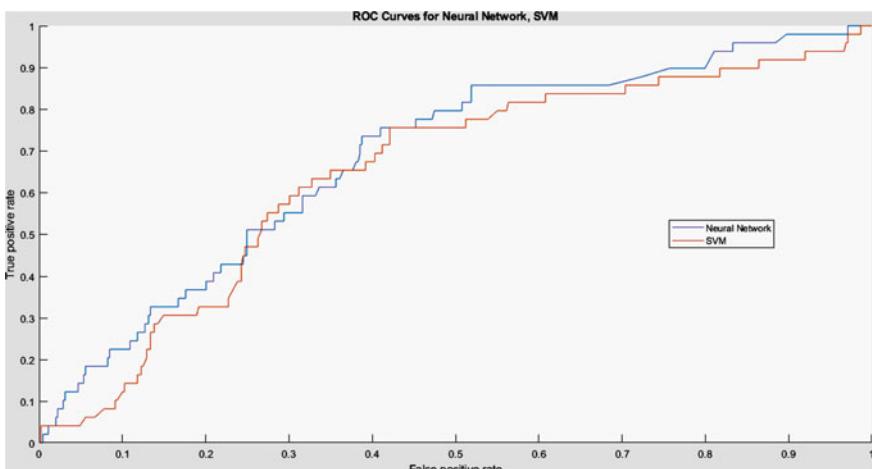
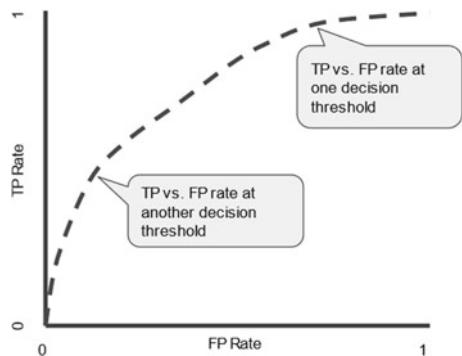


Fig. 4 Performance of ANN and SVM classification models

relationship among the features and the chances of occurrence of accident is better learnt by the ANN model. This prototype can be implemented for real-time systems in our physical world. It helps to detect the expected occurrences of road accidents in our physical world. With the increased accuracy, it can be approx. 90% accurate; hence, the rate of road accidents can be diminished to a large extent. In case it is not capable to detect the accident before time, even then this system is useful as it will help to accommodate the emergency services to the victim with ease and without delay. It is facilitated as we have mounted camera, speedometer, location tracer, pressure sensor which can recognize the accident case with location, and with GPRS, Wi-Fi module nearby ambulance is informed, and with routing algorithms, the shortest and fastest path is traced.

5 Conclusion

The entire chapter has brought a detailed discussion on the scenario of the road accidents in India: the causes, the impact and the IoT-VANET-based approach to reduce the rate of road accidents. The overall contribution of this chapter is to bring into notice of our intended readers that road accidents are causing hazardous loss to our country on daily basis. Along with that awareness, this chapter highlights that how the technological growth and advancement can be used to the betterment of human. IoT along with VANETs with the help of ML is used in this chapter to predict the plausibility of real-time accident occurrence. And, this perfect blend of technologies allows us to communicate among the vehicles, emergency services, RSUs and servers. One major contribution of this chapter is to devise OBUs which are an intelligent device to be equipped within the vehicles moving on the roads to constantly monitor the driver and the vehicles itself. Another important contribution is to provide a ML-based logic to predict the chances of road accidents in advance, means before it happens. So, there is an opportunity to avoid the collision and, hence, to reduce the road accident rate. In advance, nearby ambulance or emergency contact numbers are to be contacted. Further, with optimal routing algorithms, the fastest routes are followed to avoid all sorts of delays. In the future, this prototype can be enhanced and can be implemented as a real-time system for some commercial purposes. Finally, it will help us to design smart vehicles for the betterment of our lives and to assist the driver in the moving vehicles. Ultimately, it will save the property and lives which are always on risk while moving on the roads. While we will be building it as a commercial product, we will try to keep its financial cost feasible. In the future, the work can be extended to patent such a prototype with better accuracy.

References

1. Idris MYI, Abu Znaid AMA, Wahab AWA, Qabajeh LK, Mahdi OA (2017) Low communication cost (LCC) scheme for localizing mobile wireless sensor networks. *Wireless Netw* 23(3):737–747
2. Sadiq AS, Khan S, Ghafoor KZ, Guizani M, Mirjalili S (2018) Transmission power adaptation scheme for improving IoV awareness exploiting: evaluation weighted matrix based on piggybacked information. *Comput Netw* 137:147–159
3. Traffic accidents, NCRB 2016 Report, Chapter 1A: Traffic Accidents, Government of India
4. https://morth.nic.in/sites/default/files/Road_Accidendl.pdf
5. Olaru S, Weigh MC (2009) Vehicular Networks from theory to practice. *Vehicular Network Book*, 3rd ed. CRC Press Publishers
6. Zaldivar J, Calafate CT, Cano JC, Manzoni P (2011) Providing accident detection in vehicular networks through OBD-II devices and Androidbased Smartphones *5th IEEE Workshop On User Mobility and VEhicular Networks*. Bonn, Germany
7. Athavan K (2012) Automatic Ambulance rescue system. In: Second International Conference on Advanced Computing & Communication Technologies, pp 190–195
8. Koukoumidis RE, Peh LS, Martonosi M, Guru S (2011) Leveraging mobile phones for collaborative traffic signal schedule advisory. In: ACM mobsys11 Proceedings of the 9th international conference on Mobile systems, applications, and services, pp 127–140
9. Cai S, Becherif M, Bakhouya M, Wack M (2010) A Context Aware Embedded System for Intelligent Vehicles and Smart Roads *Proc. of ICPS, UPC Workshop*, 2010.
10. Praveen Kumar B, Mahendrakan K (2014) Prevention of accident due to drowsy by using eye blink. *IJIRSET* 3(4):12610–12616
11. Dharani M, Lalitha B (2013) Automatic Vehicle Control Using Sensors. *IJIRD* 2(3):744–750
12. Abboud M, Jaoude L, Kerbage Z (2004) Real time GPS navigation system. In: Proceedings of the 3rd FEA student conference, department of electrical and computer engineering, American University of Beirut, Citeseer
13. Chung Y (2010) Development of an accident duration prediction model on the Korean Freeway Systems. *Accid Anal Prev* 42(1):282–289
14. Zhao Y, Triantis K, Teodorovic D, Edara P (2010) A travel demand management strategy: The downtown space reservation system. *Eur J Oper Res* 205(3):584–594
15. Cheng HT, Shan H, Zhuang W (2011) Infotainment and road safety service support in vehicular networking: from a communication perspective. *Mech Syst Signal Process* 25(6):2020–2038
16. Zhao J, Cao G (2008) VADD: vehicle-assisted data delivery in vehicular Ad hoc networks. *IEEE Trans Veh Technol* 57(3):1910–1922
17. Luan TH, Ling X, Shen X (2012) Provisioning QoS controlled media access in vehicular to infrastructure communications. *Ad Hoc Netw* 100(2):231–242
18. Xu C, Zhao F, Guan J, Zhang H, Muntean G-M (2013) QoE-driven user-centric vod services in urban multihomed P2Pbased vehicular networks. *IEEE Trans Veh Technol* 62(5):2273–2289
19. Zhou L, Zhang Y, Song K, Jing W, Vasilakos AV (2011) Distributed media services in P2P-based vehicular networks. *IEEE Trans Veh Technol* 60(2):692–703
20. www.internet-of-things-book.com
21. Ghosh A, Chatterjee T, Samanta S, Aich J, Roy S (2017) Distracted Driving: A Novel Approach towards Accident Prevention. *Adv. Comput. Sci. Technol.* 10:2693–2705
22. Devi S, Neetha T (2017) Machine Learning based traffic congestion prediction in a IoT based Smart City. *Int. Res. J. Eng. Technol.* 4:3442–3445
23. Munoz-Organero M, Ruiz-Blaquez R, Sánchez-Fernández L (2018) Automatic detection of traffic lights, street crossings and urban roundabouts combining outlier detection and deep learning classification techniques based on GPS traces while driving. *Comput Environ Urban Syst* 66:1–8. [CrossRef]
24. Ryder B, Wortmann F (2017) Autonomously detecting and classifying traffic accident hotspots. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2017 ACM International Symposium on Wearable Computers, Maui, HI, USA, 11–15 September 2017; pp 365–370

25. Al Mamun MA; Puspo JA, Das AK (2017) An intelligent smartphone based approach using IoT for ensuring safe driving. In: Proceedings of the 2017 IEEE international conference on electrical engineering and computer science (ICECOS), Palembang, Indonesia, 22–23 August 2017; pp 217–223
26. Kubat M (2017) An Introduction to Machine Learning. Springer, Cham, Switzerland
27. Mohammed M, Khan MB, Bashier EBM (2016) Machine learning: algorithms and applications. CRC Press: Boca Raton. FL, USA
28. www.deeplearningbook.org
29. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. CVPR, pp 1 – 9
30. Zhang X, Yang Y-H, Han Z, Wang H (2013) Object class detection: asurvey 46(1):1–52
31. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. JCSS 55:119–139
32. Viola P, Jones M (2001) Viola-Jones object detection framework. Source:http://research.omicsgroup.org/index.php/Viola%20%93Jones_object_detection_framework Retrieve on 28th May 2017.
33. Viola J, Jones M (2001) Robust real-time object detection. In: Statistical and computational theories of vision—modeling, learning, computing, and sampling, pp 1 – 25
34. OpenCV 2.4.13.2 documentation OpenCV User Guide “Cascade Classifier Training. Source:http://docs.opencv.org/2.4/doc/user_guide/ug_traincascade.html Retrieve on 28th May, 2017.
35. Goyal S, Bhatia PK (2020) Comparison of machine learning techniques for software quality prediction. Int J Knowl Syst Sci (IJKSS) 11(2):21–40. IGI Global

FPGA Implementation of 16-Bit Wallace Multiplier Using HCA



G. Challa Ram, M. Venkata Subbarao, D. Girish Kumar,
and Sudheer Kumar Terlapu

Abstract This paper presents the FPGA implementation of modified structure of Wallace multiplier with Han-Carlson adder. As multipliers are prominent elements in the signal processing boards and microprocessing circuits, good multipliers are in need in terms of their delay and memory occupancy. Delay efficient multipliers can be constructed by adapting good adder for partial products addition. Han-Carlson adder which is a type of parallel prefix adder is used in place of ripple carry adder in traditional Wallace tree multiplier to improve the power levels and area usability of multiplier. The modified Wallace multiplier is coded in Verilog HDL and simulated using XILINX software. The performance of multiplier is validated on SPARTAN 3E FPGA kit.

Keywords Array multiplier · Han-Carlson adder (HCA) · Parallel prefix adder (PPA) · Ripple carry adder (RCA) · Wallace tree multiplier

1 Introduction

Over the past few decades, there is a rapid advancement in the fields of VLSI. A huge number of operations are performed over a single chip which includes for a high demand of components with rapid execution speed and low occupancy of space. As it is evident that multipliers are one of the important elements in many components like filters, DSP circuits and fast Fourier transforms [1]. A demand for multiplier with high performance in rose as traditional multipliers will tend to consume much power and poor execution speed. There are several types of multipliers like array multiplier, Dada multiplier, Vedic multiplier and Wallace multipliers. All these multipliers are designed to perform the multiplication operation, but their process of obtaining the partial products is different for different multipliers [2]. In traditional array multiplier, the multiplication is performed by multiplying each element of multiplier with their multiplicand which is the general arithmetic process in multiplication. All these

G. Challa Ram (✉) · M. Venkata Subbarao · D. Girish Kumar · S. K. Terlapu
Department of ECE, Shri Vishnu Engineering College for Women, Bhimavaram, India
e-mail: challaram.grandhi@svecw.edu.in

partial products are added by using simple ripple carry adder consisting of half and full adders.

As it is evident that the performance of the multiplier will lie on the choice of the adders used for addition of the products obtained in each stage. Array multiplier produces huge delay as it uses the RCA adder compared to other types of adders. While in Wallace multiplier, the process of multiplication is quite different to that of general multiplication process. In this Wallace tree multiplier [3], instead of adding two rows, the algorithm will tend to add the first three rows at once, and this addition of elements of three rows will be continued by considering the carries generated in the last stage addition. By following the technique of adding, these rows at a glance will reduce the delay in operation of multiplier. When there is a need of large bit multiplication, Wallace multiplier can be taken into account as it good delay efficiency.

Further, the speed of operation of Wallace multiplier can be improved by changing the adders used to add the partial elements produced in the multiplier [4]. The performance of the multiplier is based on the adders that has been used to add the partial products. Han-Carlson adder is used in the proposed modified structure of the Wallace multiplier. This adder will come under the category of parallel prefix adders (PPA's). In PPAs, the operation of addition has been done in three stages. As they use parallel addition technique, the speed and area usability can be improved. The modified Wallace multiplier with a Han-Carlson adder can be an efficient multiplier on the basis of speed and memory occupancy.

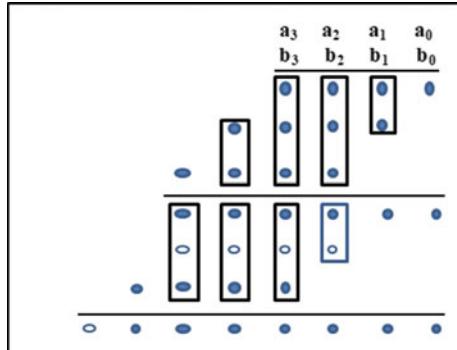
2 Array Multiplier and Wallace Multiplier

Array multiplier is simple one in which all the products are been obtained as in general arithmetic multiplication process. All the partial product elements are placed in an array. These partial elements are obtained by multiplying each bit of one input with the other, and the addition of these partial elements is done by using simple half and full adders [5]. A half adder is used for two elements and a full adder is used for three element addition. The carries generated in each addition are taken into account in next stage addition [6]. The performance of this multiplier is low when compared to other multiplier as the number of partial products is high and speed of operation is degraded.

Wallace multiplier is an efficient multiplier which is widely used in many applications as they offer good speed and area occupancy. The algorithm used for generation of partial product elements is completely different from the array multiplier which led to decrease in the delay of operation in the Wallace tree multiplier [7].

The operations performed in the Wallace multiplier are shown in Fig. 1 by considering 4-bit multiplication. The solid circle represents sum and circle represent carry [7]. The first-stage partial product elements are obtained by multiplying the first element of one input with all the elements of other input of multiplier, and this is repeated by shifting the element position. The first three rows of the products obtained

Fig. 1 Four-bit Wallace tree multiplier



are added at once unlike row by row addition in the array multiplier. The addition of three partial product rows will generate a row of partial sums and a row of carries these are further added to produce the final output [8]. Now these sum and carry row elements are added with another row of partial product elements, this process is repeated until only two rows are left [9, 10]. This addition of the elements is done by the basic full and half adders on the basis of requirement.

3 Proposed Multiplier

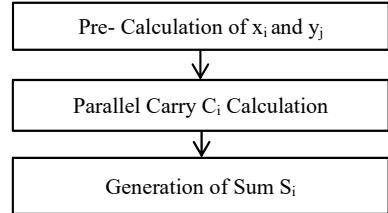
The proposed multiplier is modified structure of Wallace tree multiplier by using Han-Carlson adder. As the delay of a multiplier is sustained by the use of ripple carry adders, these adders are replaced by the Han-Carlson adders in order to decrease the delay of Wallace tree multiplier.

3.1 Han-Carlson Adder

Han-Carlson adder belongs to the family of parallel prefix adders (PPAs). There are several types of adders which perform addition on the basis of parallel prefix like Kogge stone adder, Brent Kung adder and Ladner Fischer adder [1]. These parallel prefix adders work in a three-stage process as shown in the Fig. 2.

The first stage is pre-calculation stage in which generate signal y_i and propagate signal x_i are generated for bits. In the second-stage parallel carry, calculations are performed. The operations in this stage are defined independently for every PPAs, and this stage is responsible for the reduction in the delay time [3]. The carries are generated by the help of black and grey cells in which a set of operations are performed on propagate and generate signals produced in the first stage. The equations of black are given by

Fig. 2 Stages in parallel prefix adders



$$Cx_{i:y} = x_{k-1:j} \text{ and } x_{i:k} \quad (1)$$

$$Cy_{i:j} = y_{i:k+1} \text{ or } (x_{i:k} \text{ and } y_{k-1:j}) \quad (2)$$

where $Cx_{i:y}$ and $Cy_{i:j}$ are the carry propagate and carry generate signals, respectively. Grey cell is used in order to calculate the carry generate $Cy_{i:j}$ signal and the equation is given by

$$Cy_{i:j} = y_{i:k+1} \text{ or } (x_{i:k} \text{ and } y_{k-1:j}) \quad (3)$$

Han-Carlson adder is a hybrid adder which is the combination of Brents Kung and Kogge stone adders. The second stage of the Han-Carlson adder consists of five states out of which the first is similar to Brents Kung adder, and the rest is like a Kogge's stone adder. The structure of 8-bit Han-Carlson adder is shown in Fig. 3.

3.2 Implementation of Wallace Multiplier with Han-Carlson Adder

In traditional Wallace multipliers, last stage partial products are added with half adders and full adders which cause a delay in execution [11, 12]. In the proposed Wallace multiplier, the last stage addition of partial products are done by Han-Carlson adder in place of theses half and full adders as shown in the Fig. 4. The whole structure of Wallace tree remains unchanged; the last two row elements are added by HCA adder to produce good execution speed in the multiplier [1].

Fig. 3 Stages in Han-Carlson adder with grey cell and black cells

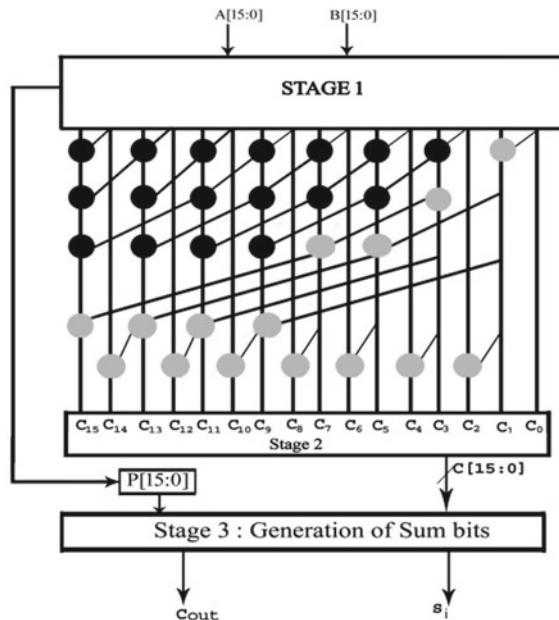
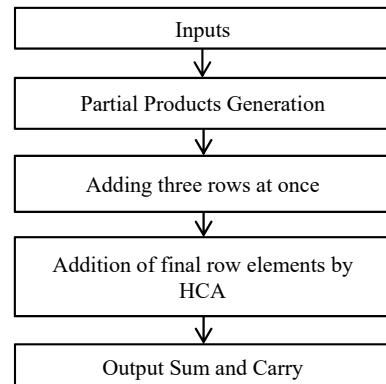


Fig. 4 Implementation of proposed Wallace tree multiplier with HCA



4 Results

4.1 Simulation Results

Results are obtained by using Xilinx Software and are implemented on FPGS device ec3s5005fg320 of SPARTAN 3E family.

The simulation results of a 4-bit Han-Carlson adder (HCA) is shown in Fig. 5. The two inputs are taken as 1010 and 1111.

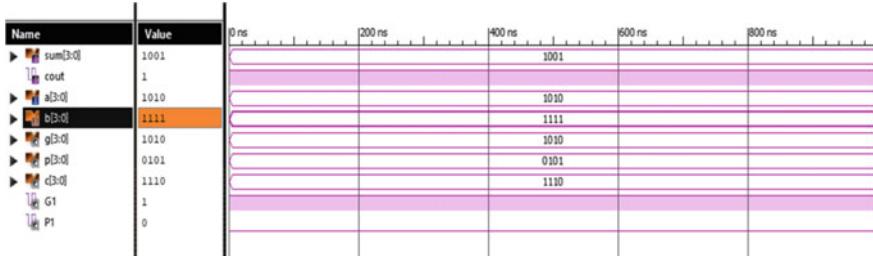


Fig. 5 Simulation results of 4-bit Han-Carlson adder (HCA)

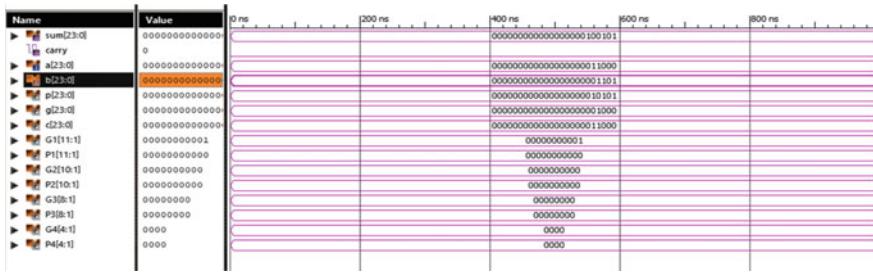


Fig. 6 Simulation results of 24-bit Han-Carlson adder (HCA)

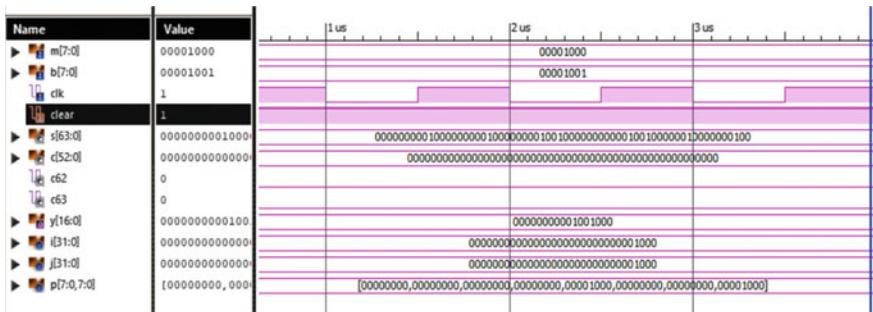


Fig. 7 Simulation results of 8-bit proposed Wallace multiplier using HCA

The simulation results of 24 bit HCA adder is shown in Fig. 6 and proposed 8-bit and 16-bit 8 Wallace multipliers using HCA are shown in Figs. 7 and 8, respectively.

4.2 Utilization Summary

The comparison between array, Wallace multiplier and proposed multiplier listed in the Tables 1 and 2, respectively.

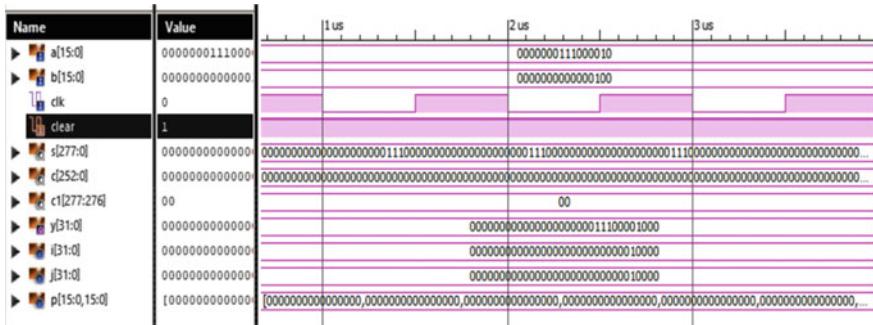


Fig. 8 Simulations results of 16-bit proposed Wallace multiplier using HCA

Table 1 Comparison between 4-bits, 8-bits and 24-bit Han-Carlson adder

Device Utilization	4-bit HCA	10-bit HCA	24-bit HCA
4-input LUT's	7	24	80
IOB's	13	31	73
Slice Registers	4	15	45
Delay in ns	8.79	14.77	24.02

Table 2 Comparison between 16-bit array, Wallace multiplier and proposed multiplier

Device utilization	Vedic multiplier	Wallace multiplier	Proposed multiplier
4-input LUT's	510	665	701
IOB's	66	66	66
Slice registers	257	360	379
Memory (kB)	454,891	453,246	452,756
Delay (ns)	58.26	42.08	36.12
Power (mW)	88.23	86.25	81.33

4.3 Hardware Implementation

The FPGA implementation of proposed multiplier is performed on SPARTAN 3E Kit. The hardware implementation of 4-bit Han-Carlson adder on FPGA SPARTAN 3E kit is shown in Fig. 9.

5 Conclusions

From the results, the proposed Wallace tree multiplier gives better enhancement in delay and area of usability than the traditional Wallace multiplier and array multiplier.



Fig. 9 FPGA implementation of 4-bit HCA on Spartan 3E Kit

The Wallace multiplier achieved a delay of 42.08 ns which is very low when compared with the delay of array multiplier. The delay for a 16-bit Wallace multiplier is given by 42.08 ns which is reduced to 36.12 ns. About 15 percent efficiency in delay is achieved by using the proposed structure with Han-Carlson's adder in place of ripple carry adders. The memory occupancy is also reduced by using the proposed design of Wallace multiplier.

References

1. Rao EJ, Ramanjaneyulu T, Kumar KJ (2018) Advanced multiplier design and implementation using Hancarlson adder. IEEE. 978-1-5386-6477-3
2. Raju A, Sa SK (2017) Design and performance analysis of multipliers using Kogge stone adder. In: International conference on applied and theoretical computing and communication technology. 978-1-5386-1144-9- 2017 IEEE
3. Ram GC, Rani DS, Balasaikesava B, Sindhuri KB (2016) Design of delay efficient modified 16 bit Wallace multiplier. In: IEEE international conference on recent trends in electronics information communication technology, May 20–21, 2016, India 978-1-5090-0774-5/16© 2016 IEEE
4. devi Ykuntam Y, Pavani K, Saladi K (2020) Design and analysis of high speed Wallace multiplier using parallel prefix adder for VLSI circuit design. 11th ICCCNT
5. Ram GC, Rani DS, Lakshmann YR, Sindhuri KB (2016) Area efficient modified vedic multiplier. International conference on circuit power and computing technologies. 978-1-5090-1276-3/16 ©2016 IEEE
6. Kesava RBS, Sindhuri KB, Rao BL, Kumar NU (2016) Low power and area efficient wallace tree multiplier using carry select adder with binary to excess-1 converter. In: 2016 Conference on advances in signal processing (CASP) cummins college of engineering for women, Pune. 978-1-5090-0849-0/16/\$31.00 ©2016 IEEE
7. Paradhasaradhi D, Prashanthi M, Vivek N (2014) Modified wallace tree multiplier using efficient square root carry select adder. (ICGCCEE/2014) 61:1–5. ISBN NO:978-1-4799-4982-3

8. Prakash E, Raju R, Varatharajan R (2013) Effective method for implementation of wallace tree using fast adders. *J Innov Res Solutions (JIRAS)* 1(1)
9. Ram GC, Rani DS, Balasakesava R, Sindhuri KB (2016) VLSI architecture for delay efficient 32-bit multiplier using vedic mathematic sutras. In: IEEE international conference on recent trends in electronics information communication technology. 978-1-5090-0774-5© 2016 IEEE
10. Vaidya S, Dandekar D (2010) Delay-power performance comparison of multipliers in vlsi circuit design. *Int J Comput Netw Commun (IJCNC)* 2(4):47–56
11. Weste N, Harris D (2004) CMOS VLSI design. Addison Wesley, Reading, MA
12. Krishna KG, Santhosh B, Sridhar V (2013) Design of Wallace tree multiplier using compressors. *Int J Eng Sci Res Technol.* ISSN: 2277-9655

A Comprehensive Study on Fruit Odour Detection and Classification Techniques Using eNose



Kalidindi Lakshmi Divya and V. Vijaya Baskar

Abstract This paper is an overview of various processes involved in the implementation of electronic nose for detection of fruit odour. Volatile gas detection techniques for fruits (especially mango) at different stages of ripening are given. These gases are detected by an array of sensors which are selected depending on the ability to detect the dominant gas component for the samples. Various pattern recognition algorithms along with different methods of training an artificial neural network (ANN) are stated. The major objective of the paper is to present a collective survey data on various procedures involved in determining the stages of ripeness in mango.

Keywords Volatile organic compounds · Gas chromatography · Sensors · Pattern recognition algorithms

1 Introduction

An electronic nose also termed as ‘Electronic olfactometry’ is a sensing device which is designed to detect odours. This is done by regenerating human senses using array of sensors and pattern recognition algorithms. It behaves in the similar way of the nose in identifying, comparing, quantifying along with data storage regarding the odours [1]. They are widely in use for applications in the field of food, medicinal drugs, issues related to crop yield in agriculture etc., to obtain accurate and fast results. The field of interest for this survey is detection of odours in mango fruit.

Any fruit in the nature acquires its natural odour due to the volatile compounds in it. Mango being the king of fruits behaves very differently for every species. Every species of this fruit has its own set of strong volatile organic compounds [2]. Mangoes undergo many changes physically and biochemically during various

K. Lakshmi Divya (✉) · V. Vijaya Baskar
Sathyabama University, Chennai, India
e-mail: kldivya@srkrec.ac.in

V. Vijaya Baskar
e-mail: vijayabaskar.ece@sathyabama.ac.in

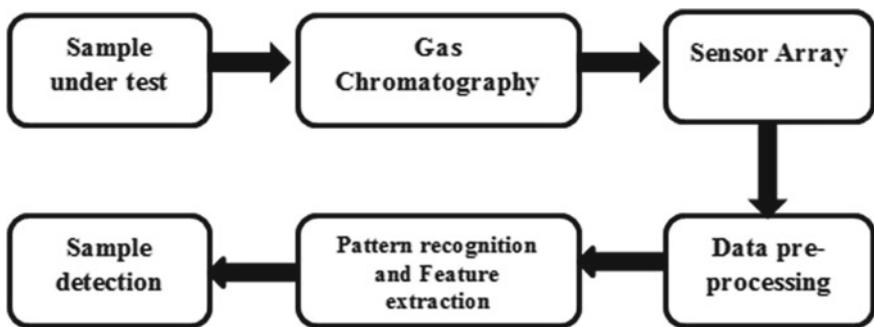


Fig. 1 Flow of odour detection in electronic nose

stages of ripening. The variating levels of volatile gases in various species of mango fruit depend on factors like pre-harvest treatment, crop maturity, process of crop handling, product storage post-harvest and processing. Majority compounds like terpene, acetone, hexane etc., are only emitted during ripening. For industrial and efficient odour analysis, we perform human sensory analysis either by chemical sensors or by gas chromatography. One of the most popular techniques of detecting biological volatile organic compounds is the gas chromatography–mass spectrometry (GC–MS) [3]. In gas chromatography, the sample under test is subjected to testing conditions for a period of time, and the release of volatile gases from it is sensed and examined.

The volatile organic compounds thus released are subjected to sensor array for detection which indeed changes the parameters of the sensor. A unique response is obtained by the electronic processing unit where a signal is transformed into a digital value. This data is fed to a pattern recognition system for detection (Fig. 1). The artificial neural networks are proved to be more efficient as they can handle the non-linear problems and reduce noise.

2 Detection of Volatile Organic Compounds

For test sample preparation, the selected species of mango has to be kept under a defined temperature for a period of time. The maturity of the fruit is determined by its colour, odour and hardness of the pulp. The sample is subjected to gas chromatography–mass spectrometry (GC–MS) setup [4], which is a method that combines the features of gas chromatography and mass spectrometry to detect various substances in the sample under test. Detection of these gases with utmost accuracy and sensitivity can be done by gas chromatography–mass spectrometry procedure. The gas chromatography–olfactometry (GC–O) technique is a union of traditional gas chromatographic analysis along with detection through sensors to study the complex

compositions of odorous substances. This technique can also be used in security purposes for detection of drugs or explosives.

There are wide range of volatile compounds in mangoes which differ in composition from one species to the other [5–7]. The compositions of these compounds change with changes in temperature, humidity, maturity and ripening stages [8, 9]. Some of the most prominent mango cultivars along with their dominant volatile gases are shown below. In most of the species, terpene hydrocarbons are the major volatile gases detected [10]. The concentrations of the volatiles also differ with their respective variety [11–13] (Table 1).

Table 1 Different species of mangoes with their dominant volatile gas and detection method

Species of mango	Strong volatile compound observed	Method used for detection
Williard (Srilankan mango) [10]	Terpinolene	Gas chromatography-flame ionization detector (GC-FID)
Karutha columban [10] Malgova [10]	Ocimene Myrcene	with Headspace–solid phase Micro-extraction (HS-SPME)
Jafna [6]	<i>cis</i> - β -ocimene	GC-MS
Parrot [6]	α -terpinolene	GC-MS
Tommy atkins [9, 11]	δ -3-Carene terpene	HS-SPME/GC-MS
Palmer [5]	δ -3-Carene terpene	
Espada [5]	α -terpinolene	
Carlota [5]	Myrcene	
Magovar (India) [5]	Hexamethyl cyclotrisiloxane	HS-SPME/GC-MS
Xiaofei' (Philippines) [8]	β -Myrcene	GC-MS
Guiya' (China) [8]	γ -Octanoic lactone	GC-MS
Haden, Smith, Florida, San Diego [12]	delta-3-carene	distillation–extraction
Keitt, and Kent, Manga blanca [12, 13]		GC
Manga amarilla, Manzano [7]		GC-MS
Delicioso, Super Haden [7]	Limonene	Distillation–extraction
Ordóñez, Filipino [7]		GC
La Paz [7]		GC-MS
Obispo [7]	Terpinolene	Distillation–extraction
Corazon [8]		GC
Huevo de toro [8]		GC-MS
Minin [8]	Alpha-phellandrene	Distillation–extraction, GC, and GC-MS

Table 2 Types of sensors, their ranges, applications and data analysis methods

Type of sensor	Range	Field of application	Suitable pattern recognition system
Metal oxide [18–20]	5–500 ppm	Food and beverage industry	PCA, ANN, DA, REGRESSION
Conducting Polymer [22]	0.1–100 ppm	Environmental parameter testing	ANN
Quartz crystal Microbalance [17]	1.5 Hz/ppm	Food analysis and lung diagnosis	ANN,PCA
Acoustic wave [21]	100–400 MHz	Quality control in food and automobile applications	PCA
Catalytic [21]	Large scale	Environmental monitoring	PCA
Electrochemical [21]	0–1000 ppm	Industrial and medical applications	Pattern recognition methods
Optical [23, 24]	DL (NH ₃)1 ppm with polyaniline coating	Biomedical applications	PCA
Metal oxide semiconductor field effect transistor [23, 24]	2.8 mV/ppm for toluene	Food industry: flavour and fragrance testing	ANN, DC, PCA

3 Sensors

The volatile organic compounds obtained by the procedure of gas chromatography cannot alone differentiate the fruit. An array of sensors are used where each sensor senses different types of aroma. Gas sensors produce finger print response to each individual volatiles which is further utilized by pattern recognition algorithms for classification [14]. Various factors that influence the selection of sensor array are their size, re-usability, fast response, sensitivity towards gas compound detection and resistivity to temperature changes [15, 16]. Depending upon the field of application, Quartz crystal microbalance sensors [17], metal oxide sensors [18–20], electrochemical sensors [21], conducting polymers [22], acoustic sensors, catalytic sensors, optical and MOSFET (metal oxide semiconductor field effect transistor) sensors [23, 24] can be described in Table 2.

4 Pattern Recognition and Feature Extraction

Pattern recognition is the process where individual patterns are identified using machine learning algorithms. Machine learning is one of the most popular mediums used for training the computer to obtain the desired output [25]. There are built-in

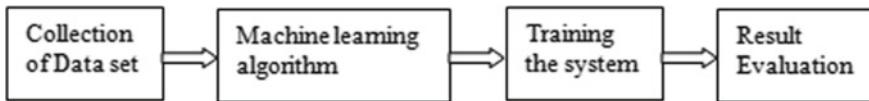


Fig. 2 Process flow of pattern recognition system

basic algorithms which can be used to classify and identify the attributes of the input data (Fig. 2). They are grouped into supervised learning, un-supervised learning, semi-supervised learning and reinforcement learning based on the nature of the input and output labels of the dataset.

Most prominently used tools for data analysis are linear discriminant analysis (LDA) aiming at dimensionality reduction, discriminant function analysis (DFA) for classification of known input labels, support vector machines (SVM's) [26–28] for linear reduction of dataset, principal component analysis (PCA) for predicting output of a response variable, stepwise discriminant analysis (SDA) for stepwise discrimination [30], generalized least squares regression (GLSR) on coefficients of linear regression partial least squares regression (PSR) for dimensionality reduction, multiple linear regression (MLR) [31] for determining a best fit hyper plane and artificial neural networks (ANNs) [32, 33]. Many experiments have been conducted on different fruits to estimate the efficient results in ripening stages.

In detecting the stages of ripeness of a fruit, supervised classifiers, such as fuzzy ARTMAP (adaptive resonance theory neural networks), LVQ (learning vector quantization) and MLP (Multilayer perceptron neural network) are used whose accuracy is more compared to back propagation algorithm [34]. Feature extraction of mango is proven performance efficient by implementation through grey-level co-occurrence matrix and classification of variety with multiclass SVM (support vector machine) and K-means clustering [35]. Clustering methods like fuzzy-C-means, PCA and KNN (K-nearest neighbours) are proved to improve the performance of pattern recognition using multiple SOM's (self-organized machines).

5 Applications

Electronic nose has wide range of applications in every field [36]. In food and beverage industry, they are efficient in quality assessment, chemical detection and shelf life [37–39]. In agriculture and forestry, it helps analyse utilization of resources and reduce wastage [40]. In medical field, they are very helpful in analysing diseases like urinary tract infections, asthma prediction, cystic fibrosis etc., [41]. They are also efficiently used in environmental parameter estimation and security systems [42]. Besides their vast range of applications, they have drawbacks like sensor drift, limited sensor sensitivity and calibration errors [43].

6 Conclusion

The human olfactory system is quite difficult and complex to replicate exactly. But an electronic nose has been a revolutionary invention in analysing a sample and obtaining sensitive, accurate, low cost and fast results.

From the study of the survey reports of various species of mangoes, it has been evident that there are efficiency in results by using gas chromatography–mass spectrometry (GC–MS) in detecting VOCs as their results are sensitive to the performance of sensors. Metal oxide sensor arrays are widely used for detection in food industries. The wide detection range of sensors encourages the future scope of developing robust nano-sensors for improved efficiency. Further, we can implement different pattern recognition algorithms to obtain a unique dataset with high end of detailing which is further used for training a system in detecting fruit ripening stage.

References

1. Karakaya D, Ulucan O, Turkan M (2020) Electronic nose and its applications: a survey. *Int J Autom Comput* 17(2):179–209
2. Brezmes J, Fructuoso ML, Llobet E, Vilanova X, Recasens I, Orts J et al (2005) Evaluation of an electronic nose to assess fruit ripeness. *IEEE Sens J* 5(1):97–108
3. Tang KT, Chiu SW, Pan CH, Hsieh HY, Liang YS, Liu SC (2010) Development of a portable electronic nose system for the detection and classification of fruity odors. *Sensors* 10(10):9179–9193
4. Brattoli M, Cisternino E, Dambruoso PR, De Gennaro G, Giungato P, Mazzone A et al (2013) Gas chromatography analysis with olfactometric detection (GC-O) as a useful methodology for chemical characterization of odorous compounds. *Sensors* 13(12):16759–16800
5. Mesquita PR, Pena LC, Santos FND, Oliveira CCD, Magalhães-Junior JT, Nascimento AS, Rodrigues FM (2020) Mango (*Mangifera indica*) aroma discriminate cultivars and ripeness stages. *J Braz Chem Soc* 31(7):1424–1433
6. MacLeod AJ, Pieris NM (1984) Comparison of the volatile components of some mango cultivars. *Phytochemistry* 23(2):361–366
7. Pino JA, Mesa J, Muñoz Y, Martí MP, Marbot R (2005) Volatile components from mango (*Mangifera indica* L.) cultivars. *J Agric Food Chem* 53(6):2213–2223
8. Hossain M, Rana M, Kimura Y, Roslan HA (2014) Changes in biochemical characteristics and activities of ripening associated enzymes in mango fruit during the storage at different temperatures. *BioMed Res Int* (2014)
9. Li Z, Wang N, Raghavan GV, Vigneault C (2009) Ripeness and rot evaluation of ‘Tommy Atkins’ mango fruit through volatiles detection. *J Food Eng* 91(2):319–324
10. Thiruchelvam T, Landahl S, Terry LA (2020) Temporal variation of volatile compounds from Sri Lankan mango (*Mangifera indica* L.) fruit during ripening. *J Agric Food Res* 2:100053
11. White IR, Blake RS, Taylor AJ, Monks PS (2016) Metabolite profiling of the ripening of Mangoes *Mangifera indica* L. cv. ‘Tommy Atkins’ by real-time measurement of volatile organic compounds. *Metabolomics* 12(3):57
12. Liu H, An K, Su S, Yu Y, Wu J, Xiao G, Xu Y (2020) Aromatic characterization of mangoes (*Mangifera indica* L.) using solid phase extraction coupled with gas chromatography–mass spectrometry and olfactometry and sensory analyses. *Foods* 9(1):75
13. Slaughter DC (2009) Nondestructive maturity assessment methods for mango. University of California, Davis, pp 1–18

14. Nouri FG, Chen Z, Maqbool M (2014) Monitoring mango fruit ripening after harvest using electronic nose (zNoseTM) technique. In: 5th international conference food engineering biotechnology, vol 65, p 8
15. Arshak K, Moore E, Lyons GM, Harris J, Clifford S (2004) A review of gas sensors employed in electronic nose applications. *Sens Rev*
16. Patel HK, Kunpara MJ (2011) Electronic nose sensor response and qualitative review of e-nose sensors. In: 2011 Nirma University international conference on engineering. IEEE, pp. 1–6
17. Tozlu BH, Okumuş Hı, Şimşek C. Selecting suitable sensor on building an electronic nose
18. Banerjee MB, Pradhan S, Roy RB, Tudu B, Das DK, Bandyopadhyay R, Pramanik P (2018) Detection of benzene and volatile aromatic compounds by molecularly imprinted polymer-coated quartz crystal microbalance sensor. *IEEE Sens J* 19(3):885–892
19. Wang C, Yin L, Zhang L, Xiang D, Gao R (2010) Metal oxide gas sensors: sensitivity and influencing factors. *Sensors* 10(3):2088–2106
20. Berma A (2010) Metal oxide sensors for electronic noses and their application to food analysis. *Sensors* 10(4):3882–3910
21. Romain AC, Nicolas J (2010) Long term stability of metal oxide-based gas sensors for e-nose environmental applications: an overview. *Sens Actuators, B Chem* 146(2):502–506
22. Wojnowski W, Majchrzak T, Dymerski T, Gębicki J, Namieśnik J (2017) Portable electronic nose based on electrochemical sensors for food quality assessment. *Sensors* 17(12):2715
23. Pelosi P, Zhu J, Knoll W (2018) From gas sensors to biomimetic artificial noses. *Chemosensors* 6(3):32
24. Gardner JW, Hines EL, Tang HC (1992) Detection of vapours and odours from a multisensor array using pattern-recognition techniques. Part 2. Artificial neural networks. *Sens Actuators B: Chem* 9(1):9–15
25. Sayago I, Aleixandre M, Santos JP (2019) Development of tin oxide-based nanosensors for electronic nose environmental applications. *Biosensors* 9(1):21
26. Fu J, Li G, Qin Y, Freeman WJ (2007) A pattern recognition method for electronic noses based on an olfactory neural network. *Sens Actuators, B Chem* 125(2):489–497
27. Behera SK, Sangita S, Rath AK, Sethy PK (2019) Automatic classification of mango using statistical feature and SVM. In: Advances in computer, communication and control. Springer, Singapore, pp 469–475
28. Goudjil M, Koudil M, Bedda M, Ghoggali N (2018) A novel active learning method using SVM for text classification. *Int J Autom Comput* 15(3):290–298
29. Burges CJ (1998) A tutorial on support vector machines for pattern recognition. *Data Min Knowl Disc* 2(2):121–167
30. Mokeev AV, Mokeev VV (2015) Pattern recognition by means of linear discriminant analysis and the principal components analysis. *Pattern Recognit Image Anal* 25(4):685–691
31. Rudas T (1984) Stepwise discriminant analysis procedure for categorical variable. In: Compstat 1984. Physica, Heidelberg, pp 389–394
32. Uyanık GK, Güler N (2013) A study on multiple linear regression analysis. *Procedia Soc Behav Sci* 106:234–240
33. Tan J, Kerr WL (2018) Determining degree of roasting in cocoa beans by artificial neural network (ANN)-based electronic nose system and gas chromatography/mass spectrometry (GC/MS). *J Sci Food Agric* 98(10):3851–3859
34. Qi PF, Meng QH, Zeng M (2017) A CNN-based simplified data processing method for electronic noses. In: 2017 ISOCS/IEEE international symposium on olfaction and electronic nose (ISOEN). IEEE, pp 1–3
35. Llobet E, Hines EL, Gardner JW, Franco S (1999) Non-destructive banana ripeness determination using a neural network-based electronic nose. *Meas Sci Technol* 10(6):538
36. Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *IEEE Trans Autom Control* 42(10):1482–1484
37. Baitello M, Wilson AD (2015) Electronic-nose applications for fruit identification, ripeness and quality grading. *Sensors* 15(1):899–931

38. Chilo J, Pelegri-Sebastia J, Cupane M, Sogorb T (2016) E-nose application to food industry production. *IEEE Instrum Meas Mag* 19(1):27–33
39. Siadat M, Losson E, Ghasemi-Varnamkhasti M, Mohtasebi SS (2014) Application of electronic nose to beer recognition using supervised artificial neural networks. In: 2014 International conference on control, decision and information technologies (CoDIT). IEEE, pp 640–645
40. Dhanekar S (2020) Smart and intelligent E-nose for sensitive and selective chemical sensing applications. *Smart Sens Environ Med Appl* 149–171
41. Wilson AD (2013) Diverse applications of electronic-nose technologies in agriculture and forestry. *Sensors* 13(2):2295–2348
42. Thaler ER, Kennedy DW, Hanson CW (2001) Medical applications of electronic nose technology: review of current status. *Am J Rhinol* 15(5):291–295
43. Capelli L, Sironi S, Del Rosso R (2014) Electronic noses for environmental monitoring applications. *Sensors* 14(11):19979–20007

Multivariate Time Series Analysis for Predicting Number of Disease Cases



Ramatulasi Tammineni and G. N. V. G. Sirisha

Abstract Dengue is pandemic disease, which is spread by mosquitoes. This disease can become lethal if not treated in time. The early prediction of dengue disease spread based on demographic features and climate conditions is helpful for the public health care department authorities. It helps in enhancing facilities in public health care system like increasing number of beds in hospitals, blood banks, spraying mosquito repellents in the surrounding areas, stocking the required medicines, etc. Machine Learning can help in developing models that can predict number of dengue cases based on historical data. Five machine learning algorithms are used to predict number of dengue cases in this paper. Those are Autoregressive Integrated Moving Average, Negative Binomial Regression, Multilayer Perceptron, XGBoost, and Linear Regression. These machine learning algorithms are applied on dataset from San Juan and Iquitos cities. Grid search is applied to tune the hyper parameters and L1 and L2 regularization are applied to avoid over fitting. Among these algorithms, linear regression algorithm with a mean absolute error of 7.969 gives the best performance for predicting number of dengue disease cases based on demographic and climatic features.

Keywords Pandemic disease · Mosquito · Dengue

1 Introduction

In India and several developing countries with large populations, healthcare is one of the main challenges to be faced. The medical resources provided by the government cannot cope with the large population. Rural areas are most affected by the lack of adequate medical infrastructure in the public health sector. The Indian healthcare system divides the population into two groups: the wealthy, and the impoverished [1].

R. Tammineni (✉) · G. N. V. G. Sirisha

Department of CSE, S.R.K.R. Engineering College, Bhimavaram, Andhra Pradesh, India

G. N. V. G. Sirisha

e-mail: sirishagadiraju@srkrec.ac.in

The first group can easily enter the private healthcare sector, while the second group has difficulty using it. Therefore, the public health department has been introduced to promote the second group of medical resources and diagnosis.

Dengue Hemorrhagic Fever (DHF) is a contagious disease caused by a virus and spread by the “Egypt” mosquito, which thrives and spreads the virus in tropical and subtropical climates. There are four serotypes of dengue virus (DENV-1 to DENV-4) [2]. Dengue fever can range in severity from a simple high fever to a severe hemorrhagic fever. The fatality rate from dengue sickness is increased due to delayed recognition of the first symptoms of dengue hemorrhagic fever. If it is not resolved as soon as possible, this disease can take the lives of patients. Severe headache, joint discomfort, skin rash, leucopenia, thrombocytopenia, and muscle soreness are all symptoms of dengue fever [3]. In some instances, the infection might evolve to aggressive dengue hemorrhagic fever (dengue hemorrhagic fever), which causes a reduction in platelet count, plasma loss, and possibly dengue shock.

2 Problem and Significance

Dengue fever is an epidemic that can kill infected people. Changes in weather conditions can affect the spread of diseases. Early detection of the disease will help the government take the necessary steps to help prevent the spread of the disease. Dengue fever is very prevalent in India and the mortality rate is also high. Detecting dengue fever cases manually is time-consuming and cumbersome. Therefore, there is a need to develop a system that can automatically predict the number of dengue fever cases that may occur in a location based on historical data.

3 Review of Literature

Chovatiya et al. have used long short-term memory (LSTM) recurrent neural network for dengue prediction [1]. The prediction accuracy of the proposed algorithm is 84%.

Manivannan and Isakki Devi have used clustering techniques for visualization of diseases [2]. The suggested method employs the k-means to cluster households by age group using dengue fever serotypes. This data set originates in Vietnam’s Ho Chi Minh City. There were 1910 records and 171 attributes in the data set.

Nakvisut and Phienthrakul have used two-step technique for prediction of dengue cases [3]. In the two-step prediction model, the first step is used to classify the data. Support vector machine regression time series prediction model is suitable for this application. Judging from the results of the two-step forecasting technique; the forecasted data is close to the real data.

Sasongko et al. have used ANN techniques for predicting number of dengue cases [4]. Techniques like Gradient Descent (GD), Levenberg Marquardt (LM), and Resilient Backpropagation (RB), are applied. In these techniques, Lavenberg

Marquardt algorithm gave better precision. Lavenberg Marquardt is used to detect dengue early. The algorithm Lavenberg Marquardt is applied on Indonesian data sets. The data is divided into training data and test data, and the amount of training data is 60%, and the test data are 40%. The proposed algorithm accuracy is 89.28%.

Nandana et al. have used the DBSCAN algorithm in predicting disease hot spot [5]. The algorithm relies heavily on the concept of density-based clustering. It aims to identify groups with different shapes and patterns. The data set is taken from Delhi from 2006 to 2015.

Tsoumakashave used a large amount of heterogeneous retail time series data collected from shopping mall from 2014 to 2017 [6]. A total of 1105 days of data set were collected, and the first 1000 days were used as the training set, and the last 105 days were used as the test set. The EMD decomposition algorithm that uses divide and conquer approach to obtain sub-sequence with different amounts of information in the source sequence and a deep learning model for each sub-time sequence is created. Then, the result of the prediction is obtained by adding the results of the posterior prediction.

Basuki and Harsono have done a comparative analysis of montecarlo linear and dynamic polynomial regression for predicting dengue cases [7]. It is determined by calculating and comparing the variance between the expected and actual number of cases. The best regression methodology for predicting dengue fever cases was determined using a mean square error (MSE) test. When these two methods are evaluated, dynamic polynomial regression produces the lower MSE when compared to polynomial linear regression. As a result, when compared to linear Monte Carlo regression, dynamic polynomial regression can predict well.

Anggraeni et al. concludes that Autoregressive Integrated Moving Average (ARIMA) model performance can be improved by the addition of Google Trends to it in forecasting number of dengue fever cases. [8]. This dataset is collected in Indonesia. However, the search query selection method in Google Trends needs to be refined in future research.

Bikku has used Multilayer Perceptron (MLP) to predict the number of disease cases [9]. The data set for this study was acquired from the UCI-ML database. They employed three standard data sets, each corresponding to a different disease. The Wisconsin Breast Cancer (WBC) data set has nine different features, 500 training samples, and 200 test samples. The second data set is the SaHeart (SHT) data set. It contains 9 characteristics, 300 training samples, the test samples are close to 160, and the third data set is “American Indian Diabetes (PID)”. There are 8 features in all, with 576 training samples. A total of 192 test samples are available. This article focuses on deep learning-based technology that can envisage diseases using medical data from the past. MLP has the highest accuracy among RNN, LSTM, and MLP.

Anggraeni et al. have used Fuzzy-Logistic Regression [10]. Data collected from Puskesmas in the Malang Regency. The data was from Malang Regency Public Health Office. From January 2010 through December 2018, the data was collected. The model’s prediction accuracy is 79.93%.

4 Methodology

Based on the person's age, gender, and climatic characteristics, a machine learning-based algorithm was built to forecast the frequency of dengue fever cases. Architecture of system for predicting no. of dengue cases is shown in Fig. 1.

4.1 Input

The data set used in this study comes from Kaggle. The data set belongs to the cities of San Juan and Iquitos. 80% of the data set is used to train machine learning models. 20% of the data set is used for testing. The data set contains 1456 tuples. 936 tuples belong to San Juan and 520 tuples to Iquitos. The test data contains 416 tuples. The training and test data sets are described by 24 variables. The predictable variable is the number of dengue cases per week for every combination of city, year, week of year. Data were collected from two different cities: San Juan (sj) in Puerto Rico and Iquitos (iq) in Peru.

Preprocessing

Data Pre-treatment is a very important step of learning the machine. Most real world data we get problematic, so we train this data before feeding on our automatic learning model.

Steps for preprocessing the data

- Removing redundancy features
- Missing values imputation
- Vectorizing Categorical data

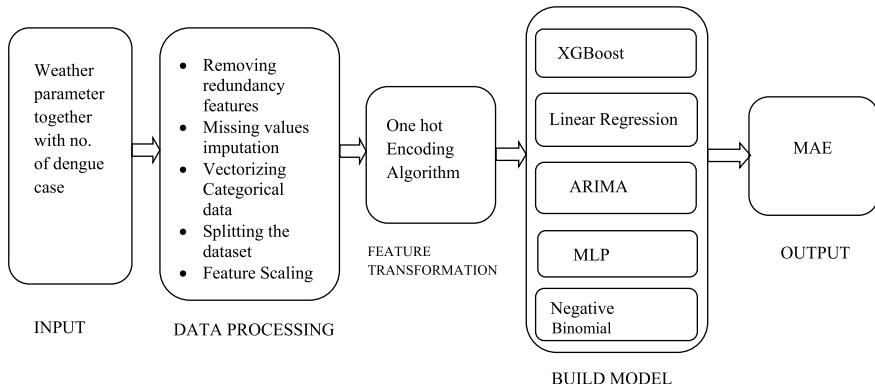


Fig. 1 Architecture of system for predicting no. of dengue cases

- Splitting the dataset
- Feature Scaling.

Removing redundancy features: Our data set has several missing values. With missing values we cannot train the model. So we need to fill the missing values with mean or median.

Replacing missing values with median value: To calculate median value sort the values in ascending order. If the values are odd, then take the middle value as missing value. If the values are even then take the average of middle two values as missing value.

Vectorizing Categorical data: Machine learning model works with numbers, so the model cannot understand the categorical data. Categorical data need to be encoded. One hot encoding is used for this purpose.

Removing unwanted data in dataset: For improving prediction accuracy some highly correlated features (i.e., absolute correlation > 0.9) are removed from the dataset. Those are ‘reanalysis_sat_precip_amt_mm’, ‘reanalysis_specific_humidity_g_per_kg’, ‘reanalysis_avg_temp_k’, ‘reanalysis_tdtr_k’ (Fig. 2).

Splitting the Dataset: Dataset is split into training set and test set. A part of the data set is used to train the model. Another part of the data set will be used to evaluate our model. 70% of the data set is used for training and 30% is used for testing.

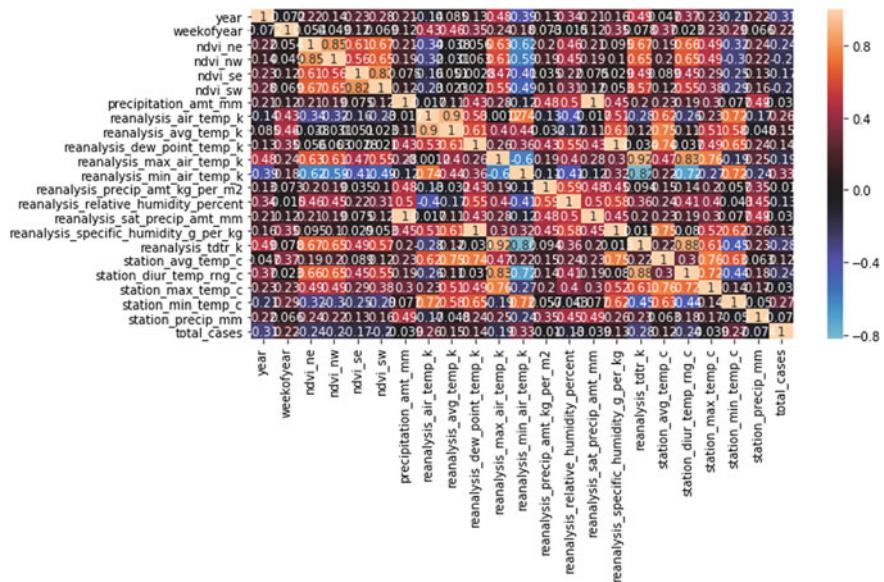


Fig. 2 Correlation matrix

Feature Scaling: This is the final stage in the data preparation procedure. All of our data is placed on the same range and scale using the element scale.

Correlation matrix: Correlation coefficients for several variables are displayed in a correlation matrix. The correlation between all possible pairings of values in a table is represented by the matrix.

Building the Model: The machine learning models are built using ARIMA, MLP, Linear Regression, Negative Binomial Regression, and XGBoost. The details of these algorithms are described in Sect. 5.

5 Algorithms Used in This Study

5.1 XGBoost

XGBoost is short for “eXtreme Gradient Boosting”. “eXtreme” refers to speed enhancements, such as parallel computing and cache recognition, which make XGBoost approximately 10 times faster than traditional gradient gain. Additionally, XGBoost includes a unique split search algorithm to optimize the tree and a built-in regularization function to reduce over fitting. The advantage of XGBoost is that less feature engineering is required, feature importance can be found, fast interpretation, outliers have minimal impact, large data sets can be handled well, good execution speed, good model performance, and not too easy to over-tune. The disadvantages of XGBoost are that it is difficult to interpret and visualize, if the parameters are not set correctly it can be over fit and it is difficult to fit due to too many hyper parameters.

5.2 Linear Regression

Linear regression is used to model the data if there is linear relationship between response and predictor variables. The term “linear” in linear regression refers to the fact that the method uses linear combinations of explanatory / predictor variables to model the data. The advantages of linear regression are the simple model, high computational efficiency, and the interpretability of the output. The disadvantage of linear regression is that it is too simple, linear assumptions are severely affected by outliers, the independence and homogeneity of variables, and the importance of characteristics cannot be determined.

5.3 Negative Binomial Regression

As the name suggests, Negative Binomial Regression is used to predict continuous variable value. It is generally preferred over Poisson regression for datasets where mean and variance of the dataset are not equal. This model is established as an effective method for predicting time series of counts.

5.4 Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is a feed forward neural network having at least one input layer, one or more hidden layers, and one output layer, each of which serves a distinct purpose. Many neurons with firing functions are present in each hidden and output layer. Due to the short amount of the data set, MLP was employed instead of DNN in this study. By carefully designing the neural network architecture, such as selecting the activation function or the number of hidden layers and neurons in each layer, it may be used to describe the complex non-linear relationship between input variables and results. Since there is no uniform method for determining the sum of neurons and layers in a neural network, the best design is usually determined through experiments or previous experience with parallel problems.

5.5 ARIMA

Automatic Regression Integrated Moving Average (ARIMA) is a forecasting model used to analyze and predict the time series. The impact of periodic, seasonal and irregular events on the data item is being measured. The data series used by ARIMA must be fixed data series. The advantage is a better understanding of time series patterns and ARIMA forecasts. The disadvantage is that it only captures the linear relationship between the data.

6 Results

This study used the data set from San Juan and Iquitos cities. 936 tuples are from San Juan and 520 tuples are from Iquitos. In this study, five types of machine learning algorithms were used to predict the number of dengue cases. Since, the target attribute being predicted is a continuous attribute; Mean Absolute Error (MAE) is used to assess the performance of models used in this paper.

The first algorithm used is the ARIMA model. Here, the output MAE of ARIMA is 60.51. The second algorithm used is MLP. The parameters used in MLP are batch

Table 1 Comparison of different algorithms

S. No.	Method	Mean absolute error (MAE)
1	ARIMA	60.51
2	MLP	19.36
3	Negative binomial regression	17.41
4	XG boost regression	12.44
5	Linear regression	7.969

size = 25, epoch = 1000. MAE of MLP is 19.36. The third algorithm used is the XGBoost regression algorithm. This is applied to a data set with default parameters. The MAE of XGBoost is 17.52.

After that the grid search is applied to the dataset to improve the output of the XGBoost algorithm. Grid search helps in identifying the best parameters. The best hyper parameter values of XGBoost found after applying grid search are gamma “0”, lambda “1”, learning rate 0.03,max_depth 5, minimum sub-item weight 2, n_estimators 50, n_thread 4, target reg"linear", silent 1, subsample 0.7. When XGBoost model is developed using these hyperparameters the MAE is decreased to 12.44. The fourth algorithm used is negative binomial regression. A negative binomial regression is applied to the data set, and then the MAE value reaches 17.41. The fifth algorithm used is the linear regression model. The MAE of linear regression model is 11.53. To improve the performance of the linear regression, L1 regularization and L2 regularization are used, and then the MAE decreased to 7.969.

This study has shown that linear regression is best in predicting number of dengue cases by taking demographic and climatic features as independent variables (Table 1).

7 Conclusion

Dengue fever is an epidemic spread by mosquitoes. The government must prepare for the emergence of dengue fever disease to save people's lives by providing medicines and other medical facilities. Availability of a model that can predict the number of dengue cases based on demographic and climatic features can assist the health care providers in upgrading the medical facilities. Since the number of features in dataset is large, correlation analysis is used to remove the redundant features. In this study, five different machine learning approaches were used to predict the number of dengue cases which are ARIMA, MLP, Negative binomial regression, XGBoost Regression, and Linear Regression. Grid search is applied to tune the hyper parameters and L1 and L2 regularization are applied to avoid over fitting. Linear regression outperformed the rest of the techniques with mean absolute error of 7.969.

References

1. Chovatiya M, Dhameliya A, Deokar J, Gonsalves J, Mathur A (2019) Prediction of dengue using recurrent neural network. In: Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), IEEE Xplore, pp 256–259
2. Nandana GM, Mala S, Rawat A (2019) Hotspot detection of dengue fever outbreaks using DBSCAN algorithm. *IEEE Trans* 158–161
3. Nakvisut A, Phienthrakul T (2018) Two-step prediction technique for dengue outbreak in Thailand. In: IEECON (2018)
4. Sasongko PS, Wibawa HA, Maulana F, Bahtiar N (2017) Performance comparison of artificial neural network models for dengue fever disease detection. In: 1st international conference on informatics and computational sciences (ICICoS), pp 183–188
5. Manivannan P, Isakkki Devi P (2017) Dengue fever prediction using K-means clustering algorithm. IEEE international conference on intelligent techniques in control, optimization and signal processing
6. Tsoumakas G (2018) A survey of machine learning techniques for food sales prediction. Published online: 14 June 2018. Springer Nature B.V, pp 441–447
7. Roziqin MC, Basuki A, Harsono T (2016) A comparison of Montecarlo linear and dynamic polynomial regression in predicting dengue fever case. In: Knowledge creation and intelligent computing (KCIC), IEEE conference, pp 213–218
8. Anggraeni W, Pramudita G, Riksakomara E, Samopa F, Radityo PW (2018) Artificial neural network for health data forecasting, case study: number of dengue hemorrhagic fever cases in Malang Regency, Indonesia. In: IEEE Conference, pp 207–212
9. Bikku T (2020) Multi-layered deep learning perceptron approach for health risk prediction. *J Big Data*, pp 1–14 (2020)
10. Anggraeni W, Sumpeno S, Yuniarno EM (2020) Prediction of dengue fever outbreak based on climate factors using fuzzy-logistic regression. In: IEEE 2020 international seminar on intelligent technology and its applications (ISITIA)

Difference Differential Op-Amp Based CMOS Adiabatic All Pass Filter



P. Mullangi and John Blessy Mandru

Abstract Research on active filters has multiplied. Active Filters are a form of a digital circuit that makes use of active components like an Amplifier. Amplifiers engage with a channel configuration can work at the channel's presentation and consistency, where we attempt no longer to utilize segments like inductors that are marginally highly-priced than different active and uninvolved parts. An enhancer forestalls the heap impedance of the accompanying degree from influencing the features of the channel. This paper provides a survey of the all pass Filters. In no time term, all pass Filters could have a load's more tremendous diploma than it has now. The properties of superior all pass filters are inspected, and a huge define of the sort of uses in automated keeping apart is given. A distinctive structure gratifying the all-skip property are amassed, accentuating number one misfortune lessness. Applications are then specified in score separating, vital sifting and channel banks, multi-fee keeping apart, range and accumulating defer balance, and Hilbert adjustments. In all cases, the underlying misfortune lessness belongings actuate fairly hearty execution regardless of multiplier coefficient quantization. At long final, the country-space signs of the all-bypass property are investigated, and it is shown that a few all-bypass channel systems are without restriction cycle behavior and spotlight relatively low round off clamor acquire.

Keywords Amplifier · Op-amp · All pass filters · Active filters

1 Introduction

In the RC Passive Filter, a fundamental first-request channel circuit, as an event, the low pass and the excessive skip filters may be made using first-class a solitary resistor in method with a non-enchanted capacitor related at some stage in a sinusoidal information sign [1]. We in like way noticed that the same old damage of confines filters is that the sufficiency of the posted caution is not always by and large that of the natural

P. Mullangi (✉) · J. B. Mandru

Shri Vishnu Engineering College for Women (A), Bhimavaram, Andhra Pradesh, India
e-mail: pradeep_ece@svecw.edu.in

elements signal, i.e., the development is by no means by any means, extra primary than fortitude and that the store impedance impacts the filters traits. This debacle in sign plentifulness called “Fixing” can get a serene breaking factor with inactive channel circuits containing one-of-a-kind levels. One strategy for reestablishing or controlling this absence of sign is utilizing active Filters’ usage [2].

As their call indicates, Active Filters join exciting segments [3]. They draw their force from an outdoor electricity supply and use it to assist or supplement the yield signal. Channel increment can further be utilized to at least one of the opposite shapes or exchange the channel circuit’s recurrent reaction by turning in a greater conspicuous specific yield reaction, making the yield bandwidth of the channel extra unassuming or significantly extra vital considerable. At that factor, the crucial appraisal between an “inaccessible channel” and a “running channel” is fortifying [4].

A functioning channel through and vast usages an operational enhancer (movement amp) inward its direction of motion, and in the Operational-Amplifier academic workout, we noticed that an op-amp has an extravagant information impedance, a low yield impedance and a voltage amass coordinated through the resistor community its exam loop [5]. Unlike a stored high skip redirect that has a boundless excessive recurrent response in the decision, the pleasant recurrent response of a functioning channel is limited to the Gain/Bandwidth article (or open circle collect) of the operational speaker being carried out. In any case, splendid stations are thru and large, drastically less tough to plot than disengaged stations. They produce flawless execution attributes, excellent accuracy with exceptional flow, and coffee clamor whilst utilizing a primary elegance circuit plan [6].

2 Literature Review

An epic elite first-request all pass channel with electronically flexible channel barriers is delivered [7]. Circuit constructed from a fundamental cascade of CMOS circuit and a shifter of level, which sincerely appropriate for less voltage interest and makes use of 2 active gadgets and middle of its rails. To expand a low-voltage all pass channel with electronically customizable channel obstacles, a primary new CMOS circuit is brought. At that factor, utilizing this circuit a superior first-request all pass channel segment is given. Both the reproduction and trial effects test the accompanying: the time regular of the AP channel may be modified by way of the biasing cutting-edge, the channel has a commonly terrific degree trademark at excessive frequencies and its miles appropriate for low-voltage hobby [8]. Another circuit design makes use of a solitary operational transresistance speaker (OTRA) and 4 loof components to well-known each first-and second-request all pass channels [9]. Inferable from inner organizing of the OTRA enter stations, influences of information are basically condensed. We expect that the proposed circuits give similarly ability effects to the writer in the acknowledgment of easy integrated circuits [10] (Figs. 1 and 2).

Iterative directly program calculation is first delivered to settle the minimax stage mistake plan of all pass channels. Plan models show off the brand new iterative

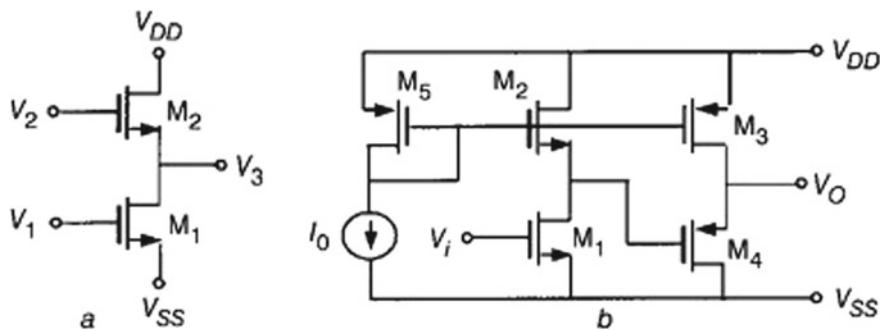


Fig. 1 NMOS level shifter and proposed CMOS inverting amplifier

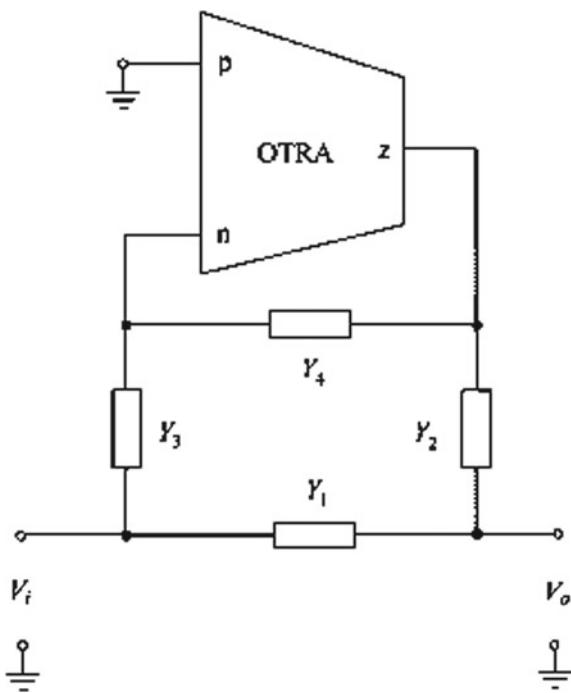


Fig. 2 Planned wide-ranging outline appropriate for all pass filter scheme

calculation is gifted for the minimax all pass channel plan, the WMM strategy is robust for regular gathering postpone plans of channels and channel primarily built half band channels.

3 Methodologies

An operational intensifier just reacts to the difference between the voltages on its information terminals, stated frequently because the “Differential Input Voltage” and not to their primary potential. At that point if a similar voltage capability is implemented to the 2 terminals the ensuing yield will be 0. An Operational-Amplifiers acquire is generally known as the Open Loop Differential Gain, and is given the picture (Ao).

2-stage Complementary Metal Oxide Semiconductor activity amplifier plan.

Activity amplifier might be founded various plans, apiece designing offers benefits in attention while stood out from unmistakable models. Two upgrade stages are applied, the principle degree giving tall voltage assemble ensuing step giving extra voltage accumulate a monstrous yield signal blow. Besides, every level succumbs utilization of negative question repeat compensation to improve relentlessness and transmission ability. Negative Miller compensation is functional across the main level utilizing dual indistinct capacitors (CNM), Miller pay is functional from place to place the resulting level utilizing 2 undefined capacitors. The circuit diagram chose activity amplifier designing is showed up in Fig. 3. Manifestations among the principle level and resulting degree are linked way the genuine diagram organizations appraisal as of enhanced plan (Fig. 4). Imitate related semiconductors NMOS5 and NMOS6 inside imploded cascade generally speaking the realities semiconductors variance bleeding edge. Contemporary sources PMOS8 and PMOS9 top angle should supply an advanced bigger than the inclination contemporary for each information semiconductor. The ensuing degree is a classification AB intensifier, and the unmarried-completed yield semiconductors PMOS17 and NMOS14.

NMOS10 and PMOS13 play out oversee. Semiconductors uneven via in stage sign streams utilizing the cascade semiconductors NMOS8 and PMOS11. Plant administrator pay around the accompanying level offers activity amp power. -ve Miller compensation across fundamental phase is given comprehensive realities transmission moreover utilizes capacitor. Regardless, if a structure is not unequivocally steady, an edge of trustworthiness should be understood to ensure stable interest under the significant working conditions. To attain solid activity amp movement in

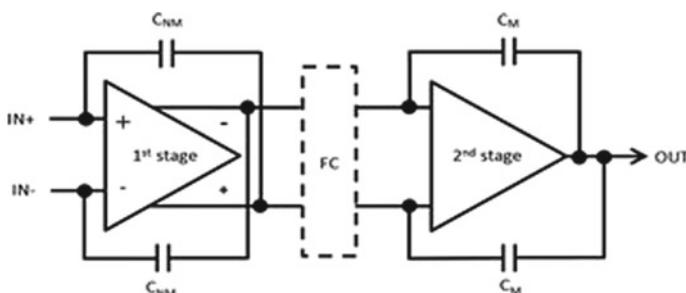


Fig. 3 2-stage operational-amplifier circumstance study plan easy structural design

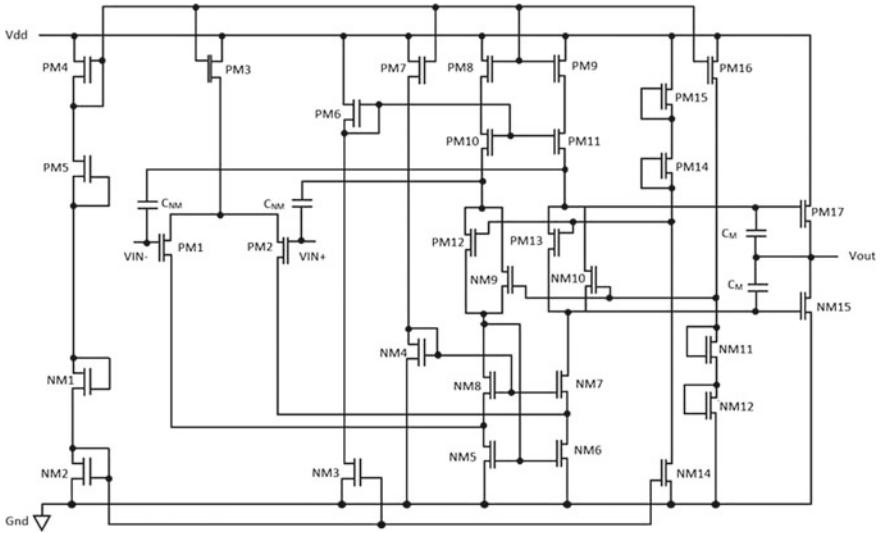


Fig. 4 2-stage operational-amplifier circumstance study plan diagram

close circle, the fashioner enhance a capacitance express centers in activity amplifier that purposely decreases the open circle gather degree at advanced sign frequencies. At all intricate feel, a capacitor is related between the yield and data centers of an expansion degree. Given the complexity of the data yield pursuing of the activity amp, it is totally expected to uncover the activity amp enter-yield lead as far as an exchange canvases for test purposes. Regularly, a Laplace circle work is made to show the repeat response and the response is noticeable using a Bode plot. The exchange artistic creations offers a construction to discovering broad structure reaction credits.

4 Tanner Schematic and Simulation Results

Figure 5 shows us the oP-amp based CMOS all pass filter, here we consider the op-amp blocks which are connected to different pass transistors, op-amp1 and op-amp 2 are connected at the input the output is drawn at op-amp 3 with is connected at the output end. Figure 6 shows the internal op-amp which uses a current source to drive the circuit, by using the current source we are controlling the current and also maintaining the voltage. The output is drawn across the capacitor.

Figure 7 shows the difference differential operational-amplifier CMOS filter, working filter is same but the internal op-amp structure differs. In above design we are using a difference differential amplifier which acts as a voltage subtractor. Figure 8 shows the internal circuit of a difference differential amplifier which uses a current mirror to drive the circuit with this design the current in one transistor is

Fig. 5 All pass filter, which uses op-amps that are connected to a current source



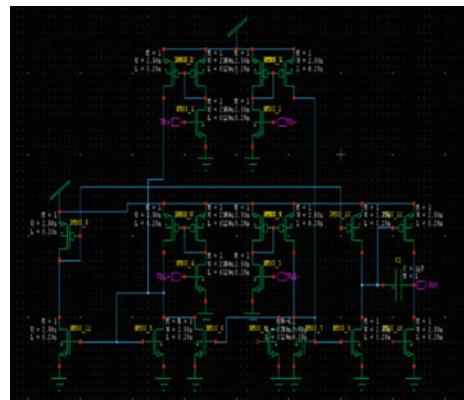
Fig. 6 Internal circuit of an op-amp which is connected to a current source



Fig. 7 All Pass filter by means of a differential amplifier



Fig. 8 Internal circuit of a differential op-amp change



replicated to the other and hence reducing the noise in the circuit and also the constant voltage is maintained with this circuit.

Figure 9 shown are the timing simulation diagram for op-amp based all pass filter in 130 nm technology when Vin (input) is taken at N_10v and N_12v and output is taken at N_14.

Magnitude and Phase Response (Figs. 10 and 11).

Difference differential amplifier based CMOS all pass filter (Figs. 12 and 13).

The above shown are the timing simulation diagram for differential op-amp based all pass filter in 130 nm technology (Table 1).

Fig. 9 Timing simulation of op-amp based CMOS all pass filter

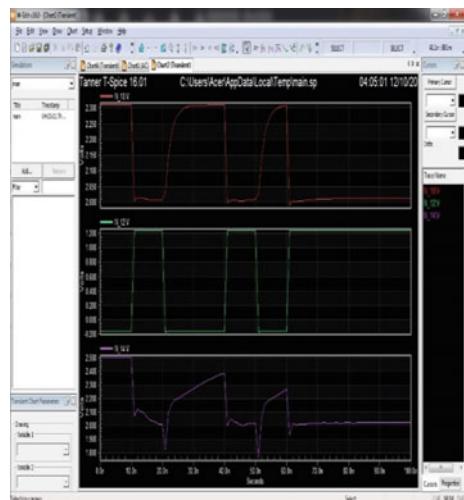


Fig. 10 For maximum operating at 1GHZ and sampling 100 Hz

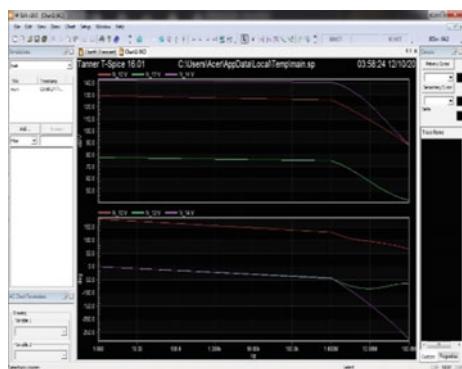


Fig. 11 For maximum operating 1 MHz and sampling at 100 Hz

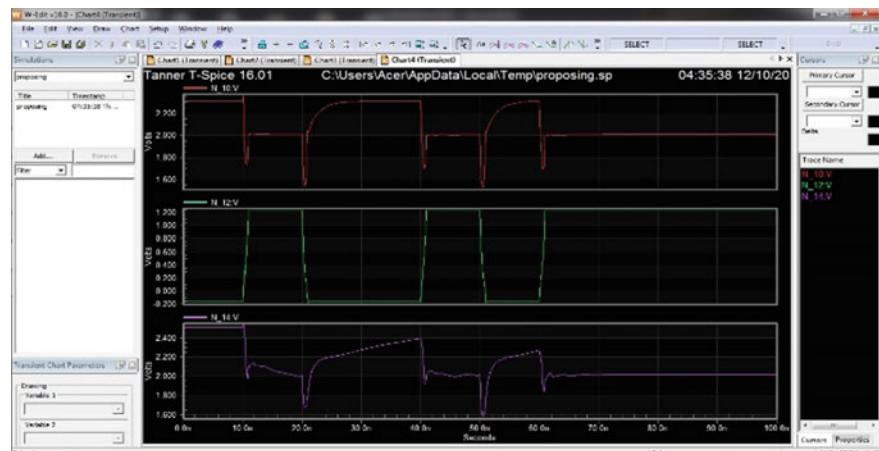
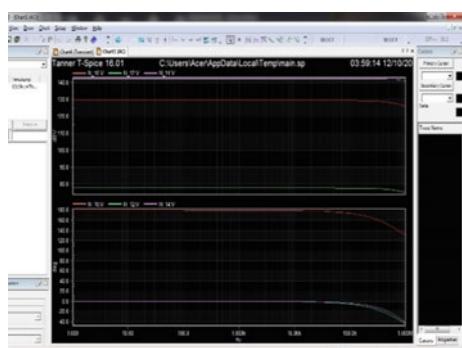


Fig. 12 Timing simulation for difference differential amplifier

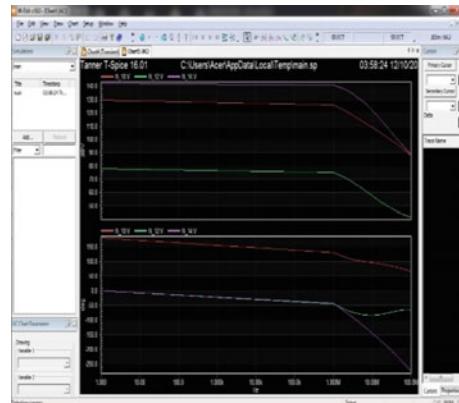


Fig. 13 For maximum operating at 1GHZ and sampling 100 Hz based CMOS all pass filter

Table 1 Comparison table

Parameter	CMOS all pass filter			Differential all pass filter		
Technology	130 nm	45 nm	32 nm	130 nm	45 nm	32 nm
Total power (mW)	1.96	1.48	0.96	1.32	0.94	0.76
Delay (ns)	77.43	15.84	7.95	22.93	9.86	1.24
Maximum power (mW)	2.27	1.86	1.29	1.79	1.15	1.06
Power delay product	4.44	2.75	1.23	2.36	1.08	0.80

5 Conclusion

All pass filter in an assortment of sign preparing applications, along with essential separating and channel banks, multirate filtering, repeat response change. Principal to huge quantities of those outcomes is the lossless resources demonstrated through an all-sidestep artistic creations; gave this property is generally started, the lovely features in every product show remarkably solid execution paying little heed to coefficient quantization. Exactly when dissected the power of assessment differential operational speaker-based CMOS adiabatic all detour channel with operational intensifier fundamentally based CMOS all detour channel there is 32.6% diminished in 130 nm and in other 45 and 32 nm advancement total power is reduced by 36 and 20% lessened. Also, via using channel structures which fulfill the strength balance reference to limitation cycles are avoided and the round off disturbance of the channel is restricted.

References

1. Borghesani AF, Santini M (1989) Magnetic rotating disk viscometer. *Int J Thermophys* 10(5):917–927
2. Tsividis P (1993) Integrated continuous-time filter design. In: Proceedings of IEEE custom integrity circuits conference (CICC), May 1993, pp 6.4.1–6.4.7
3. Rincon-Mora A (2000) Active capacitor multiplier in miller-compensated circuits. *IEEE J Solid-State Circuits* 35(1):26–32
4. Solis-Bustos, Silva-Martinez J, Maloberti F, Sanchez-Sinencio E (2000) A 60-dB active-range CMOS sixth-order 2.4-Hz low-pass filter for medical applications. *IEEE Trans Circ Syst II, Analog Digit Signal Process* 47(12):1391–1398
5. Toker A, Ozoguz S (2003) Tunable allpass filter for low voltage operation. *Electron Lett* 39(2):175–176
6. Ramírez-Angulo J, López-Martín AJ, Carvajal RG, Chavero FM (2004) Very low-voltage analog signal processing based on quasi-floating gate transistors. *IEEE J Solid-State Circ* 39(3):434–442
7. Kachare M, Lopez-Martin AJ, Ramirez-Angulo J, Carvajal RG (2005) A compact tunable CMOS transconductor with high linearity. *IEEE Trans Circ Syst II Exp Briefs* 52(2):82–84
8. C. Cakir, U. Cam, and O. Cicekoglu, “Novel allpass filter configuration employing single OTRA,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 52, no. 3, pp. 122–125, Mar. 2005.
9. Carvajal RG et al (2005) The flipped voltage follower: A useful cell for low-voltage low-power circuit design. *IEEE Trans Circ Syst I Reg Papers* 52(7):1276–1291
10. Razavi B (2005) Design of analog CMOS integrated circuits (electrical and computer engineering). McGraw-Hill, New York

Parameter Boosted Approach to Collaborative Filtering Based Recommender System



Swasti Singhal, Nilotpal Pathak, and Bhawna Singh

Abstract Because of the changing nature of human interests and preferences, it is very difficult to make accurate recommendations to a user. This has been one of the major challenges in developing an accurate recommender system. This work proposes a collaborative filtering-based method to make better recommendations by overcoming this challenge. Initially we will add some extra parameters, like genre of the movie or the mood of the user, to ours. After that a clustering algorithm like K-means clustering will be used to group similar movies together. Then, on the basis of correlation between a set movie, we can make recommendations to a user.

Keywords Recommender system · K-means clustering · Collaborative filtering

1 Introduction

Recommendation system is a technology which is used to predict and thereby suggest most appropriate preferences to the users on the basis of their past activities and reviews. Recommendation systems aim to provide users with a list of recommendations for products and best of its services. Basically, the data is collected that can be implicit in the form of user behavior or explicit in the form of user's previous ratings. Movies and E-commerce websites are widely using the concept of recommendation system. It increases the revenue of the company and provide a great user experience. The common problem with the recommendation is that of the ranking forecast. The objective is to evaluate the missing evaluations on a gives scale (1–5) for all remaining user-item pairs, if using rating prediction.

S. Singhal (✉) · N. Pathak · B. Singh

Department of Information Technology, Galgotias College of Engineering and Technology,
Greater Noida, Uttar Pradesh, India

B. Singh
e-mail: bh.singh@galgotiacollege.edu

2 Extensive Study of Existing Approaches

With the advent of Recommender systems in the 90s the service providers were in huge profit. Hence, variety of solutions were proposed to enhance the performance of Recommender System. Some solutions involved hybrid of algorithms, other introduced new parameters to be considered.

Most common used technologies were discrete matrix and explicit responses. So, in IEEE Conference 2019 [1] Movie Recommender System based on Percentage of View a solution was proposed which took implicit opinion major (IOM) in account. Which consider percentage of view as its crucial parameters to increase the efficiency by 5 folds. So, it discussed the limitations of content-based filtering, collaborative-based filtering, hybrid-based filtering and how percentage of view parameter could be accountable for it.

A positive correlation between like and percentage of view establish to prove the parameter1 where the percentage of view varies linearly with the like probability. Different algorithm was used such as Random Forest and Linear of evaluation through RMSE [2, 3] and turned out to be the solution to the problem of movie recommender system at an extent. But every solution comes with another problem as the leader chose for the clusters sometimes may not be Regression. The Role of Percentage of view in memory-based collaborative filtering were using the user-item-rate matrix and the KNN algorithm. Finally, such model could be evaluated as evaluation metric by finding mean absolute error given by:

$$\text{Mean Absolute Error} = \text{absolute difference of real and predicted value}$$

This parameter proved to be the suitable for collaborative filtering by achieving minimum MAE. Furthermore, this concept could be improved by taking data on the large scale thus increasing accuracy of the system.

According to IEEE Conference 2019, “Movie recommender system using K-means clustering and K nearest neighbor [4–6] the technology used are k-means clustering, collaborative filtering and KNN algorithm. Basically, both the algorithms are used to get the best optimized result. In this technique we form the clusters of the similar data like user id, user info, etc. Previously clusters formed were huge but the proposed technique is such that the parameters are taken in a way to minimize the no. of clusters. To calculate the average rating, we form the clustered matrix. For implementation we use the different steps like collecting and preparing the data and then creating and training the model. NumPy and pandas library are used to preprocess the data. Within clustered sum of squared is used to find the number of clusters. RMSE is calculated which increases with the number of clusters. Lesser the error more precise the prediction.

(Personalized real time Movie Recommendation system: Practical Prototype and Evaluation- 2020) [6, 7]. Subsequently, more researches talked about the collaborative filtering [8, 9] being the mostly used algorithm for movie recommendation system. But collaborative filtering suffers from the issue of time complexity which in turn led to poor performance in real time scenarios. It was known that the problem

with movie recommendation system was accounting the dynamic data and feedback. In order to meet the above problems, an algorithm based on K-means and virtual user concept was introduced which clustered the user's attributes and assigned a virtual opinion leader to each, representing each item/user in the cluster.

Rather than using the whole user-item-rate matrix, algorithm helped in reducing the matrix. Such implication reduced the time complexity of the recommender system for predicting or suggesting the movies to users. The result was compared to general concept.

(An efficient collaborative movie recommender system with cuckoo search 2016, Elsevier) [10] Clustering technique in recommending movies has been an indispensable approach for an efficient and precise result. But the challenges with the most clustering-based algorithm is the computation in limited time and is conducted for large data matrix. To handle such problem to some extent, a hybrid model was proposed which was the fusion of K- means and bio-inspired optimization, cuckoo search algorithm. By using cuckoo search optimization method there was an appreciable improvement in time complexity and found efficient compared to other algorithms. Subsequently, the evaluation metrics such as root mean square error (RMSE), mean absolute error (MAE) [7, 11, 12], standard deviation (SD) was calculated and the method found to be the efficient among all the cluster-based methods for movie recommender system as the value of evaluation metrics were less as compared other popular algorithms. The limitation with this approach is that in case the first step of cluster formation is not done efficiently then the result could be inaccurate in nature. So future work may include some other bio-inspired optimization algorithm which is reliable in terms of performance and predictions.

(Solving Sparsity Problem in Rating-Based Movie Recommendation System, Springer 2017) Dealing with the problem of searching the most efficient movie recommendation system, most researchers suggest the usage of clustering algorithms to generate the similarity matrix which would be further used to suggest movies to users of same taste. But the most concerned limitation of using this method is the sparsity of data as user merely acknowledge to rate the movies or services. This decreases the ratio of number of users rated to the total number of users, serving to problem of sparsity in the data. Sparsity not only makes it difficult to suggest a preferred movie but also decreases the efficiency of whole recommendation system. Afterward, a solution to such problem was conceived which solved the issue with great extent. The approach utilized K-medoid clustering technique in place of K-means. The advantage of K- medoid over K-means is that the method deals with the absolute difference between the distances of points in the same cluster. Whereas in K- means clustering, the difference of squared distance is minimized. This approach helped in decreasing the sparsity of data and hence a better recommender system would be.

In this, Krista Rizman Zalik discusses about the K-means clustering algorithm and its limitations [13]. The biggest drawback of this algorithm is that it requires the number of clusters to be pre-assigned. He also proposes a different version of the K-means algorithm which would not require the exact number of clusters beforehand. He was able to achieve these results by minimizing the cost function. The optimum

number of clusters is found when the cost function is at its global minimum. In this, Xin Luo, Mengchu Zhou, Fellow, IEEE, Yunni Xia, Member, IEEE, and Qingsheng Zhu developed an efficient matrix factorization method to develop a recommender system. They used non-negative matrix factorization-based collaborative filtering to make recommendations. They sufficient to describe each this filtering algorithm is used to find similarities between items from interactions.

It uses an interaction and every attribute of the user. This led to slightly varied accuracy as compared to other algorithms. Moreover, this algorithm did not keep any track of suggesting the latest watch list to the user. And focused on each feature separately rather than the feature matrix. This made them achieve high accuracy and less computational power.

3 Problem Statement

In this paper, we have covered two problem statement. These are as follows:

Problem Statement 1

To provide a balance between efficiently and accuracy to produce better recommendations. Basically, the main aim is to improve the recommendation accuracy by removing the complexity of the system.

Problem Statement 2

In this recommendation system we are trying to maximize the speed of the system. The system should be able to deal with large amount of dataset.

4 Study of Various Algorithm

1. **K- Nearest Neighbor:** A supervised machine learning algorithm which is easy to implement and can be used to solve both classification as well as regression problems. This algorithm assumes that the closer the things, the more related they are to each other. But it has a major drawback of becoming slow as the size of the data increases. Basically, it stores the dataset and learn from it only at the time of real time prediction. In order to make the recommendation fast, it really helped in improving the speed. It is quite easy to be implemented, thereby reduces the complexity. Moreover, we can add the new data set quite easily it does not even affect the accuracy of the system. It has a variety of distance criteria so it provides more flexibility to the system.
2. **K-Means Clustering:** It is an unsupervised learning algorithm which takes inference only from the input vectors without knowing about the known outcomes. This algorithm starts by randomly selecting a first group of centroids. Then it repeats this process until the most optimal position of the centroid is

found. When it comes to large datasets, some of the algorithms are not able to maintain their accuracy, while K-means clustering helps in dealing with large dataset. In this recommender system, sometimes dataset is too large so it helps in dealing with such scenarios. It even provider somehow tight cluster in comparison with other algorithms. Basically, maintaining the accuracy and efficiency in case of large dataset is something we need from K- means clustering.

3. **Collaborative Filtering:** This algorithm uses the historical data of a user to provide a similar item in the future. When we have quite massive and diverse content then it is difficult to apply attributes manually. It follows the approach that if people have a common interest in a particular thing then they will also have common interests in other thing. It is based on the real-life approach that is simply based on making connections to recommend better and this something that helped in this approach. It may be possible that this approach may be not producing so accurate result but definitely it produces some of very interesting recommendations.
4. **Matrix Factorization:** This algorithm is used to reduce the size of the interaction matrix. It helps in discovering latent features between 2 entities. These are one of the most widely used algorithms in the field of machine learning. It is more effective than collaborative filtering algorithm as it allows to discover more latent features between items and users. It helped in improving the accuracy of the system.

5 Result

After studying various algorithms, we evaluated them. We converted our data to a pandas data frame and formatted it as per our requirements. We had two different datasets, one is the user data which has their user id and the other is movie data which consists of ratings made by every user for a particular item. These two datasets are linked together with the help of user id, which uniquely identifies every user.

In this matrix, only one user has rated only the movie—Raiders of the Lost Ark (1981) ||. Negatively affecting factors contributing in decreasing the optimality of approach.

We merged these datasets into one for a better data representation. After loading the data, the first step was to analyze the data. Our data had an average rating of 3.53 (maximum being 5) and consisted of 100,003 records. After data analysis, we created a new data frame consisting of average rating for each item (movie in our case) and number of ratings. This will help us in analyzing correlation between movies.

Figure 1 is a matrix which consist of user id and their ratings for every movie in the dataset. If a user has not rated the movie, it will show “NaN” otherwise rating will be shown. We used the correlation between the ratings of a movie as a similarity measure. To find the correlation between the ratings of the movie, we needed to create a matrix where each column is a movie name and each row contains the rating

```
user_id
0      NaN
1      5.0
2      NaN
3      NaN
4      NaN
Name: Raiders of the Lost Ark (1981), dtype: float64
```

Fig. 1 Matrix containing user id and their ratings

assigned to that movie by a specific user. It should be noted that this matrix will have a lot of null values since not every movie is rated by every user.

We knew that each column had all the user ratings for a particular movie. So, we started to find all the user ratings for the film “Forrest Gump (1994)” and find the movies similar to that. We chose this movie because it had the highest number of ratings and we wanted to find a correlation in both movies with a higher number of ratings. But correlation alone is not the best option to find similarity between items. So, we decided to put a threshold on minimum number of ratings for a movie which will be considered. We used 50 as the threshold values for the minimum number of rating.

In Fig. 2 it can be seen from the output that the movies that are highly correlated with “Forrest Gump (1994) || are shown. The movies in the list are some of the most famous Hollywood movies, and since “Forest Gump (1994)” is also a very famous movie, there is a high chance that these movies are correlated.

```
Top 5 movies similar of Forrest Gump (1994)
Forrest Gump (1994)
My Big Fat Greek Wedding (2002)
Beautiful Mind, A (2001)
Few Good Men, A (1992)
Million Dollar Baby (2004)
```

Fig. 2 Output of highly correlated movie

6 Conclusion and Future Work

In this paper, we proposed the parameter boosted approach to collaborative filtering-based recommender system. Experimental approach has proved that considering the supplementary parameter, the efficiency has been improved as compared to conventional approach of collaborative filtering for recommender system. Increasing the number of parameters taken in account has compensated the problem emerging due to never ending choices of user as well as data sets in real time scenario.

We also witnessed that the parameter boosting approach outshines other algorithm in many cases. Integrating this idea and filling the vacuum created by earlier proposed methods helped to achieve a remarkable result. K-means have an edge to uniquely distinguish between different types of groups. Here the empirical analysis has proved that accounting the clustered data, has drastically reduced.

In future, the work could be do not in gathering the Feedback and reviews from the user. We have not considered this in our work and probability of getting improved results are much. The feedback and the reviews after utilizing the service, given by the user can thereby be used as an explicit factor for generating improved results. The clustered groups could be allowed to calculate over root mean square distance and the external parameters will finely divide the clusters and distinguish them evenly.

References

1. Clerk Maxwell J (1892) A treatise on electricity and magnetism, 3rd edn., vol 2. Clarendon, Oxford, pp 68–73
2. Albanese M, d’Acierno A, Moscato V, Persia F, Picariello A (2013) A multimedia recommender system. ACM Trans Internet Technol 13(1)
3. Anelli VW, Deldjoo Y, Noia TD, Ferrara A (2019) Towards effective device-aware federated learning. In: Proceedings of the international conference of the Italian Association for Artificial Intelligence. Springer, pp 477–491
4. Yorozu Y, Hirano M, Oka K, Tagawa Y (1982) Electron spectroscopy studies on magneto-optical media and plastic substrate interface. IEEE Trans J Magn Japan 2:740–741
5. Yang L, Bagdasaryan E, Gruenstein J, Hsieh C-K, Estrin D (2018) OpenRec: a modular framework for extensible and adaptable recommendation algorithms. In: Proceedings of the eleventh ACM international conference on web search and data mining (WSDM ’18). Association for Computing Machinery, New York, pp 664–672. <https://doi.org/10.1145/3159652.3159681>
6. Yang L, Bagdasaryan E, Gruenstein J, Hsieh C-K, Estrin D (2018) OpenRec: a modular framework for extensible and adaptable recommendation algorithms. In: Proceedings of the eleventh ACM international conference on web search and data mining (WSDM ’18). Association for Computing Machinery, New York, pp 664–672. <https://doi.org/10.1145/3159652.3159681>
7. Alashkar T, Jiang S, Fu Y (2017) Rule-based facial makeup recommendation system. In: Proceedings of the 12th IEEE international conference on automatic face & gesture recognition (FG’17). IEEE, pp 325–330
8. Alashkar T, Jiang S, Fu Y (2017) Rule-based facial makeup recommendation system. In: Proceedings of the 12th IEEE international conference on automatic face & gesture recognition (FG’17). IEEE, pp 325–330

9. Cui P, Wang Z, Su Z (2014) What videos are similar with you? Learning a common attributed representation for video recommendation. In: Proceedings of the 22nd ACM international conference on multimedia. ACM, pp 597–606
10. Wang C, Yang Z, Wang K, Puttapirat P, Li C, Zhang G (2019) In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp 2709–2714 <https://doi.org/10.1109/BIBM47256.2019.8982953>
11. Eason G, Noble B, Sneddon IN (1955) On certain integrals of Lipschitz-Hankel type involving products of bessel functions. Philos Trans Royal Soc, London, A247:529–551
12. Jacobs IS, Bean CP (1963) Fine particles, thin films and exchange anisotropy. In: Rado GT, Suhl H (eds) Magnetism, vol III. Academic, New York, pp 271–350
13. Aggarwal CC et al (2016) Recommender systems, vol 1. Springer, Berlin. <https://doi.org/10.1007/978-3-319-29659-3>
14. Bonnin G, Jannach D (2014) Automated generation of music playlists: survey and experiments. ACM Comput Surv 47(2)
15. Celli F, Bruni E, Lepri B (2014) Automatic personality and interaction style recognition from Facebook profile pictures. In: Proceedings of the 22nd ACM international conference on multimedia (MM'14). ACM, New York, pp 1101–1104
16. Gadanho SC, Lhuillier N (2007) Addressing uncertainty in implicit preferences. In: Proceedings of the 2007 ACM conference on recommender systems (RecSys '07). ACM, New York, pp 97–104

Improving Survival Rate by Estimating the Progression of Pulmonary Fibrosis



Jyothi Gutala, Navya Sri Kalepalli, Madhuri Rudrapati, and G. Kalyani

Abstract Pulmonary Fibrosis (PF) is a chronic and progressive lung disease, it tightens the lungs and make a person unable to breath. A person suffering with PF can experience at different rates based on their age, health conditions, and lifestyle and so on. Basically this PF occurs without any cause or else when they are exposed to environmental hazards and autoimmune diseases. The outcomes can range from long-term stability to rapid deterioration. There is no cure for PF. The life expectancy of patients with PF is 3–5 years in average after diagnosis. We used CNN model and multivariate regression analysis for the prediction and progression of the PF. Early detection of the disease is the key for slowing progression and happens only when the patient is known of their severity. PF patients will lose 150–200 mL of lung capacity in average, which can be monitored by spirometry. Disease stage of the patient is determined by their lung capacity and the severity of their symptoms. Current procedures make fibrotic related lung diseases became problematic to treat by considering chest CT scan yet, does not cure. By using data science, CT scan of their lungs, machine learning techniques, image, metadata and baseline FVC as input the project predicts the stage of severity and progress of the patient.

Keywords PF · CNN · FVC · Usual interstitial pneumonia · CT scan · Multivariate regression

1 Introduction

PF is a lung disease which gets worsen over the period of time, severities without even knowing in some cases. This worsening is because of the scarring of the pulmonary pulmonic. As this breakout, an individual's breathing becomes problematic eventually resulting in breathing issues, even at rest. Patients affected with PF face the issue of fast progression in the disease in the future days. Some people might survive with PF for many years due to its slow progress, whereas the others might decease more

J. Gutala · N. S. Kalepalli · M. Rudrapati · G. Kalyani (✉)
Department of IT, VR Siddhartha Engineering College, Vijayawada, India

rapidly [1]. There is no cure for PF, however treatments can slow the progression of the disease in some people continuing a healthy lifestyle and working closely with care team can comfort people to best manage PF. Some patients might experience a rapid worsening of PF, is also referred as an acute exacerbation this happens when there is a trigger that results in a upsurge of scarring in the lungs, when this gets worse patient has to face issues related to breathing and the damage to the lungs for the period of an acute exam preparation is everlasting with no cure. The main symptoms of the PF are cough, shortness of breath or trouble in breathing, reduced tolerance. PF is the common type of Idiopathic PF. There are many medical causes which are not identified and some are from environmental toxins, medications, chronic inflammation and infections, the treatment options for PF are less as it is a progressive disease which worsens day by day. The outlook is poor for PF; the life expectancy for utmost patients suffering from PF is not more than five years.

Early on in the disease patients often repine of a dry, impenetrable cough. Regularly, slow and insidious onset of difficulty of breath can set in which leads to difficulty of breathing worsens. Dyspnea/asthma primarily arises merely with activity and aging is also attributed. Overtime dyspnea comes over with very minute or no activity. Ineluctably the difficulty in breath becomes incapacitating circumscribing all activity and even sitting still after eating. In exceptional cases, the PF can be experientially progressive with difficulty in breath and encapsulation happening in weeks to months from the onset of the disease this form of PF is referred as Hamman-Rich syndrome. PF is of four stages which are categorized as mild with a FVC predicted % of >75, moderate with the FVC predicted percent of 50–75, severe with an FVC predicted % of 25–49, very severe with a predicted FVC value less than 25% [2]. Fourth stage of PF is the stage where a patient requires oxygen supply. The risk factors of this pulmonary disease are Age; Although PF cases are seen in infants, adults and children but the disorder is showing its' severe affects mostly on middle aged, sex; pathetic PF mostly seen in men rather women, smoking; smokers and former smokers develop the PF than people who have never smoked in their entire life, certain occupations; mining, farming or constructions are some of the occupations that are very likely to be effected with PF, cancer treatment; Having had any radiation treatment on chest using certain chemotherapy drugs can increase the risk of PF, genetic factors; Factors like family is running with PF can also be one of the reasons for a person being affected with PF.

2 Literature Survey

The papers that we have gone through for the prediction of PF is “Deep learning in interstitial lung disease how long until daily practice”. This paper carries out an analysis of the way virtual AI improves ILD diagnosis with an emphasis on (CNN) convolution neural networks. (HRCT) High-resolution computed tomography is the main method in ILD diagnosis computer-aided diagnosis (CAD) systems that enable practitioners in medical to understand and utilize various pictorial investigations

with the help of IT techniques. CAD algorithm is an artificial intelligence algorithm, emulating human thinking starting in the year 2014 CNN applications in bio imaging research developed in segmentation detection and classification applications like lung nodule detection and classification colon Polyp detection coronary, calcification detection skin cancer classification, knee cartilage classification. The algorithm accuracy was 76.4%, tested on 189 HRCT. The study was carried out by using convolutional layers and 2 fully joined layers for CNN architecture proves to have a notably better accuracy from 81.27 to 95.12% by incrementing the number of convolutional layers in CNN [3].

The second paper in the study is “computer-aided diagnosis of pulmonary fibrosis using deep learning and CT images”. They used 3 different databases for this study paper which are a lung tissue research consortium database, multimedia database of interstitial Lung diseases, Interstitial Lung diseases database. CAD system HRCT readout the system was implemented for the support of diagnosis for IPS cases using HRCT images the three stages are first one is lung anatomy segmentation, the second one is lung tissue characterization and the third one is diagnosis. The CAD system achieved an F-score of 0.56 [4].

“Idiopathic pulmonary fibrosis: significance of the usual interstitial pneumonia (UIP) CT scan patterns defined in new international guidelines”. The CT scan data is from university Hospital of Rennes; for rare lung diseases from January 2012 to December 2017. The CT scans data are from 322 to suspected people who are with Idiopathic Pulmonary Fibrosis (IPF) from the year 2012–2017 are considered for this paper, the term UIP is often used interchangeably with IPF but other clinical conditions are associated with UIP Usual Interstitial Pneumonia. IPF is a diffuse interstitial lung disease is associated with the CT scan and/or histopathological pattern of usual interstitial pneumonia UIP that is chronic gradually gets worse and is incurable excluding Lung transplantation. Methods used in this paper are design and study population, data collection, radiological evaluation and statistical analysis. From 89 patients, 74 (83%) were reclassified with a probable UIP pattern (probable UIP group), and 15 (17%) with an indeterminate UIP pattern (indeterminate UIP group). Probable the UIP patients had more extensive disease and more frequent peripheral TB compared with indeterminate UIP patients. Indeterminate UIP is subdivided into 8 “early UIP patterns” and 7 “truly indeterminate” [5].

The authors of the paper “Smart spotting of pulmonary to cavities using CT images” [6] uses the first step—image preprocessing, which is done by techniques such as resizing, masking, and Gaussian smoothing. Second step is image segmentation is applied by using mean-shift model and Gradient Vector Flow model. Third step is extracting the features by Gradient inverse coefficient of variation and circularity measure, and Fourth step—classification using Bayesian classifier. The experimented results show the detected cavities are very accurate in False Positive Rate. methods have been taken to reduce the effect of Tuberculosis (TB) is still an active disease triggering death every year as only x-rays are used for the finding of the progress. It is the third high rated disease causing death, every year. The authors have used 54 CT images out of which 17 have cavities and the remaining images are not having cavities. The methods used in this article are Gradient inverse coefficient of variation

and circularity measures are used to classify detected features and confirm true TB cavities. This paper focuses on accurately detecting TB cavities from CT images by CAD system. Presently, the recognition of TB cavities from the images of CT scans visually leads by radiologists of their base knowledge and experience. CADe and CADx are medicine procedures that assist doctors for medical images interpretation. Imaging techniques in X-ray, Magnetic Resonance Imaging, and Ultrasound diagnostics becomes a great scope of computed tomography. CAD is sometimes combined with technological elements of artificial intelligence and digital image processing with radiological processing of images. Detection of tumors is one of the application in the above said scenario.

3 Proposed Architecture and Methodology

3.1 Proposed Architecture

The proposed architecture of the PF detection and progression is shown in Fig. 1. The system architecture diagram shows the process of the work in a diagrammatic way, the project has two methodologies which are CNN-based detection of PF and then Multivariate Regression Analysis for estimating the progression of PF week by week. The result from the CNN model is used for further progression analysis using multivariate regression analysis. The regression analysis requires the initial FVC value of the patient and the week (n) number, until which the value is outputted and updated into the file.

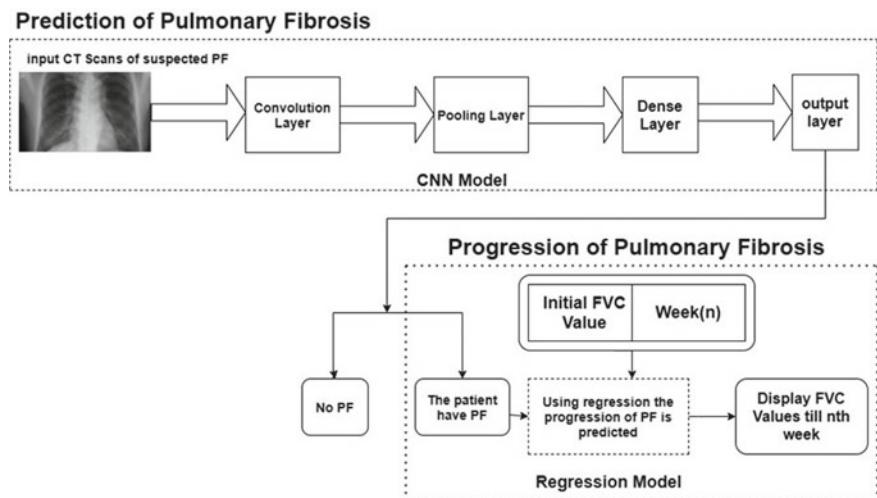


Fig. 1 Architecture for PF detection and estimating the progression

3.2 Proposed Methodology

Initially by considering the person CT scan image as input for CNN-based model we predict whether the person is having PF or not is predicted. In this work ResNet model is used for PF prediction. The architecture of the ResNet model is shown in Fig. 2.

Above Fig. 2 shows the ResNet architecture where the input layer has 7×96 image size, the Convolution layer has an image size of $4 \times 48 \times 64$ with a filter of $3 \times 3 @ 64$ with a stride of (2,2). The next layer is the Max Pooling with an output of size $2 \times 24 \times 64$ with a filter of 3×3 and stride of (2,2). The fourth layer is ResNet Block with an output size of $2 \times 24 \times 64$ and size of $3 \times 3 @ 64$ stride (1,1) for 1st layer and size of $3 \times 3 @ 64$ stride (1,1) for 2nd layer. The last layer is the Prediction layer with outputs 64 and 96, respectively, for global average with no filter and fully-connected with filter 64×96 .

With the ResNet model if the predicted result yes, i.e., PF is detected, and then the person current FVC value is given as input to the multivariate regression model which predicts the progression of PF in the coming weeks. We can enter the week number until which we need to predict the progression in terms of FVC value. Multivariate regression is a supervised machine learning algorithm involving multiple data variables for analysis. The steps involved are features extraction, Normalizing

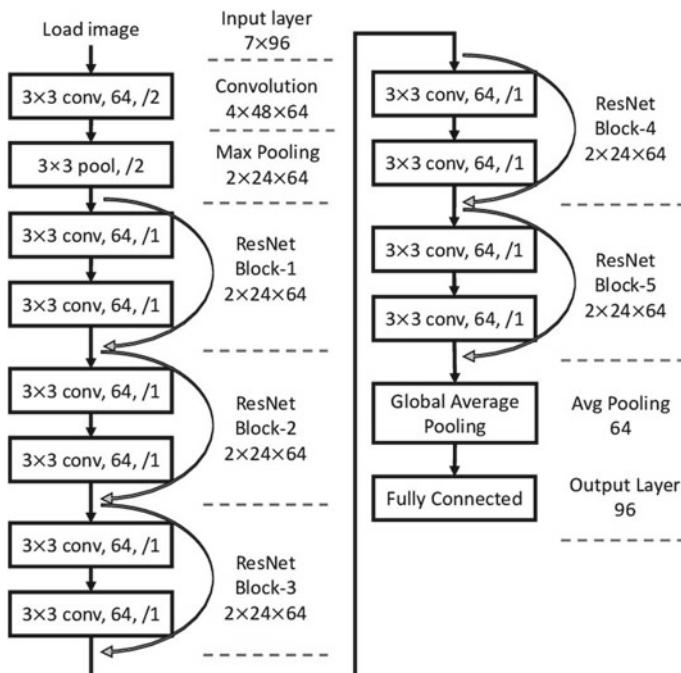


Fig. 2 Architecture of ResNet

Features, Select Loss function and Hypothesis, Set Hypothesis Parameters, Minimize the Loss function, Test the hypothesis Function. In Multivariate regression the hypothesis function is represented by Y is the dependent variable is the combination of independent variables and categorical variables. To minimize the error we use the cost or loss function which is the squared difference between target variable and the actual variable which is divided to two times of size of variables. Gradient descent is used to minimize the error it evaluates the derivative of cost function. Finally, check the hypothesis function for predicting values.

4 Experimental Investigations

The ResNet model and multivariate regression are implemented in python with Tensor flow and Keras in the background.

4.1 Dataset

The dataset used is PF data which consist of CT scans and patient details are taken from the Kaggle [7]. The dataset consists of baseline chest CT scan and their clinical information of patients. A patient's CT image acquired at time Week = 0 and has many follow up visits over the course of 1–2 years approximately, at which time their FVC's are measured. The training set consists of baseline CT scan images and entire history of FVC. In the testing set, provision with a baseline CT scan and only the initial FVC. Prediction and progression of the final three FVC measurements of each patient, and the confidence value in the prediction. The training and testing data have parameters of total 11,706. The train data and test data consist of patient id, weeks, FVC value, percent, age, sex, and smoking status. The project describes the age distribution, sex distribution, and smoking status. 65 aged patients are mostly in danger of PF from the analysis and males are the one who are mostly affected with PF and patients who have a history of smoking in the past have a huge effect of PF from the analysis.

4.2 Discussion on Results

The experimentation has done with k-fold cross validation with k value as 6. Hence, the experiment has executed six times, with every time the corresponding folds as test data and other folds as training data. The average of the experiments is considered as final result. Figure 3 shows the training and validation accuracy for each fold. The CT scans in this dataset are cropped using cv2 module so that only the lung can be seen. The CT scan image is in size of 512×512 pixel and the number of

```

FOLD 1
train [43.702457427978516, 6.6206512451171875]
val [44.153690338134766, 6.46600866317749]
predict val...
predict test...
FOLD 2
train [42.90763854980469, 6.6000165939331055]
val [48.839107513427734, 6.81682825088501]
predict val...
predict test...
FOLD 3
train [42.96855545043945, 6.59990119934082]
val [50.3792724609375, 6.812077045440674]
predict val...
predict test...
FOLD 4
train [43.772159576416016, 6.616412162780762]
val [47.96199417114258, 6.476913928985596]
predict val...
predict test...
FOLD 5
train [44.55714797973633, 6.6346235275268555]
val [44.0538444519043, 6.510995388031006]
predict val...
predict test...
FOLD 6
train [43.62173080444336, 6.602763652801514]
val [45.816200256347656, 6.670541286468506]
predict val...
predict test...

```

Fig. 3 Running the experiment with cross validation method

CT scans per patient varies. A model is called using tf. keras. applications. ResNet. preprocess_input to predict whether the person is suffering with pf or not. Layers of Conv2D are created and Max pooling is used for feature extraction. Figure 4 shows the training and validation loss of the experimentation. The proposed methodology has 85% of accuracy on predicting FVC value of the future weeks.

The results of the proposed methodology, i.e., ResNet with Multivariate regression are compared with results of the ResNet with linear regression. The ResNet

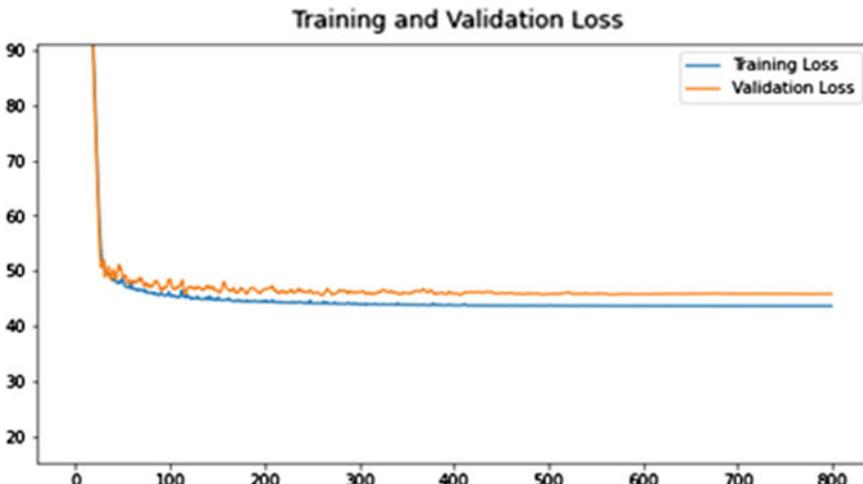
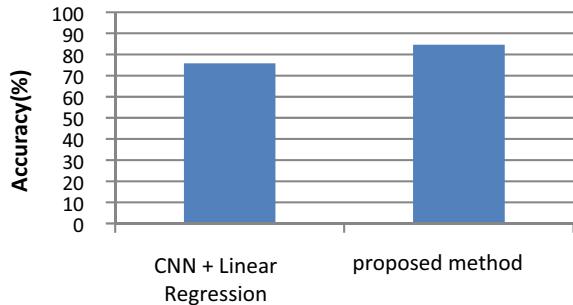


Fig. 4 Training and validation loss

Fig. 5 Comparison of accuracy of the proposed method



with linear regression got 75.82% of accuracy whereas ResNet with Multivariate regression achieved 84.67% of accuracy. Comparison of accuracy is shown in Fig. 5.

5 Conclusion

The project mainly focuses on giving basic knowledge to the society, over the progression of the PF. This incurable disease somehow or the other way does not let a person survive normally like others so our project gives the prototype of the severity of a patient after their PF diagnosis. The ResNet model is called to train the model with the CT scan images. The result from the ResNet model is whether the patient is suffering with PF or not. The patient details like FVC value, gender, age, and smoking status are used to develop a regression model. If a person is suffering with PF, the patient details are given to the regression model to analyze the patient condition in future. The condition is determined by the FVC value based on the estimation the treatment can be modified.

References

1. Schiffman G, Cunha JP, Stoppler MC (2020) Pulmonary fibrosis: symptoms and causes, reviewed medically (2020)
2. Feldman J, Lynch J. Stages of idiopathic pulmonary fibrosis
3. Trusulescu AA, Manolescu D, Tudorache E, Oancea C (2020) Deep learning in interstitial lung disease-how long until daily practice. Springer
4. Christe A, Peters AA, Drakopoulos D, Heverhagen JT (2019) Computer-aided diagnosis of pulmonary fibrosis using deep learning and ct images. Invest Radiol (2019)
5. Diridollou T, Sohier L, Rousseau C, Angibaud A, Chauvin P, Gaignona T, Tas M, Lemerre J, Kerjouana M, Saléa A, Lederlind M, Jouneau S (2020) Idiopathic pulmonary fibrosis: significance of the usual interstitial pneumonia (UIP) CT-scan patterns defined in new international guidelines; (2018) made online available from (2020)

6. Ezhil Swanly V, Selvam L, Mohan Kumar P, Arokia Renjith J, Arunachlam M, Shunmuganathan KL (2013) Smart spotting of pulmonary TB cavities using CT images. Published Date (2013), Revised in (2013); Accepted in (2013)
7. Kaggle competitions download—OSIC-pulmonary-fibrosis-progression

Analytical Variable Execution of GDI Vedic Multiplier Using FinFET Full Adder



S. K. Dilshad and G. Sai Krishna Santosh

Abstract In integrated circuits (ICs) system, computational performance is limited by its performance, and since the execution time is dominated by the multiplication factor because of that high speed multiplier is much more important in DSP applications. This paper presents an efficient multiplier design using FinFET technology using gate diffusion input (GDI). The quantum-based threshold voltage has been significant when the FinFET is scaled to sub 18 nm; these effects are modeled by solving the transistor channel with the change of material parameters. Adder is a fundamental block in numerous digital circuits, fluctuations degrading the device performance with the aim that altering FinFET design parameters to get rid of variations. The proposed FinFET Vedic GDI adder is analyzed using physics-based hot carrier effect/compact model for prediction of FinFET performance degradation due to the interface state, to reduce cost and time-time market in advanced process development. This design uses less number of transistors than the conventional CMOS-based multiplier designs. By employing the FinFET-GDI technique, there is an extensive decrease in power and delay of the circuit. Design also solves the threshold drops problem of original GDI cell resulting in better output signals. The proposed methodology is implemented in tanner tools using TSMC library. Design technology co-optimization (DTCO) obtains high speed, low power and reliability of practical MOSFETS. The simulation results show the desired accuracy, speed, power and reliability of proposed adder structure by lowering the fluctuations of digital circuits.

Keywords Fluctuations · Reliability · DTCO · Process variations · FinFET · Vedic multiplier · GDI · Full adder

S. K. Dilshad (✉) · G. S. K. Santosh

Department of ECE, KL Deemed to be University, Guntur, Andhra Pradesh, India

G. S. K. Santosh

e-mail: gksantosh17@kluniversity.in

1 Introduction

In present generation, the growth of integrated circuit devices has increased a lot. The VLSI applications are given as digital signal processing and microprocessors. These applications are most widely used to perform arithmetic operations. Along with that multiplication, subtraction operations are also performed. By using some modules, all these operations are performed on the 1 bit full adder circuit. Approximate adders are implemented in a way to balance the trade-off between accuracy V_s performance/power. The areas of image and video processing appear as good case studies for the use of approximate results. The circuit design is addressed at different levels. The design parameters of the circuit produce good trade-off in terms of speed and area [1]. Multiplication has been given careful consideration by utilizing analysts, because of the reality expansion which is essentially bitwise activity among two subject components and the additional complex tasks, reversal and might be finished with a few multiplications. Montgomery's multiplier is classified into three types; they are bit-serial, bit-parallel and digit-serial architectures. Bit-parallel shape is rapid; however it is far steeply priced in phrases of vicinity. Bit-serial structure is region efficient, but it is far too sluggish for plenty packages. The digit-serial structure is flexible which may change the space and velocity; consequently, it achieves a moderate pace, reasonable price of implementation, and hence, it is most appropriate for practical use. Montgomery presented a technique for figuring modular multiplication productively. He introduced to move the portrayal of numbers from the Z_n to an alternate area called Montgomery Residual portrayal or Montgomery Domain [2, 3].

Here, for the purpose of security, the computers and communication system brought with a demand from private sector [4]. The Montgomery multiplication is the calculation that permits effectively for registering. The expense of the particular duplication is equivalent to three whole numbers which increases in addition to the expense of the change in the Montgomery area. Yet, in the event that the large-scale task is an exponentiation at point the change cost is insignificant contrasted with the quantity of augmentations executed in the Montgomery area. In the process of Montgomery multiplication, preprocessing unit and post-processing units are used [5]. The preprocessing unit produces N-Residue operands, and in the same way, post-processing unit will eliminate the constant factor $2n$. Hence, to form N-Residue operands in the system, modular exponentiation is used.

The adoption of lower technology nodes has put the VLSI industry in a position to deal with various challenges. When compared to planar CMOS devices, the deployment of footed quasi-resistance (FinFET) has increased the power and performance characteristics but by creating new technical and collaborative challenges such as device types, material physics, electro migration, lithography process and double patterning. The gate of FinFET device is wrapped around the drain-source which is a conducting channel. This yields significant improvement in performance by providing better electrical properties, low threshold voltages, increased electrostatic control, higher density and reduced dynamic and leakage power. FinFET can easily

be replaced by the bulk CMOS devices due to its benefits like high ON and low OFF currents, intra-die variability, low power consumption, high integration density and reduce in short channel effects. The VLSI design cycle primarily comprises two design stages: frontend design and the backend design.

Rest of the paper is organized as follows. Sect. 2 gives the detailed analysis of conventional methods and its problems. Section 3 gives the detailed analysis of proposed Vedic multiplier with respect to the various block level operations using FinFET full adder formulations. Section 4 gives the detailed analysis of Xilinx ISE-based software simulation of the proposed method with area, power and delay analysis with simulation waveforms and comparative analysis with various literatures. Section 5 concludes the paper with possible future studies.

2 Literature Survey

The foremost step in physical design is the floor plan. Floor plan stage involves relative placement of the macros, sub-blocks, IPs, I/O pads and pins. It also specifies the aspect ratio, die size, shape and core utilization area of the chip. Soft, hard or partial placement blockages are added to reduce the density in specified areas [1]. This stage includes creating prototype that uses multiple iterations of floor plan to be performed to obtain less congestion, wiring length and with the focus on routability. This stage is followed by power planning where the power grid (PG) network is built by adding power rings and stripes to allocate the power supply to each component of the chip [3, 4]. The next stage is the placement which is a process of placing the standard cell instances on the die. The purpose of placement stage is to place the cells in design without overlap, minimize the area of layout and reduce the timing of critical nets. The register grouping can be done after placement stage to reduce area and power [5]. The next step is to deliver the clock signals equally through clock tree synthesis (CTS) to all sequential components present in the system. The aim of CTS is to control the skew, insertion delay and optimize clock and data path to achieve better power, performance and area (PPA) [6]. The CTS is followed by routing where the precise paths for each net are connected through metal layers. The routing is carried out by considering logical connectivity well-defined in the net list without violating the rules of design [7]. The interconnection of instances, i.e., routing, takes place through two steps: first the global routing and later the detailed routing. In a design, there should be n-well continuity, so filler cells, i.e., a type of special cell used in physical design (PD), are added to avoid the gaps [8, 9]. Few other special cells such as tie high and tie low cells have its usage in design to sustain the constant V_{DD} and V_{SS} , respectively. Tap cells are used to avoid latch up problem, end cap cells avoid cell damages at the end of the row, decap cells act as charge reservoir to resolve dynamic IR drop, spare cells are used at engineering change order (ECO) stage. The aim of post layout optimization/verification is to reduce delay, perform DRC, LVS checks, fix the violations, check for timing constraints, fix setup and hold violations and minimize the effects of IR drop and electro migration.

To perform timing closure with no violation buffer sizing, V_T swapping, increasing drive strength, insertion of buffers or repeaters and few other techniques are carried out. The design is ready for fabrication, after layout and verification. Traditionally, the layout data to be fabricated is sent on a tape. Thus, this data release event is called as tape out. The VLSI chips manufacturing process consists of sequential steps such as lithography, etching, deposition, chemical mechanical polishing, oxidation, ion implantation and diffusion. The correctness and efficacy of the test are most necessary for quality goods. In a fabrication center, the fabrication of wafer takes place which is then transformed into individual chips. Every chip is afterward packaged and tested to make sure it follows all the specific model requirements and works as intended. The semiconductor industry relies profoundly on EDA, and scripting forms an essential part of EDA. Perl is the most common approach for this job, but TCL is also widely used.

3 FinFET-Based GDI

To design low power circuits is highly challenging, and one does not know whether the timing constraints can be satisfied until the final routing has completed. In deep submicron (DSM) processes of FinFET, the most influential on consumption of total power is the leakage power. Depending on unique design and application, devices have different GDI goals, and one of these parameters is always a tradeoff. From RTL to signoff stage, multiple optimization techniques are performed—upscale, downsize, V_T swap and many more, and all these factors impact or tweak design PPA. In a VLSI IC, electrical energy during circuit operations is converted to heat energy, and the rate at which this energy is removed from the source and converted into heat is known as power dissipation. To avoid the increase in chip temperature, heat energy must be dissipated from the chip which otherwise can affect the circuit leading to its failure. The two major sources of power dissipation are the static power and dynamic power dissipation. The dynamic power dissipation is mainly due to switching activity of the capacitance and short circuit current, while static power dissipation is because of sub-threshold, gate and junction leakage currents.

FinFET-GDI technique was mainly focused on reduction of power consumption, area and complexity of digital combinational circuits. FinFET-GDI uses a CMOS inverter circuit to implement various complex logic functions listed in Table 1 using only two transistors based on the inputs given to G, P and N nodes. The transistors

Table 1 Modified-GDI cell logic functions

(G)	(P)	(N)	Outputs
A	B	B^\dagger	$A^\dagger B + AB^\dagger$
A	0	B	AB
A	1	0	A^\dagger
A	B	C	$A^\dagger B + AC$

PMOS and NMOS will share a single gate terminal input G, while N and P are inputs for NMOS and PMOS source terminals, respectively, as shown in Fig. 1a.

However, the drawback of the original FinFET-GDI cell was dropping of threshold which tends to reduction in performance and increases static power dissipation. Improving this, FinFET-GDI approach was proposed [2] by connecting the PMOS and NMOS transistors bulk nodes to supply voltage and low constant voltage, respectively, shown in Fig. 1b.

Using the modified-GDI cell, we can implement AND gate and XOR gates by giving input to respective nodes in M-GDI cell as per Table 1. Here, Fig. 2a shows

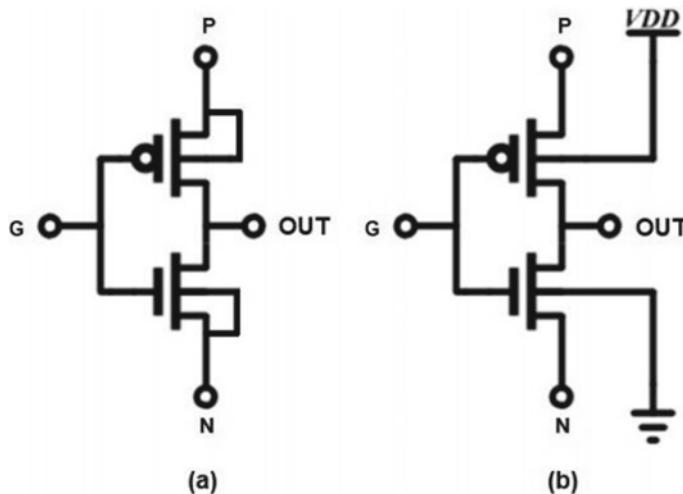


Fig. 1 a GDI cell, b FinFET-GDI cell

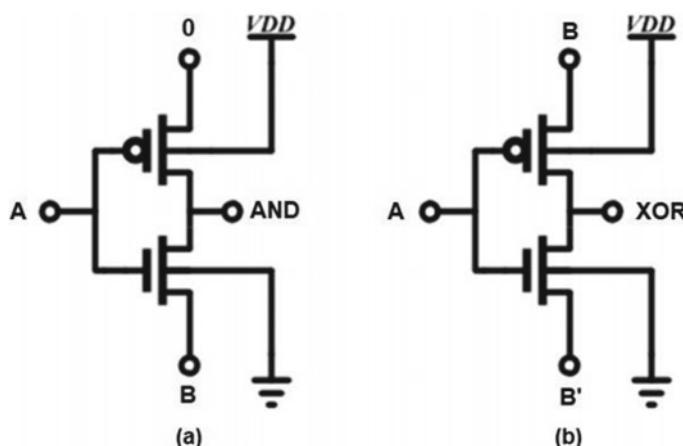


Fig. 2 a AND gate using FinFET-GDI, b XOR gate using FinFET-GDI

the AND gate, and Fig. 2b shows XOR gates designed using M-GDI.

4 Multiplier Using FINFET-GDI

The multiplier is the essential building block in the DSP processor DIP processor, ALUs and all the types of integrated circuits. Thus, the efficient design of multiplier makes the design to area, loop- = [']\delay and power efficient. Multiplier mainly focuses on speed. It is much faster than the conventional multipliers, which are the aspect most sought in today's world. By further incorporating GDI technique, the amount of transistors used are scaled down and improve the speed even further. Using GDI cells leads to unintended threshold drops. To resolve this problem, we implemented the multiplier using FinFET-GDI technique. Consider the two 2-bit inputs as (A_1, A_0) and (B_1, B_0) . First, the LSB of 2 binary numbers is vertically multiplied. The LSB of partial product result is taken as output. Secondly, 2 binary bits are multiplied crosswise, result is added to carry generated previously, and LSB of partial product result is taken as output. At last, 2 binary bits MSB are multiplied vertically and appended to the preceding carry. The resultant 2 binary bits are taken as result. The output of the Vedic multiplier is given by following equations

$$S_0 = (A_0 \times B_0) \quad (1)$$

$$S_1 = (A_1 \times B_0) + (A_0 \times B_1) \quad (2)$$

$$S_2 = (A_1 \times A_0 \times B_1 \times B_0) + (A_1 \times B_1) \quad (3)$$

$$S_3 = (A_1 \times A_0 \times B_1 \times B_0) \quad (4)$$

Consider that (A_1, A_0) and (B_1, B_0) are input bits and S_0, S_1, S_2 & S_3 be output bits. From Eq. 1, S_0 will be in logically high state when both A_0 and B_0 input bits are high. From Eq. 2, S_1 will be logically high when one of either partial products $(A_1 \times B_0)$ or $(A_0 \times B_1)$ are logically high but not both. From Eq. 3, S_2 will be logically high either when all the input bits are logically high or when partial product of A_1 and B_1 is high. Finally, from Eq. 4, S_3 will be logically high only when all the input bits are high.

The 2-bit FinFET-based multiplier uses two half adders and four AND gates to give a 4-bit output which is the product of two 2-bit inputs. The bit FinFET-based block diagram of the Vedic multiplier is shown in Fig. 3. This 2-bit multiplier can be used as building block of 4-bit and further higher-order Vedic multipliers. AND gates and half adders are used for generating product and addition of the generated partial products. Half adder used in Vedic multiplier can be implemented in various methods such as using multiplexers, transmission gate logic (TGL), GDI technique

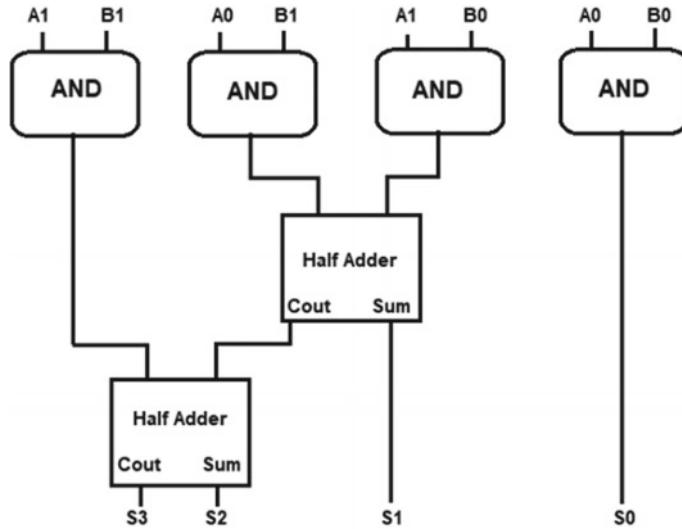


Fig. 3 2-bit FinFET-GDI-based Vedic multiplier using URDHVA-TRIYABHYAM

and pass transistor logic (PTL). Using GDI or modified-GDI technique to implement half adder significantly improves the area and power efficiency.

The use of FinFET-GDI technique lets us simplify complex designs and scales down the amount of transistors used in the circuit. Full adder implementation using FinFET-GDI-based 4-Transistor XOR gate and AND gate is shown in Fig. 4. FinFET-GDI-based half adder consists of six transistors. XOR gate used in the half adder uses the four transistors, and full adder consists of ten transistors. 4-Transistor XOR design is implemented by taking an inverter circuit, common gate node is given with first input signal (A), PMOS source node is given with second input signal (B), and NMOS source node is given with complement of second input signal (B'). To generate complement of signal B, a NOT gate is used. AND gate is implemented based on GDI technique where inverter common gate node is used as first input signal (A) port, and PMOS, NMOS source nodes are used as second input (B) signal port.

Figure 5 represents the N-bit FinFET-GDI Vedic multiplier utilizing these GDI AND gates and N-bit RCA adders developed by FinFET-based full adders.

Step 1: Every multiplier consists of partial products as fundamental outcomes. The partial products are generated by performing the bitwise AND operation between A inputs to the B inputs in a chronological order. Here, the FinFET-GDI AND gate will be utilized to perform the AND operation as mentioned in Fig. 2a. For implementing N-bit multiplier, we require N^2 number of AND operations so that of FinFET-GDI AND gates.

Step 2: After developing the partial products, the partial products are grouped together to perform the multiplication operation, respectively.

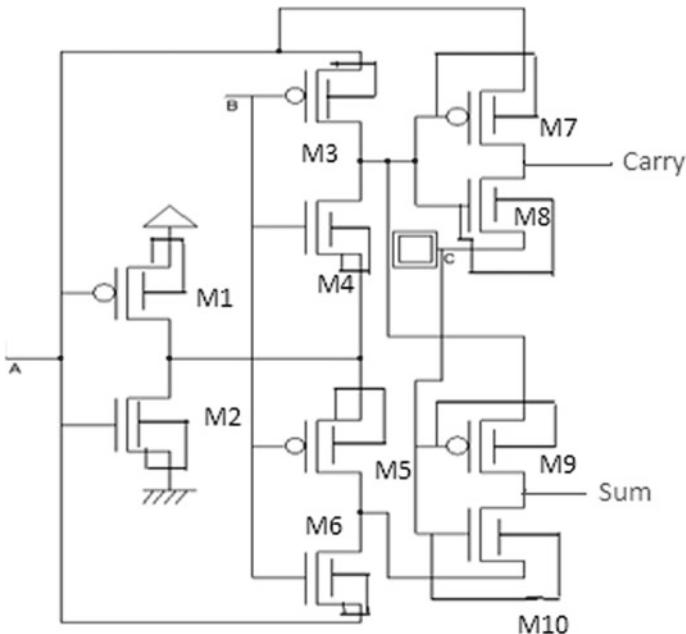


Fig. 4 Full adder using FinFET-GDI

Step 3: Fig. 6 shows the detailed analysis of partial products addition for N-bit multiplier. For developing the N-bit multiplier, we require N-1 number of adder with each of N-bit.

- For first N-bit RCA, the input will be M_1 , M_2 , and Cin become zero, which generates S_1 as its sum output and C_{out_1} as its carry output.

$$M_1 = \{1'b0, \dots, PP_{0,N-1}, PP_{06}, PP_{05}, PP_{04}, PP_{03}, PP_{02}, PP_{01}\};$$

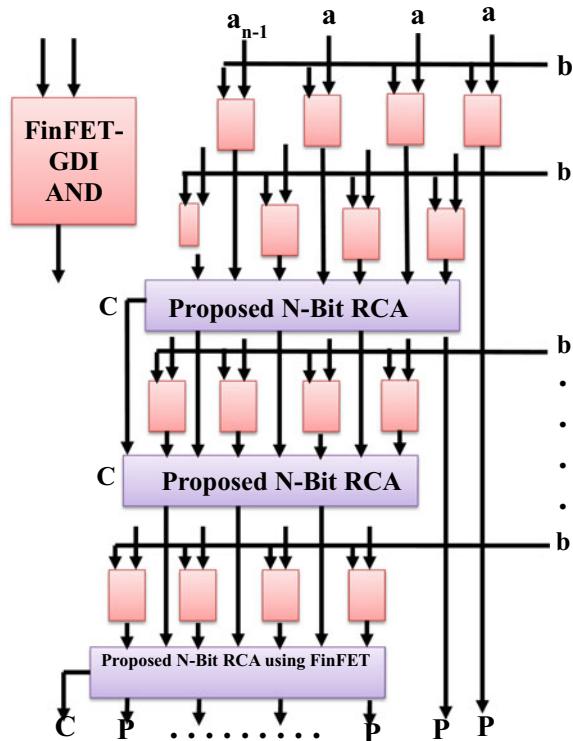
$$M_2 = \{PP_{1,N-1}, \dots, PP_{16}, PP_{15}, PP_{14}, PP_{13}, PP_{12}, PP_{11}, PP_{10}\};$$

$$S_1 = \{C_{out_1}, S_{1,N-1}, \dots, S_{16}, S_{15}, S_{14}, S_{13}, S_{12}, S_{11}, S_{10}\};$$

- For the second stage onwards, the addition operation will maintain same structure of inputs until $N - 1$ th stage. For an instance, inputs for adder 2, S_1^* = and M_3 becomes inputs and S_2 becomes output. S_1^* is generated from S_1 , but it does not contain LSB bit.

$$S_1^* = \{C_{out_1}, S_{1,N-1}, \dots, S_{16}, S_{15}, S_{14}, S_{13}, S_{12}, S_{11}\};$$

Fig. 5 Proposed FinFET-GDI-based N-bit Vedic multiplier



A_7	A_6	A_5	A_4	A_3	A_2	A_1	A_0
B_7	B_6	B_5	B_4	B_3	B_2	B_1	B_0
<hr/>							
PP_{07}	PP_{06}	PP_{05}	PP_{04}	PP_{03}	PP_{02}	PP_{01}	PP_{00}
PP_{17}	PP_{16}	PP_{15}	PP_{14}	PP_{13}	PP_{12}	PP_{11}	PP_{10}
PP_{27}	PP_{26}	PP_{25}	PP_{24}	PP_{23}	PP_{22}	PP_{21}	PP_{20}
PP_{37}	PP_{36}	PP_{35}	PP_{34}	PP_{33}	PP_{32}	PP_{31}	PP_{30}
PP_{47}	PP_{46}	PP_{45}	PP_{44}	PP_{43}	PP_{42}	PP_{41}	PP_{40}
PP_{57}	PP_{56}	PP_{55}	PP_{54}	PP_{53}	PP_{52}	PP_{51}	PP_{50}
PP_{67}	PP_{66}	PP_{65}	PP_{64}	PP_{63}	PP_{62}	PP_{61}	PP_{60}
PP_{77}	PP_{76}	PP_{75}	PP_{74}	PP_{73}	PP_{72}	PP_{71}	PP_{70}
<hr/>	P_{15}	P_{14}	P_{13}	P_{12}	P_{11}	P_{10}	P_9
					P_8	P_7	P_6
						P_5	P_4
						P_3	P_2
						P_1	P_0

Fig. 6 Operation of Vedic multiplier

$$M_3 = \{P P_{2,N-1}, \dots, P P_{26}, P P_{25}, P P_{24}, P P_{23}, P P_{22}, P P_{21}, P P_{20}\};$$

$$S_2 = \{C_{\text{out}_2}, S_{2,N-1}, \dots, S_{2,6}, S_{2,5}, S_{2,4}, S_{2,3}, S_{2,2}, S_{2,1}, S_{2,0}\};$$

Finally, inputs for $N - 1$ th stage adder S_{N-2}^* and M_N are inputs, and P becomes output.

$$S_{N-2}^* = \{C_{\text{out}_{N-2}}, S_{N-2,N-1}, \dots, S_{N-2,6}, S_{N-2,5}, S_{N-2,4}, S_{N-2,3}, S_{N-2,2}, S_{N-2,1}\};$$

$$M_N = \{P P_{N-1,N-1}, \dots, P P_{N-1,6}, P P_{N-1,5}, P P_{N-1,4}, P P_{N-1,3}, P P_{N-1,2}, P P_{N-1,1}, P P_{N-1,0}\};$$

$$\text{Output } P = \{C_{\text{out}_{N-1}}, \dots, S_{N-1,6}, S_{N-1,5}, S_{N-1,4}, S_{N-1,3}, S_{N-1,2}, S_{N-1,1}, S_{N-1,0}\};$$

5 Results and Discussion

The simulation has been carried out in tanner EDA. The technology used is TSMC process 18 nm technology. All the proposed designs have been programmed and designed using tanner EDA software; this software tool provides the two categories of outputs named as simulation and synthesis. The simulation results give the detailed analysis of proposed design with respect to inputs, output byte level combinations. Through simulation analysis of accuracy of the addition, multiplication process estimated easily by applying the different combination inputs and by monitoring various outputs. Through the synthesis results, the utilization of area with respect to the programmable logic blocks (PLBs) and lookup tables (LUT) will be achieved. The time summary with respect to various path delays will be obtained, and power summary is generated using the static and dynamic power consumed.

From Tables 2 and 3, it is observed that the proposed **FINFET-GDI multiplier** is area, power and delay efficient compared to the literatures CMOS [8], GDI [1] and PTL [7].

6 Conclusion

This paper presents an efficient multiplier design in FinFET-GDI 18 nm technology on tanner EDA. This design uses less number of transistors than the conventional Vedic multiplier design. By employing the FinFET-GDI technique, there is an extensive decrease in power and delay of the circuit. There is an improvement of 30% in power and 28% in delay compared to N-bit multiplier using GDI technique. This design also solves the threshold drops problem of original GDI cell resulting in better output signals.

Table 2 Performance measurement of various multiplier techniques

Multiplier bit length	Parameter/method	No. of transistors	Nodes	Delay (nS)	Power (μ W)
8-bit	CMOS [8]	191	358	28	82
	GDI [11]	249	463	25	132
	PTL[12]	322	596	48	114
	FINFET-GDI	70	126	14	54
16-bit	CMOS [8]	308	801	44	98
	GDI [11]	433	928	38	226
	PTL[12]	332	1281	68	146
	FINFET-GDI	180	320	21	69
32-bit	CMOS [8]	1081	1602	116	143
	GDI [11]	1362	2395	169	260
	PTL[12]	1289	3283	229	280
	FINFET-GDI	562	1001	45	90

Table 3 Comparison of proposed multiplier with different length of bits

	64-bit	128-bit	256-bit
#LUTs	826	1280	1502
#Slices	1397	1670	1943
Time (ns)	58.78	73.936	85.842
Power (μ W)	127	161	186

References

- Morgenshtein A, Fish A, Wagner IA (2002) Gate-diffusion input (GDI): A power-efficient method for digital combinatorial circuits. *IEEE Trans Very Large Scale Integer Syst* 10(5):566–581
- Patel CR et al (2020) Inverted gate Vedic multiplier in 90nm CMOS technology. *Am J Electr Computer Eng* 4(1):10–15
- Sahu SR, Bhoi BK, Pradhan M (2020) Fast signed multiplier using Vedic Nikhilam algorithm. *IET Circ Devices Syst*
- Masurkar E, Dakhole P (2016) Implementation of optimized Vedic multiplier using CMOS technology. In: International Conference on Communication and Signal Processing (ICCPSP 2016), pp. 840–844
- Krishna BV et al (2020) Analysis of Vedic multiplier for conventional CMOS, complementary pass transistor logic (CPL) & double pass transistor logic (DPL) logics. *PalArch's J Archaeol Egypt/Egyptol* 17(7):5649–5656
- Lachireddy D, Ramesh SR (2020) Power and delay efficient ALU using Vedic multiplier. In: Advances in electrical and computer technologies. Springer, Singapore, pp 703–711
- Saritha P et al (2020) 4-bit Vedic multiplier with 18nm FinFET technology. In: 2020 International conference on electronics and sustainable communication systems (ICESC). IEEE
- Bianchi V, De Munari I (2020) A modular Vedic multiplier architecture for model-based design and deployment on FPGA platforms. *Microprocessors Microsyst* 103106

9. Sivanandam K, Kumar P (2019) Design and performance analysis of reconfigurable modified Vedic multiplier with 3-1-1-2 compressor. *Microprocessors Microsyst* 65:97–106
10. Safoev N, Jeon J-C (2020) Design and evaluation of cell interaction based Vedic multiplier using quantum-dot cellular automata. *Electronics* 9(6):1036

Classified Authentication System with IoT and Dashboard



Chilakala Sudhamani, K. Bharath Kumar, G. V. Hari Prasad, and N. Renuka

Abstract There has been significant concern about conceptual security in both family units and workplaces. In this digital era, conceptual security is replaced by Internet of things (IoT). The proposed authentication system is composed of NI myRIO embedded controller with the proximity detection sensor, which senses the presence of the person in front of the device setup. After sensing a person in front of the system, GSM module is utilized to send the unique OTP to the owner of the system for accessing; we will provide an option of entering the OTP through the keyboard. Once after entering the password, on board button on NI myRIO has to be pressed to make an entry of the password into system. If password matches with master database OTP, then the authentication will be provided to the end user along with LED glowing status, and motor rotation will be turned ON to open the door. If the password does not match with the master database password, then the buzzer will be blowing, which indicates the authentication failure or access denied. In the above two cases, the camera would capture the image of a person for future use, and it is stored in the NI myRIO. The whole data is updated to cloud through Dweet IoT service.

Keywords Internet of things · Security · Automation

C. Sudhamani (✉) · K. Bharath Kumar
CMR Technical Campus, Hyderabad, Telangana 501401, India

K. Bharath Kumar
e-mail: bharathkumar.ece@cmrtc.ac.in

G. V. Hari Prasad
CMR College of Engineering and Technology, Hyderabad, Telangana 501401, India
e-mail: drgvenkatahariprasad@cmrcet.org

N. Renuka
RVR and JC College of Engineering, Guntur, Andhra Pradesh 522019, India
e-mail: nrenuka@rvrjc.ac.in

1 Introduction

Digital India concept was proposed by our Prime Minister Narendra Modi. Nowadays, everyone is interested in automation of their surroundings like automatic driving, automatic ON and OFF of lights, washing machine, rice cooker, door opening etc. So, we considered security as a major problem in this automation. Security in today's lives play a very vital role in authenticating the data or belongings. Authentication system is defined as the process of recognizing the user's identity. Therefore to enhance the security of an entrance system or accessing system, an authentication system using sensors and microcontrollers is proposed. In any authentication system, sensors and controllers are used at the transmitter side and monitors, and buzzers are used at the receiver side. The main purpose of these sensors is to provide a secured authentication system for a home or office.

Authentication to offices or homes can be provided by several controlling devices like Arduino, Global Positioning System (GSM), Internet of things (IoT), radio frequency identification (RFID), and LabVIEW software. GSM-based door lock system was proposed to control the open and closing functions of a door by a simple short message service (SMS) [1, 2]. In this paper, authors used Arduino microcontroller for controlling the transmitter and receiver sensors. IoT-based smart door lock system is proposed to authenticate the commercial offices and homes [3–5]. IoT-based authentication systems are used to control the door locking systems remotely also. A smart door lock system is a system which uses digital password for opening and closing the door.

In [6], a locker with high security system based on RFID, fingerprint, password, and GSM technology is connected with the LPC2148 microcontroller. In [7], authentication system is designed using IR sensor. This IR sensor works by using select-light sensors to detect the selected light wavelength. In [8], RFID-based authentication system along with 8051 microcontroller and GSM module is proposed. In [9], authors describe about the various modern methods used for the smart entrance system such as graphical password system with one-time password (OTP), in which there are three levels of authentication: text authentication (LEVEL-1), image authentication (LEVEL-2), and OTP authentication (LEVEL-3).

In this paper, we are using advanced technology of LabVIEW software along with NI myRIO as a controller, and it is enabled with Dweet IoT technology. In the proposed authentication system, we are using OTP-based authentication, and for facial recognition, camera is installed. It provides a secured and safe door entrance system. In this system, an incoming request with a set of credentials is identified if any person is trying to open the door. These credentials are compared with the database of the authorized users information and then gives an access. Hence, the proposed system ensures the authenticated persons to access the device through OTP, and all the data can be updated to cloud through Dweet IOT service and can be viewed as a dashboard.

Rest of the paper is organized as follows: Classified Authentication System in Sect. 2, flowchart in Sect. 3, results in Sect. 4, and conclusion in Sect. 5.

2 Proposed Model—Authentication System

The components used are laptop with LabVIEW Software, NI myRIO, IR sensor, LEDs, motor, motor driver, buzzer, GSM module, and camera. Here, the NI myRIO acts as embedded microcontroller which takes the input from IR sensor and uses GSM module to send the OTP, and output is indicated through red and green LEDs along with the motor and buzzer. Camera is also interface with NI myRIO for clicking the pictures of end user. Finally, Dweet IOT is used as an online dashboard to see the last few transactions (Fig. 1).

2.1 MyRio

NI myRIO is introduced by NI (national instruments) which is a type of microcontroller, and it is being preferred because of its faster data acquisition. It can operate at a frequency of 667 MHz. It is being programmed by using NI LabVIEW software. It consists of analog and digital input and also output channels. In our project, we use myRIO to analyze the inputs and to produce the outputs according to the inputs. NI myRIO takes 5V power supply as source of power. One end of the USB from GSM is connected to myRIO to transfer the data, i.e., to communicate with it. An additional slot on myRIO is connected with the camera. It works as transmitter and also as receiver with GSM module. The hardware of NI myRIO is shown in the Fig. 2.

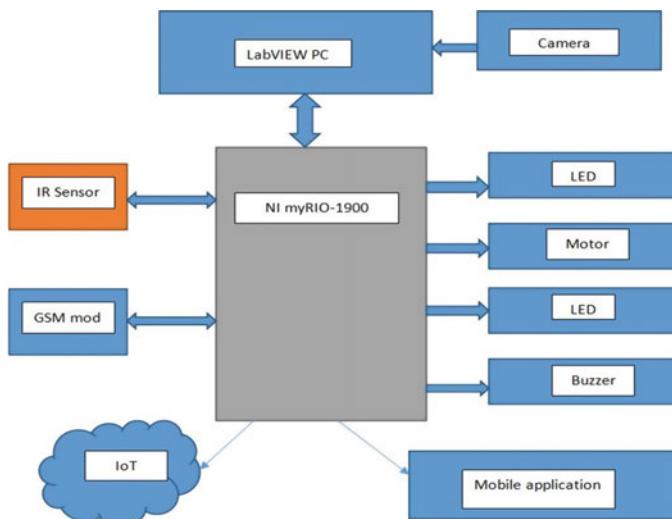


Fig. 1 Classified authentication system—block diagram

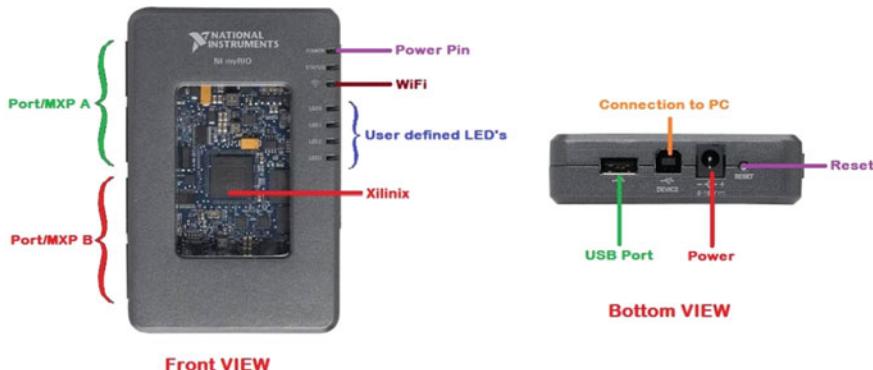


Fig. 2 NI myRIO—hardware representation

In this proposed system, myRIO takes inputs from power supply, IR sensor, GSM model and analyzes the inputs and produces the outputs through buzzer and led.

2.2 IR Sensor

IR sensor is an electronic sensor. It is used to measure the infrared light coming from an object. Every object radiates some infrared light which will be detected by the IR sensor. It detects the objects by measuring the heat radiated from the respective object. We can use IR sensor as transmitter and also as receiver, which is shown in Figure 3a. It can measure the obstacle in a maximum range of 30ft. It has its frequency range in between 300 GHz and 400 THz. It consists of LED and photo-diode. IR sensor is connected to analog input channel of myRIO. In classified authentication system, the IR sensor is used to detect whether the person is present or not in front of the

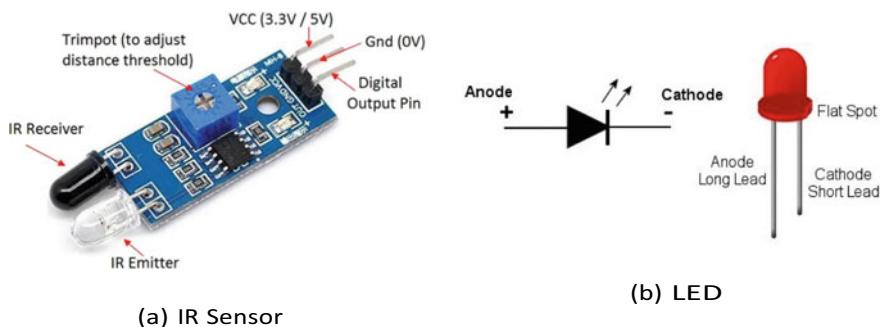


Fig. 3 a IR sensor, b LED

door. Here, it detects the person as an obstacle. Here, IR sensor detects obstacles and sends the information to myRIO.

If obstacle is not detected, then we get the output which states to approach the sensor to get the authentication codes. If the obstacle is detected, then myRIO continues to the next process. Hence, we could say IR sensor is used for the initial processing of authentication data.

2.3 LED

Light emitting diode (LED) is mainly used to emit the visible light, which is shown in the Figure 3b. LED's are used to emit light when they are charged with electric current. The emitted light is in different visible colors. The color emitted by them depends on the wavelength of light which is emitted. LED lighting differs from source to source because each source does not provide same amount of energy to it. LEDs are connected to the digital output channels in myRIO.

In this system, LEDs are used to indicate the successful and unsuccessful authentication. If a successful authentication happens, then LED glows in green color, and for an unsuccessful authentication, LED glows in red color.

2.4 Motor

An electronic motor is a device which converts the electrical energy into mechanical energy and is shown in Figure 4a. This motor works on the principle of “Fleming left hand Rule.” The basic working principle of a DC motor is whenever a current carrying conductor is placed in a magnetic field, it experiences some mechanical force. According to Fleming’s left hand rule, direction of this force can be decided, and its magnitude is given by the formula $F = BIL$, where B = magnetic flux density,



(a) DC Motor



(b) Motor Driver

Fig. 4 a DC motor, b motor driver

I = current, and L = length of the conductor within the magnetic field. In this system, DC motor is used in order to open the locker shaft if the access is granted.

2.5 Motor Driver Circuit

Motor driver is the breakout board which consists of a motor driver IC. Motor driver is used for the running of the motor which acts as a current amplifier. The motor driver acts as amplifier switch between the circuit and the motor. It takes the low current input signal and amplifies it and supplies the high power signal to motor in order to drive it. In this system, motor driver is used to drive the motor efficiently, which is shown in Figure 4b.

2.6 Buzzer

A buzzer is basically an active and passive device. It a small and an efficient component which add sound features such as buzzing or beeping noise. It consists of two pins which could be used an indicator, which is shown in Fig.5(a). It is considered as an audio signaling device with an integrated structure of electronic transducers. The buzzer could be either active or passive. Basically, the buzzer is made up of piezo electric material and a vibration plate along with the electrode for conduction. It works on principle of piezo electric effect and reverse piezoelectric effect. In this system, we are using buzzer in order to indicate the invalid access. Whenever the entered secret pin is wrong or validation is unsuccessful, then the buzzer turns ON.

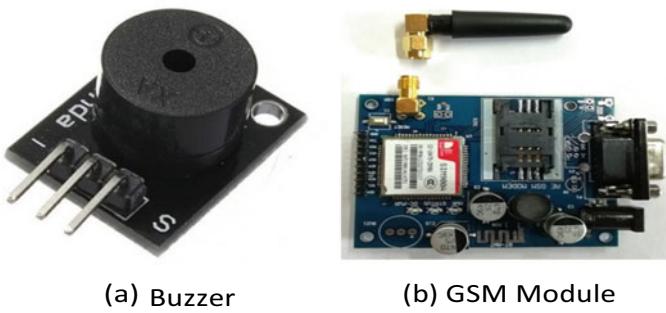


Fig. 5 **a** Buzzer, **b** GSM module

2.7 GSM Module

GSM module is a hardware device that uses mobile telephone technology which will provide a data link to a remote network, which is shown in Fig. 5b. From the view of mobile phone network, they are essentially identical to an ordinary mobile phone, including need for SIM to identify them to network. In this system, GSM is used to transfer the obtained secret pin to the owner of the system for granting the permission to the end user.

3 Classified Authentication System—Flow Chart

Work flow of a classified authentication system is shown in Fig. 6. This flow chart explains how an authentication system works step by step.

- Initially, IR Sensor will sense the person who is in front of the door. If the person is not detected, then the system will stop functioning.
- Whereas, if the obstacle is detected, then the GSM module is triggered, and it sends the unique OTP to the owner of the system.

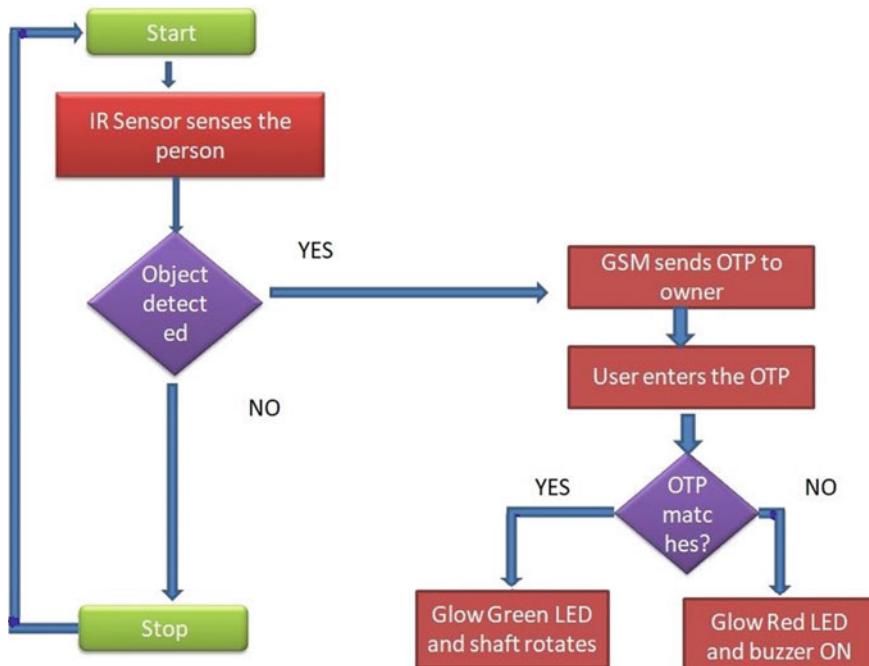


Fig. 6 Flow chart of classified authentication system

- This OTP is shared to the person who is accessing the system by the owner, and it is entered by the person using a keypad.
- This entered OTP is compared with the master database OTP and then gives the output.
- If the OTP entered does not match with the master OTP, then the RED LED glows along with the buzzer sound.
- If the OTP is matched, then the GREEN LED glows and the motor shaft.

4 Results

Classified authentication system is implemented using NI LabVIEW software and hardware. NI myRIO is used as a basic component for controlling the transmitter and receiver devices. LabVIEW software is used to write a program code for activating all devices and sensors and also to verify the output before implementing using a hardware devices. The software implementation for authentication system is shown in Fig. 7a, and hardware implementation using myRIO is shown in Fig. 7b. This part is called as a front panel. In this front panel, right-half is used to implement the code and left-half is used to check the outputs.

In this proposed authentication system, we are using IR sensor as a transmitter device to identify the obstacles in front the door. It is classified into two cases:

- CASE 1: Obstacle not detected: Whenever the person comes in front of the IR sensor, then it will detect the obstacle. If the obstacle size is less than 3 cm or the obstacle is not detected, then the message is displayed on the front panel as “Please Kindly approach the nearby sensor to get the authentication codes,” and the system is reset in order to function for the next case. Case 1 result in the front panel is shown in Fig. 8.
- CASE 2: Obstacle detected: If the obstacle is present in front of the IR sensor, then it detects the obstacle. If the obstacle is detected, then it will triggers the GSM module to generate a unique secret code. The secret code is generated by multiplying the random number with 10 power 6, and it will be shared with the owner. This secret pin will be shared by the owner to end user and is entered by the end user through the keyboard. System waits for 10–20 s for the process to

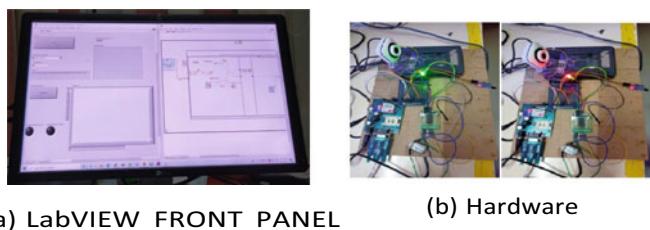
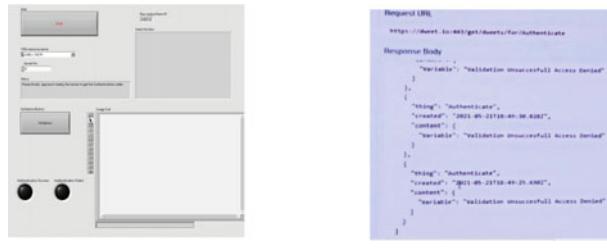


Fig. 7 **a** LabVIEW front panel, **b** hardware implementation using NI myRIO



(a) Obstacle not Detected (b) Last five trails stored in Cloud

Fig. 8 **a** Front panel when obstacle not detected, **b** Last five trails stored in cloud rotates and opens the system or unlocks it

complete and generate a validate message. If the person will press on the validate button, then both the conditions are checked across each other. Here, again, we have two cases:

1. Initially, the shared secret pin with the end user is entered by him through a keyboard and should press the validate button for which the system waits for 10 s to complete the process. If both the conditions are verified, i.e., the entered secret pin by the end user is same as the system generated pin and validate it by the user, then the green LED glows as well as the motor turns ON. At the same time, the camera clicks the image of the person accessing the system and is stored in the cloud and gives a message “Validation successful access granted” which is displayed on the front panel code along with the status validation button and image output, which is shown in Fig. 9a. After completion of the process, the whole system is reset for the further usage.
2. Initially, the shared secret pin with the end user is entered by him through a keyboard and should press the validate button for which the system waits for 10 s to complete the process. If the entered secret pin by the end user is not

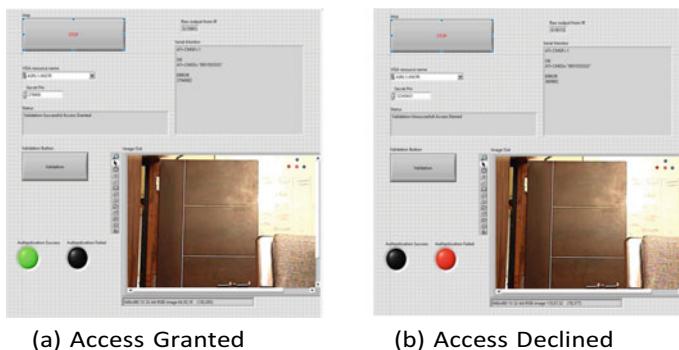


Fig. 9 Front panel when obstacle is detected and access granted **(a)** when obstacle is detected and access granted **(b)** when obstacle is detected and access declined

same as the system-generated pin, then RED LED glows. At the same time, the camera clicks the image of the person accessing the system and is stored in the cloud and gives a message “Access declined” which is displayed on the front panel code along with the status validation button and image output, which is shown in Fig. 9b. After completion of the process, the whole system is reset for the further usage.

We can store the images of a person who is trying to open the door in cloud data base using dweet.io. Figure 8b illustrates the output of dweet.io, and it helps us to read the last five cached dweets along with the time stamp and the status which could be very helpful.

5 Conclusion

The proposed “Classified Authentication system with IOT and Dashboard” can improvise the security of our belongings and data. With this system, a person can secure everything without his/her physical presence at that particular place and can give access only to those who are valid users. We can reduce the thefts, burglar effects to a large extent, and it can be implemented at offices, banks, homes etc.

Acknowledgements This work is supported by grant *no.SR/FST/College017/2017* under DST FIST program-2017, received by Department of Electronics and Communication Engineering, CMR Technical Campus, Hyderabad.

References

1. Najma A, Mohana B, Mohan Babu A (2019) GSM based door lock system. *Int J Innov Technol Exploring Eng* 8(6):229–231
2. Ibrahim A, Paravath A, Aswin PK, Iqbal SM, Abdulla SU (2015) GSM based digital door lock security system. In: 2015 IEEE International conference on power, instrumentation, control and computing, pp 1–6
3. Satoskar R, Mishra A (2018) Smart door lock and lighting system using internet of things. *Int J Computer Sci Inform Technol* 9(5):132–135
4. Sennan S, Srinivasan P (2018) Internet of things based digital lock system. *J Comput Theoretical Nanosci* 15(9)
5. Adiono T, Fuada S, Anindya SF, Irfan Ganj P, Fathany MY (2019) IoT-enabled door lock system. *Int J Adv Computer Sci Appl* 10(5):445–449
6. Raghu Ram G, Subramanian Sarma G (2013) Locker opening and closing system using RFID, fingerprint, password and GSM. *Int J Emerg Trends Technol Computer Sci* 2(2)
7. Ajmera P (2017) A review paper on infrared sensor. *Int J Eng Res Technol* 5(23)
8. Gandhimathi H (2019) RFID password and OTP based door lock system using 8051 microcontroller. *Int J Eng Res Technol* (2019)
9. Ishthaq Ahamed K, Manjunath M (2013) Implementation of multi level security system using click based authentication. *Int J Eng Res Technol* 2(8)

Waste Management Data Analytics and Solution for Domestic Waste Management to Improve Soil Quality



Suwarna Gothane, K. Srujan Raju, and B. Kavitha Rani

Abstract Solid waste management is crucial in India. In metropolitan India, nearly, 377 million population produces 62 million tons of municipal solid waste per annum, 70% is collected, 20% is treated, and 50% is neglected. With the current rate of increase in population, it will be incremental till 2025. Domestic house waste per person is also important parameter in metro cities. Exposed garbage every day results in contaminated land. Here, objective is to suggest an approach to develop an IoT-enabled waste management system to improve soil quality for agricultural activities and minimize human intervention helps to reduce human time and effort. Approach helps in control pollution and diseases caused due to environmental pollution. Here, we have performed data analytics using Weka tool and Microsoft Excel. We have used monthly-recycling-and-waste-collection-statistics data set for performing analytics and noticed contribution of curb waste maximum compare to other types of waste. There is a demand to provide technology for waste management.

Keywords Internet of things · Waste management · Data analytics

1 Introduction

Survey carried out in 2007 shows for Indian urban cities with a inhabitants more than a million estimate MSW composition by weight 41% biodegradable, 40% inert, 6% paper, 4% plastic, 4% textiles, 2% glass, 2% metals, and 1% leather. As per 2014 India Planning Commission MSW research, 51% of waste is organic, 32% is non-organic, and 17% is recyclable waste. India is called as trash bomb as 80% daily garbage remains uncovered. Thus, there is a problem of waste management. In

S. Gothane (✉) · K. Srujan Raju

Department of Computer Science and Engineering, CMR Technical Campus, Hyderabad 501401, India

e-mail: suwarnagothane.cse@cmrte.ac.in

B. Kavitha Rani

Department of Information Technology, CMR Technical Campus, Hyderabad 501401, India
e-mail: kavitharani.cse@cmrte.ac.in

future, at times polluted untreated waste at places level up to nearly three crore MT. Waste not taken care becomes poisons to environment. Issue of Municipal solid waste is noticed globally. To overcome unhygienic environment and increasing quality of living beings, waste management is the important task need to focus. To reduce organic waste and provide society better life, we are proposing development of IoT-enabled waste management system to improve soil quality for agricultural activities. This technique reduces organic waste and offers benefits to society to breathe in healthy and pollution less area in pleasant environment. Approach reduces air, water, and land pollution. It helps to boost up countries' economic development.

2 Review of Work

Bakhshi et al. worked on waste collection and management for urban environments. Author used IoT Raspberry Pi and ultrasonic sensors on waste-bins on few places for content checking and used machine learning technique analytics for monitoring status of bin. Approach increased fuel competence by 45% and reduced collection to 18% [1]. Jain et al. developed smart waste management system for solid waste collection efficiently. Waste checking is monitored using sensor, and data are sent to GSM/GPRS system. Designed application assists in information gathering on waste [2].

Jayaprakash et al. developed compost bin for Indian household kitchen. It holds separate place for compost starter with mixing blade, air filter setup, and a compost collection tray. Natural air purifier reduces awful smell and acts as a purifier. Users liked the approach with benefits for its features, portability, and ease of maintenance [3].

Ng et al. in UK conducted organic food waste management review, analysis, and focused on alternative waste management strategies. Author used gasification technology to transfer waste into electricity [4].

Gupta et al. identified the dimension of the garbage with the assistance of ultra-sonic sensor and sent information to the approved office for garbage collection through GSM framework interfaced through the Arduino microcontroller. A GUI is created to screen the ideal data identified with the garbage bins for various chosen areas. Through the GSM at control room messaged is shown using LCD and approved individual advise the drivers to gather the garbage on schedule for safer garbage collection [5].

Fujdiak et al. performed efficient-municipal waste collection with Internet of things (IoT), using genetic algorithm. The proposed approach works best for garbage-truck with simulations implemented on mentioned area [6]. Jaware et al. developed smart composting system using IoT with minimal cost infrastructure and without human intervention [7].

3 Technology Importance and Challenges

Food waste situation observed worst when overflowing of waste due to inefficient monitoring. Measure needs to take to reduce landfill waste harmful to environment and living beings. Composting domestic waste offers helpful for fertilizer use in gardening and agriculture and as cooking gas. Produced compost ward off plant disease and provides cooking gas for the society. It reduces loads of government work, time, human resources, and promises us with healthy and pollution free environment.

Environment growth and living beings are affecting as a result of inhabitants and changed practices. Telangana State Government Initiatives focused on Swachha Bharat Mission and solid waste management rules. In National Green Tribunal on Waste Management at New Delhi, several officials and ministry were participated. With a key focus on solid waste management, dump yard land where 8450 tons of waste generated in the state only 8223 tons was being collected; 95.9% collected from household. An action plan was formulated in tune with solid waste management rules, 2016. The meeting discussed preparation of detailed project reports for the municipalities, land acquisition for the dumping yards, setting up waste to energy plants, creating public awareness on waste burning, etc. Government of Telangana initiated Haritha Haram in large scale tree planting program to enhance the state greenery. In 2018, Telangana Government Chief Minister K. Chandrasekhar Rao gave an indication in banning single-use plastic. Telangana State Government driving constant efforts for environment protection.

The challenges noticed for organic waste management focus on

- Investigate plans on waste management—using IoT to receive waste and generation of information.
- Physical infrastructure of waste bin—for different types of waste.
- Process for correct disposal—using IoT technologies and considering RFID support.
- Different types of sensors—temperature, humidity, chemical, pressure, capacity weight.
- Software analysis—how information will be processed and used.

4 Proposed Architecture

Detail methodology of organic waste management is require to initiate since process starts from data detection, transmission, analysis, processing of collected data, and obtaining final result. Efficient methodology solution developed for organic waste management is represented in Fig. 1.

Detail project architecture for development of IoT-enabled waste management system to improve soil quality for agricultural activities is represented in Fig. 2.

Step 1: Collection of organic waste/data detection

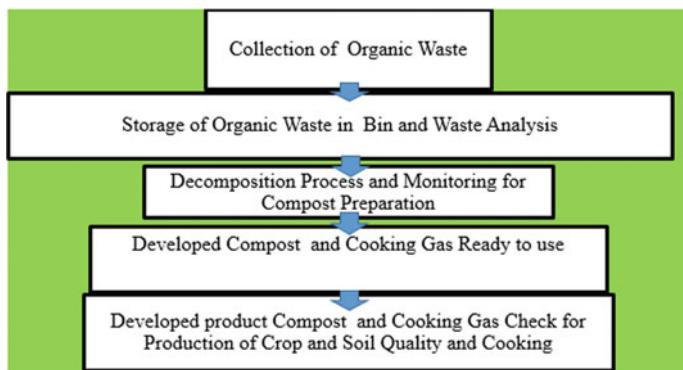


Fig. 1 Detail methodology organic waste management



Fig. 2 Proposed project architecture

Organic waste can be collected from daily household kitchen waste, school college canteen and gardens, vegetable markets, and corporate office. Waste can be food waste, paper, dry leaves from the garden, vegetable market leftover.

Step 2: Storage of organic waste in separate bins/data transmission

Different organic wastes separated in 4 categories of bins, namely Bin 1 contains plants waste, Bin 2 contains vegetable and fruits waste, tea waste, coconut waste, etc. Bin 3 contains leftover cooked food, ripped fruits, and stale vegetables, and Bin 4 contains paper waste.

Step 3: Waste analysis/data analysis

Collected waste further used mainly for 2 purpose, i)to feed it to cattles: leftover cooked food and ripped fruits and stale vegetables, ii)transfer it to compost bin for creation of compost.

Step4: Developed product /processing of collected data

In compost bin process, waste undergoes through process of creation of compost by decomposing for few days and accelerating decomposition process to obtain faster final product gas and compost. Bin will use IoT technology for understanding composition of compost such as to identify carbon and nitrogen in kitchen waste. Proper composition is considered in our approach to perform faster decomposition process. Composition process starts with different sensors such as temperature sensor and humidity sensor. IoT alerts when compost is ready to use with specification of product. Final product is obtained is cooking gas and manure.

Step 5: Usage of product obtaining final result

Developed product compost can be tested for crops to gain agricultural productivity and increase quality of soil. Further product tested in terms of beneficiary and increasing productivity of crops and cooking gas usage for performing cooking.

Proposed project development of IoT-enabled waste management system helps to improve soil quality for agricultural activities is associated with science and technology. India is progressive to fight disease, poverty, and illiteracy. It calls for innovative technology to reduce pollution in world's biggest cities, detection of deadly viruses, and environment protection. Detection of waste product analysis requires steps transformation, processing, and generation of final product using IoT devices such as sensors and microcontroller Raspberry Pi 3.

Solid waste management is key problem like sanitation and public health. Considering a problem of waste management, waste neglected, and poisonous for land, air, and water. So as to overcome unhygienic environment and increasing quality of living beings, waste management project got attempted. We proposed development of IoT-enabled waste management system to improve soil quality for agricultural activities. Work talks about welfare for farmers and crop yield to be productive. We proposed ecofriendly solution for organic waste management and taken an step toward improvement of quality of soil. Finally, we can receive product in form of compost after decomposition process and cooking gas. IoT-enabled approach reduces

organic waste and offers benefits to our society to breath in healthy and less polluted area and further develops countries economic growth. It provides economical and feasible solution. The developed product is applicable to gated community, private offices, markets, schools' colleges. Deployment of this technique reduces transportation time to carry out waste. This reduces traffic on road and fuel cost along with human workers who operated the vehicle of waste collection. Before decomposing the good cleaned can be fed to cows, pets, and poultry farm, and after decomposing, gas can be used for cooking and manure for field. Compost is use for farming and horticultural sectors. This process reduces pollution and thus can be efficient to waste lots of space. Concept in agriculture applications using suggested technology for environmental protection found economical, efficient, user friendly, less costly for waste management system.

5 Implementation

For data analytics, we have used Microsoft Excel and Weka tool for retrieving results. Dataset is downloaded monthly-recycling-and-waste-collection-statistics-1 with 4 attributes date, month, type, total in tons, and 1140 instances. Further, count function is applied under various categories of data to understand waste contribution in tons is maximum. It has found on average curb garbage is maximum. Overall type of waste and total in tons table is shown in Table 1.

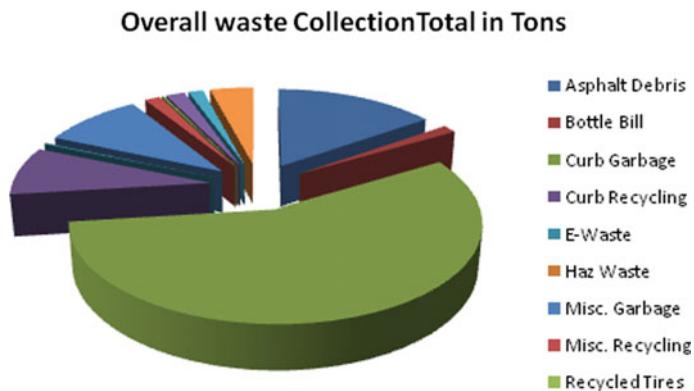
Overall category wise graph with count total in tons is shown in Fig. 3.

Year wise garbage type waste collection and grand total are shown in Table 2.

Garbage type versus total in tons graph is shown in Fig. 4.

Table 1 Overall type of waste and total in tons table

Type of waste	Total in tons
Asphalt debris	208,000
Bottle bill	18,703
Curb garbage	772,574
Curb recycling	113,186
E-waste	2781
Haz Waste	130
Misc. garbage	130,371
Misc. recycling	20,388
Recycled tires	2405
Scrap metal	24,423
Sidewalk Debris	18,908
Yard waste	56,014
Grand total	1,367,883

**Fig. 3** Category wise overall waste collection graph**Table 2** Year wise garbage type waste collation and grand total

Garbage type	Calendar year								Grand total
	2011	2012	2013	2014	2015	2016	2017	2018	
Asphalt debris	28,000	60,000	60,000	60,000	0	0	0	0	208,000
Bottle bill	0	0	0	3804	3804	3804	3804	3487	18,703
Curb garbage	102,404	95,045	113,856	113,109	88,936	87,325	89,231	82,668	772,574
Curb recycling	8864	13,177	14,106	14,669	14,944	15,472	16,501	15,453	113,186
E-waste	0	215	412	469	428	482	392	383	2781
Haz waste	0	6	7	37	24	19	17	20	130
Misc. garbage	23,736	20,948	2185	2381	27,042	19,843	16,418	17,818	130,371
Misc. recycling	1	14	63	4912	5158	4920	5210	110	20,388
Recycled tires	364	343	335	212	279	256	319	297	2405
Scrap metal	0	0	4057	4140	4142	4456	3976	3652	24,423
Sidewalk debris	3900	5000	5004	5004	0	0	0	0	18,908
Yard waste	2475	4480	5523	5868	11,433	10,800	13,300	2135	56,014
Grand total	169,744	199,228	205,548	214,605	156,190	147,377	149,168	126,023	1,367,883

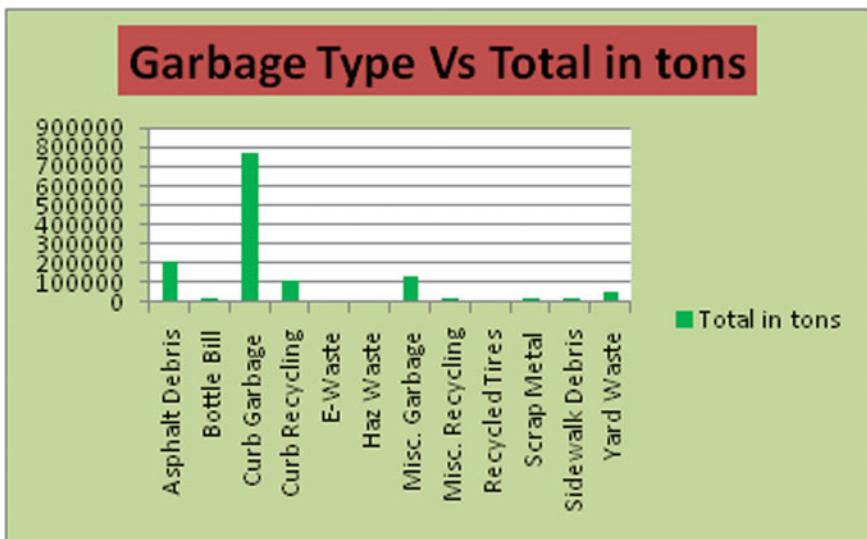


Fig. 4 Garbage type versus total in tons graph

We have used Weka classifier and obtain clusters using EM method for 13 instances and 10 attributes using EM scheme and obtained mean and std. deviation. Mean and std. dev. clustering result are shown in Table 3.

Using training set time taken to build model (full training data) is 0.01 s and clustered instances received with log likelihood: -114.53743.

Table 3 Mean and std. dev. clustering result

Year	Mean	Std. dev.
Yr2011	26,114.4615	49,521.9413
Yr2012	30,650.4615	56,001.4933
Yr2013	31,622.7692	59,542.6268
Yr2014	33,016.1538	61,152.0266
Yr2015	24,029.2308	44,671.8605
Yr2016	22,673.3846	42,490.8708
Yr2017	22,948.9231	43,094.8651
Yr2018	19,388.1538	37,637.6104
Total	210,443.5385	389,298.4271

6 Conclusion and Future Work

Proposed approach helps society by composting organic waste and transforming it into form of fertilizer. Considering cost issues with few automatic and high-end compost bins in future there is a need for IoT-enabled product for waste collection and management for superior environments and beneficiary to society as it is attempt toward smarter city and smarter city management. Approach proves to be economical, efficient, user friendly, less costly for waste management system. Solutions help in environment protection by promoting low cost of service for the waste management. In future, after testing developed product, new measures can be initiated as an incremental model for waste management. Deep study needs to perform for lifecycle of rechargeable batteries to incorporate in system. Batteries are fixed with containers and function of cell and interaction of the cells needs to study. Further development of new generation of batteries benefits many applications.

References

1. Bakhshi T, Ahmed M (2018) IoT-enabled smart city waste management using machine learning analytics. In: 2018 2nd International conference on energy conservation and efficiency (ICECE). Proceedings IEEE Xplore, 03 Dec 2018
2. Jain A, Bagherwal R (2017) Design and implementation of a smart solid waste monitoring and collection system based on internet of things. In: 8th International conference on computing, communication and networking technologies (ICCCNT), 3–5 July 2017. Proceedings IEEE Xplore, 14 Dec 2017
3. Jayaprakash S, Lohit HS, Abhilash BS (2018) Design and development of compost bin for indian kitchen. Int J Waste Resour 8(1). <https://doi.org/10.4172/2252-5211.1000323>
4. Yanga KSNA, Yakovlevab N (2019) Sustainable waste management through synergistic utilisation of commercial and domestic organic waste for efficient resource recovery and valorisation in the UK. J Cleaner Prod 227:248–262
5. Gupta H, Chaturkar A, Bagde A, Raj H, Hemane A, Deshmukh P (2019) IOT based garbage monitoring system. In: National conference on advances in engineering, technology and applied sciences (NCAETAS-2019). IJSRSET, vol 5(6)
6. Masek RFP (2016) Using genetic algorithm for advanced municipal waste collection in smart city. In: 10th International symposium on communication systems, networks and digital signal processing (CSNDSP)
7. Jaware G, Didwaniya A, Jatkar C (2018) Smart composting system using IoT. Int J Computer Eng Appl XII

Low-Cost Irrigation and Laser Fencing Surveillance System for the Paddy Fields



Abdul Subhani Shaik, Manir Ahmed, and Merugu Suresh

Abstract The sector that contributes most in the Indian GDP is the agricultural sector. However, the use of science and technology in this sector of India is very poor. Most irrigation systems used in India are manually operated which consume more water and required more manpower. Today, science and technology have increased enormously in a number of areas, including agriculture, healthcare, and so forth. In India, some traditional technologies are available in the field of farming like drip irrigation, sprinkler irrigation, etc. In this paper, sensors and Internet of things (IOT) techniques are combined with the traditional technology to make them intelligent. IOT helps access and make major decisions through different values of sensors such as soil humidity and water level sensors. The main focus of this paper is to reduce waste water and to minimize irrigation manual labor to save the farmers time, cash, and power. To monitor and control of unauthorized entry of animals in to the paddy fields, a laser fencing system is also been included in this paper.

Keywords Internet of things · Irrigation · Laser fencing · Soil moisture sensor · LDR

1 Introduction

Agriculture sector is the main source of the food and other manufacture goods. It is considered as a foundation of life. Agricultural sector growth is necessary for the economic growth of the nation. In India, more than 50% of work force employed in agricultural section and contributing 17–18% to country's GDP. However, most of the irrigation are controlled by manually which have several disadvantages such as wastes water resources and consumes manpower. [1–5]. Moreover, manual irrigation process, farmers spray more water than needed (over-irrigation), and as a result not

A. Subhani Shaik (✉) · M. Ahmed · M. Suresh

Department of ECE, CMR College of Engineering and Technology, Hyderabad, Telangana 501401, India

e-mail: sasubhani@cmrcet.ac.in

only productivity are challenged but also water and energy are wasted. In Medchal-Malkajgiri district of Telangana State, India, agriculture is the biggest consumer of water, amounting to up to 70% of the total use in the district. Therefore, proper irrigation and field protection method for cultivation of crops play an important role in such region.

Most of the farmers suffering by animals and birds which can cause a significant damage to the crops, they are growing in the paddy field. Animals like bandicoots and elephants can affect the large damage to the crops if they are not taken care earlier. Farmers are losing 10% of revenue due to such animals in India. Therefore, modern technology like webcam, buzzer, laser, body temperature sensors can be used to protect crops from animals and birds [6–11]. Some wireless technology can be utilized in the irrigation as described in the previous work [12–16].

This paper attempts to develop a low-cost automated irrigation system. The proposed irrigation system is continuously monitoring the moisture level of soil and the growth of the plant. The proposed system uses soil moisture sensor to continuously monitoring the moisture. The signals coming from the sensors are feed to the Arduino board. Based on the sensor information, the Arduino board takes the decision whether to switch ON/OFF the motor. The IOT techniques are used as a communicator between sensors and Arduino board. Arduino board performs several works to monitoring the paddy field such as recording age of crops, moisture of soil, display data on LCD, play different sounds. In this paper, an automated irrigation system is proposed which can monitoring the water level in the field with respect to the growth of the plant. The proposed method is also developed a laser fencing to avoid entry of unauthorized animals to the field.

2 Previous Works

A traditional level basin method is observing in India from the ancient days [15]. In this method, water is continuously supply from one end of the field until other side of the field becomes wets. Such method cannot wet the field evenly. Some portion of the field is overwatered often. Furrow irrigation method has several advantages than level basin method [15, 16]. Furrow irrigation technique basically used in vegetable production. Furrows are nothing but some sloping channel formed in the soil. In furrows irrigation, water is release in furrows. Here, vegetable plant is not directly contact with water which is basically need for vegetable trees growth. This method also saves the water as compare to level basin method. The disadvantage of such methods is that they required more water and wastage of water can be seen.

Two modern irrigation methods are introduced in Indian agriculture recently [15, 17, 18]. These methods are (i) drip irrigation and (ii) sprinkler irrigation. In drip irrigation [19], water is supply to the plant with a narrow tube slowly. It falls water in the root as much required for the plant. Thus, no wastage of water and overwatering can be seen in this irrigation method.

Sprinkler irrigation method [19–21] supplies the water through a pressurized pipe network which sprays the water into the air. It is similar like artificial rains. It makes artificial rain two or three times in a day to plant. This method is easy to installation, and wastage of water is also less. It saves manpower too. Some machine learning techniques [22–24] can be combined with these modern irrigation methods to make it more robust. Thus, in this paper, sprinkler irrigation method is combined with Arduino nano device to make the irrigation automatic.

In India, some of the farmer use bamboo fencing which is not proper for protecting animal like elephant, bandicoot, etc. Wire and plastic fence are long lasting and strong too. But it is not affordable for the poor farmer. Electric agricultural fences are risk for farmer as well as animals. In this paper, a laser fencing is proposed which is cost effective and easy to use.

3 Proposed Irrigation System

The block diagram of components used in the proposed irrigation system is shown in Fig. 1. The proposed system is a combination of both hardware and software component. The heart of the system is Arduino nano which try to control all internal and external operation. The hardware component consists of sensors such as soil moisture sensor, water level sensor, laser generator, laser detector, buzzer, and Wi-Fi module. The software component consists of an android-based application connected to the Arduino board and Internet of things (IOT). This paper proposed an automated irrigation process on agricultural areas through the control of the soil water level in relation

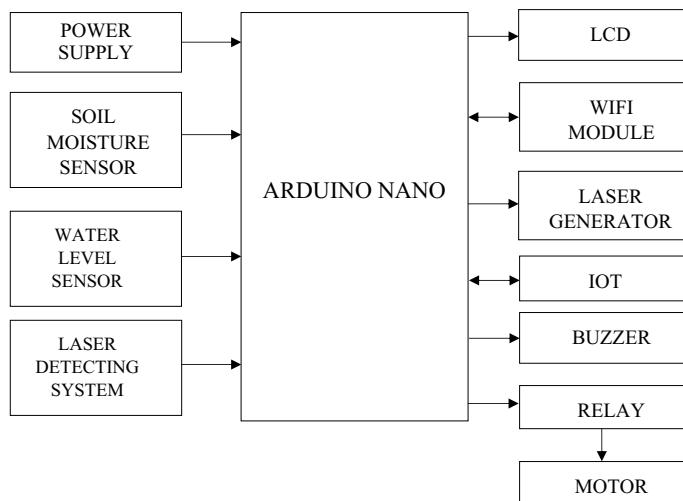


Fig. 1 Block diagram of the proposed system

to the crop age and the adaptive water sprinkling to simulate the impact of rainfall. The proposed method also integrates the laser detection and generation module to protect animal to enter the field. The working principle of all the components is described below.

Arduino Nano: As the name suggests, Arduino nano is a small microcontroller board with a dimension from 4.5 to 1.8 cm. Arduino Nano consists with different components such as microcontroller (ATmega328P), flash memory (32 KB), EEPROM (1 KB), and SRAM (2 KB). It has 16 MHz clock speed, 5–12 V input voltage, and 8 analog I/O pins.

Wi-Fi Module (ESP 8266): The Internet of things is a network of physical devices, cars, home appliances, and other elements integrated with electronics, software, sensors, actuators. Such connectivity allows them to communicate and exchange data between each other. The Internet of things is a network. The IOT enables the remote sensing or control of things through existing network infrastructures, which creates chances to directly integrate the physical world with the computer system. The ESP8266 supports APSD for VoIP applications and Bluetooth interfaces; it features an automatic RF which may be used under any operating situation and does not require external RF parts.

Soil Moisture Sensor: The water content is measured by soil humidity sensors in the ground. As a way of optimizing irrigation and warning plant stress on dry or wet, the soil moisture sensor is used. The data reading of the sensor is sent to the Arduino using ESP 8266 Wi-Fi module. The soil moisture is inversely proportional to the resistance of the soil. The sensors provide reading in terms of voltage. The voltage supply to the sensors is 5 V (recommended 3.3–5 V).

Water Level Sensor: These sensors are used to monitor the water level in the water tank or the water level in the paddy field. Once the water level falls below the threshold, then it sends message to Arduino to switch ON/OFF the motors.

Relay: A relay is a switch that runs electrically. Relay is extensively employed in sending signals from one circuit to another circuit. In the proposed irrigation system, relay is used in between Arduino and motors. Based on the information reached from the Arduino, relay do switch ON/OFF motor.

Liquid Crystal Display (LCD): LCD is used with Arduino to display reading. Different values come from soil moisture sensors, water level sensors, net growth of the crops, temperature, etc., are display on the LCD.

Buzzer (Speaker): Buzzer is used to produce burst like sound whenever some unauthorized animal enters to the field. This sound can alert both animals as well as farmers.

Laser Generation and Detection: This component can detect the animal present in the field. Once it detects animal, it sends message to the Arduino to make sound in the buzzer.

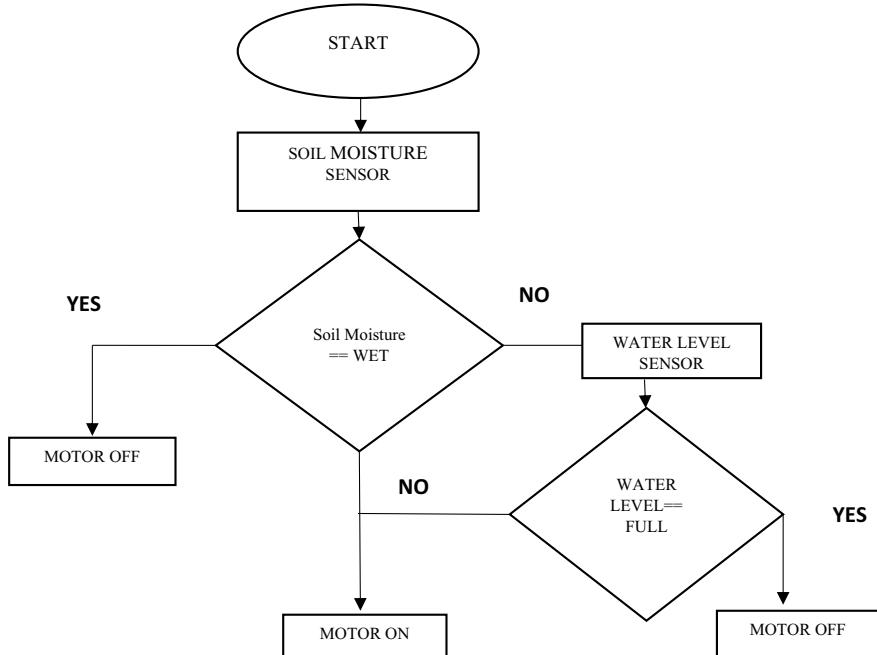


Fig. 2 Flowchart of the proposed water level monitor system

Working principle

The proposed system is aimed to eliminate and control the unwanted flow of water and to monitor the unauthorized entry of animal in to the fields. Initially, soil moisture and water level sensors are continuously monitored and transmitted the information to the Arduino. Android collects data from the associated IP address and calculates the dryness of soil and water level in the tank.

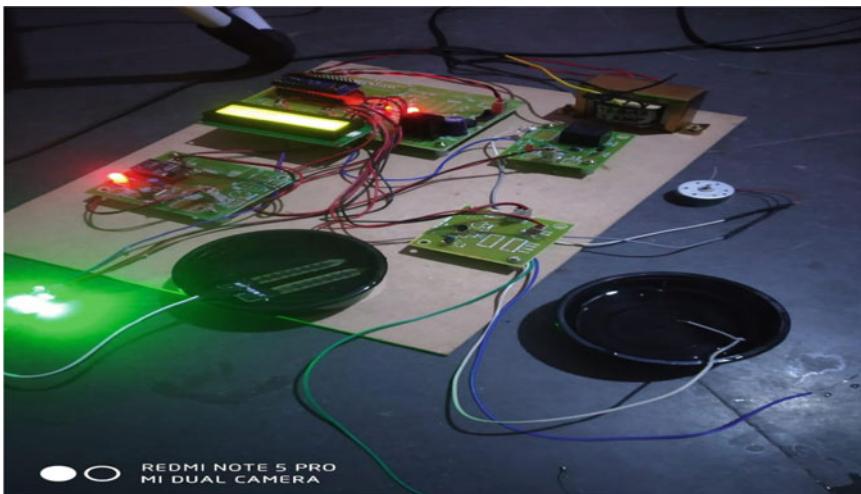
Based on the information, it makes switch ON/OFF the motors. Thus, the mechanism automatically switches the engine on whenever moisture levels in the ground reach below the threshold. The motor switches off automatically when the water level reaches its regular level. The flowchart of the working principle of the proposed system is shown in Fig. 2. Moreover, the proposed system used laser beam to detect animals. Once it detects animal, it sends information to Arduino. Arduino then switch ON the buzzer to keep away the animal.

4 Experimental Results and Discussion

The proposed system basically has two independent units. One is to detect the moisture level, temperature, humidity, and water level, and other one is laser fencing.

Combination of these units provides an automated agriculture irrigation and crops protect system. This system is design and test both in laboratory as well as physically environment in Medchal-Malkajgiri district of India. The laboratory and physical diagram are shown in Fig. 3.

To test the system in laboratory, a similar environment is created. The moisture level, water level, laser detection, temperature, and humidity are measures in such environment. It is observed that all the components are working properly. Some outputs of the experiment are shown in the Fig. 4.



(a)



(b)

Fig. 3 Sample of proposed system **a** laboratory circuit arrangement and **b** physical field arrangement

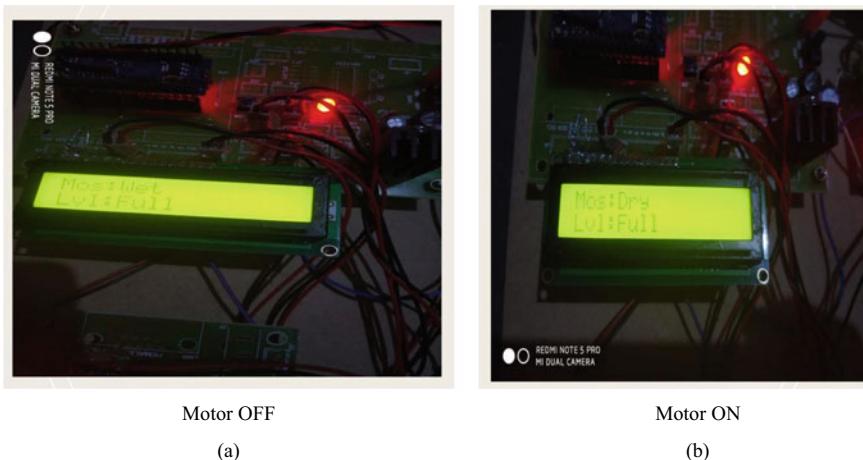


Fig. 4 Display reading on LCD in laboratory arrangement

The supply to the system and motor is 5 V and 230 V, respectively. The soil moisture sensor provides the analog reading in the range of 0.5–5 V. Different reading is taken by the soil moisture sensor in a sunny day is listed in the Table 1. The area of the physical field is 200 square meters. Sprinklers are set up at every 10-m distance in the field. It is observed that the proposed system spray water in the field twice in a day.

Table 1 Reading from Soil moisture sensor in a day

S. No.	Sensors reading (volts)	Time of recording	Dryness	Sprinkler
1	0.9	04.00 AM	Wet	Off
2	1.1	07.00 AM	Wet	Off
3	2.4	10.00 AM	Wet	Off
4	3.7	12.00 PM	Dry	On
5	0.1	01.00 PM	Wet	Off
6	0.6	03.00 PM	Wet	Off
7	1.2	05.00 PM	Wet	Off
8	2.5	07.00 PM	Wet	Off
9	3.6	10.00 PM	Dry	On
10	0.2	11.00 PM	Wet	Off
11	0.4	02.00 AM	Wet	Off
12	0.6	03.00 AM	Wet	Off

5 Conclusion

In order to optimize water supplies for agricultural productivity, the intelligent irrigation system is developed which is also cost effective. Depending on soil moisture levels, the system makes the water sprinkler ON and OFF. This paper developed an intelligent irrigation system to monitor and control the moisture level of soil automatically. This technology is, therefore, a remedy for the challenges in the current irrigation procedure facing by farmers. All observations and tests suggest that this paper represents a solution to the problems with irrigation and field observation. Practical observation also suggests that this implementation helps to increase harvesting and overall productivity.

Acknowledgments This research work has been carried out in the Research & Development Center, CMR College of Engineering and Technology, Hyderabad, India and is supported by the project titled Science, Technology and Innovation Hub (STI Hub), Department of Science and Technology, Government of India (File No. DST/SEED/TSP/STI/2019/94).

References

1. Klute A (1986) Methods of soil analysis. Part 1: physical and mineralogical methods. American Society of Agronomy, Madison, Wisconsin, p 1188
2. Knight JH (1992) Sensitivity of time domain reflectometry measurements to lateral variations in soil water content. *Water Resour Res* 28:2345–2352
3. Magagi RD, Kerr YH (1997) Retrieval of soil moisture and vegetation characteristics by use of ERS-1 wind scatterometer over arid and semi-arid areas. *J Hydrol* 188–189:361–384
4. Marthaler HP, Vogelsanger W, Richard F, Wierenga JP (1983) A pressure transducer for field tensiometers. *Soil Sci Soc Am J* 47:624–627
5. Bircher S, Skou N, Jensen KH, Walker JP, Rasmussen L (2011) A soil moisture and temperature network for SMOS validation in Western Denmark. *Hydrol Earth Syst Sci Discuss* 8:9961–10006
6. Putjaika N, Phusae S, Chen-Im A, Phunchongharn P, Akkarajitsakul K (2016) A control system in an intelligent farming by using arduino technology. In: 2016 Fifth ICT International Student Project Conference (ICT-ISPC), Nakhon Pathom, pp. 53–56
7. Abdullah A, Enazi SA, Damaj I (2016) AgriSys: a smart and ubiquitous controlled-environment agriculture system. In: 2016 3rd MEC international conference on big data and smart city (ICBDSC), Muscat, pp 1–6
8. Umair SM (2010) Automation of irrigation system using ANN based controller. *Int J Electr Computer Sci (IJECS-IJENS)* 10(02)
9. Kumar S (2013) Advance technique for soil moisture content based automatic motor pumping for agriculture land purpose. *Int J VLSI Embedded Syst (IJVES)* 04
10. Chikankar PB, Mehetre D, Das S (2015) An automatic irrigation system using ZigBee in wireless sensor network. In: 2015 International conference on pervasive computing (ICPC), Pune, pp 1–5
11. Kestikar CA, Bhavsar RM (2012) Automated wireless watering system (AWWS). *Int J Appl Inform Syst (IJAIS)* 2–3
12. Anurag D, Siuli R, Somprakash B (2008) Agro-sense: precision agriculture using sensor-based wireless mesh networks. In: Innovation in NGN, kaleidoscope conference, Geneva, pp 12–13

13. Arun C, Sudha KL (2012) Agricultural management using wireless sensor networks—a survey. In: 2nd International conference on environment science and biotechnology (IPCBEE), Singapore, p 48
14. Bogena HR, Huisman JA, OberdErster C (2007) Evaluation of a low cost soil water content sensor for wireless network applications. *J Hydrol*
15. Hussain R, Sehgal J, Gangwar A, Riyag M (2013) Control of irrigation automatically by using wireless sensor network. *Int J Soft Comput Eng* 3(1)
16. Aziz IA, Hasan MH, Ismail MJ, Mehat M, Haron NS (2009) Remote monitoring in agricultural greenhouse using wireless sensor and short message service (SMS). *Int J Eng Technol (IJET)* 9(9):1–12
17. Pavithra DS, Srinath MS (2014) GSM based automatic irrigation control system for efficient use of resources and crop planning by using an android mobile. *IOSR J Mech Civil Eng (IOSR-JMCE)* 11(I): 49–55
18. Kumbhar SR, Ghatule AP (2013) Microcontroller based controlled irrigation system for plantation. In: Proceedings of the international multi conference of engineers and computer scientists
19. Dursun M, Ozden S (2011) A wireless application of drip irrigation automation supported by soil moisture sensors. *Sci Res Essays* 6–7:1573–1582
20. Kim Y, Evans RG, Iversen WM (2008) Remote sensing and control of an irrigation system using a distributed wireless sensor network. *IEEE Trans Instrument Measure* 57(7)
21. Shaik AS, Karsh RK, Islam M (2022) A review of hashing based image authentication techniques. *Multimed Tools Appl* 81, 2489–2516. <https://doi.org/10.1007/s11042-021-11649-7>
22. Ahmed M, Laskar RH (2019) Eye center localization in a facial image based on geometric shapes of iris and eyelid under natural variability. *Image Vision Comput* 88:15. <https://doi.org/10.1016/j.imavis.2019.05.002>
23. Ahmed M, Laskar RH (2019) Eye detection and localization in a facial image based on partial geometric shape of iris and eyelid under practical scenarios. *J Electronic Imaging* 28(3):18, 033009. <https://doi.org/10.1117/1.JEI.28.3.033009>
24. Ahmed M, Laskar RH (2021) Evaluation of accurate iris center and eye corner localization method in a facial image for gaze estimation. In: *Multimedia System*. Springer, p 20. <https://doi.org/10.1007/s00530-020-00744-8>

Hybridization of Modified Cuckoo-Moth Flame Optimization for Effective Route Recovery of Networks



Satish Kumar Patnala, Venubabu Rachapudi, S Anjali Devi,
Sasibhushana Rao Pappu, and T. Subha Mastan Rao

Abstract In the mobile ad-hoc networks (MANETS), route failure is one of the reasons for route rediscovery, and the route failure may be caused by frequent mobility of nodes. The route failures in route discovery mechanisms lead to loss of data and overheads in communication. So, mobility is to be considered, while routing is performed in the MANETs. In this paper, a hybrid optimization based on the lifetime prediction algorithm, for route recovery in MANET, is being proposed. Hybrid optimization (MCMMF) is a combination of modified cuckoo and modified moth flame optimization algorithm. This algorithm considers the parameters, expiration time of a link (LET), signal strength of a link (LRSS), residual energy (RE/RS), and the availability of the bandwidth (ABW) to predict the lifetime of the node and the link. Using the predictions to decide a node status, the limits have been fuzzified, the fuzzy rules have been formulated, and all nodes in the network exchange this information. Thus, every node's status is verified before the transmission of data. The route recovery performance is found in a way that the respective routes are redirected to the stronger nodes or normal nodes in the case of a weak node. Simulation results using MATLAB2019b have shown that the modified cuckoo-modified moth flame (MCMMF) hybrid optimization algorithm produces the better results by comparing with the modified cuckoo search and modified moth flame optimization algorithms.

S. K. Patnala (✉)

Department of Information Technology, Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam, Andhra Pradesh, India
e-mail: psatishkumar.it@anits.edu.in

V. Rachapudi · S Anjali Devi

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

S. R. Pappu

Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh, India

T. S. Mastan Rao

Department of Computer Science and Engineering, CMR Technical Campus, Hyderabad, Telangana, India

Keywords MANETS · WSN · Fuzzy logic · Cuckoo optimization · Moth flame optimization

1 Introduction

With the outstanding mobile network technology development, wireless networks are becoming increasingly well known. Wireless networks can be divided into two major types—networks based on infrastructure (e.g. cellular networks) and ad-hoc mobile networks (MANETs). In the infrastructure-based networks, stationary base stations are responsible for coordination of communications among the mobile devices (nodes). The MANETs have many wireless nodes (devices) that can communicate with each other nodes (devices) without a network's base station or any other centralized administration. The wireless network's topology can vary unpredictably and rapidly because of the random movement of mobile nodes [1].

Reliability is one of the main issues in MANETs routing. Shortest paths may be considered to route the packets in MANETs. But this mechanism can fail quickly, because some wireless links on the shortest path may break, after the path has been established because of mobility of nodes [2]. Without the use of a base station or centralized administration, the nodes in MANETs can communicate with each other [3]. In an ad-hoc network, a packet sent by a node can be reached to the neighbouring node that is located in a distance up to the transmission radius [4]. The design of the routing algorithm should maximize the batteries' lifetime since node's energy is consumed while transmitting, receiving and relaying the packets between nodes [5].

2 Related Work

In general, to provide the packet routing, the network depends on the assistance of a node. Routing is one of the most basic operations in mobile ad-hoc networks. The routing algorithm must be robust, adaptive and reliable [6]. Prediction is considered in making of decision to transmit the data. Data packets cannot be forwarded from source to destination node when the pre-configuration threshold value is more than the prediction error [7]. A revolutionary-fuzzy prediction in the field of inter-domain routing of broadband network connections with QoS requirements is presented [8]. In order to better the performance of mobile ad-hoc networks, some of the challenging tasks like packet loss, routing overhead are to be considered [9].

The greater mobility of nodes leads to link breakage and that finally escalates the complexity of routing. Also, breakage of link increases the routing overhead and will reduce the network efficiency because of the escalated frequency of the route discovery. So, the main factor becomes a concept of link wreckage in MANET. These kind of link breakages are also the reason for periodic path failures and can cause

re-route constructions. Re-routing in MANETs is cost effective and can flood the network because of the dearth of infrastructure [10].

The route rediscovery also leads to the huge routing overhead and greater latency. Hence, efficiency of the network is reduced by the re-route discovery process. The network breakage due to frequent mobility of nodes or energy consumption restricts the routing in MANETs [11]. Priyadarshini has proposed a new algorithm to predict the lifetime of the node and link lifetime, which uses parameters related to draining rate of energy and relative mobility estimation rate. To forward the data packets, least dynamic route has been selected. This protocol outruns the current protocols similar to the lifetime prediction routing (LPR) and DSR protocols by considering the parameters throughout, failure in routing, overhead routing, loss of packet ratio and delivery of packet ratio [12]. Han has proposed a stable routing for MANETs by considering incalculable topology changes and recurrent link failures. By calculating the displacement between two adjacent nodes, the link accessibility is predicted over in some time. Based on the relative mobility of the nodes, an analytical expression for link availability has been derived. Route failure, channel's condition needs to be verified while designing to avoid packet loss and power loss. A new optimization technique, i.e. modified moth flame optimization algorithm technique, is imposed to identify the node and link lifetime based on the movement of the node and the energy consumption of the node [13].

Evolutionary algorithms and clustering mechanisms play an important role in finding flawless solutions for partition of a network. A number of revolutionary algorithms had been proposed in the past like the genetic algorithm, artificial neural network system, particle swarm intelligence and clustering based on particle swarm optimization that gives the top overall solutions[14].

3 Node Status Estimation Based on Fuzzy

This method uses some fuzzy logic techniques to detect the node's status. The interference based on the fuzzy rule is determined in Fig. 1, and it is as follows:

Fuzzification

In this process, the accurate input values are gathered from the chosen input values, and then the decrees to which the values belong to one of the likely fuzzy set are calculated. The input variable's expiration time of the link (LET), received signal strength of the link (LRSS), bandwidth availability (ABW) and residual energy (RE/RS) are fuzzified, and the outcome of fuzzification can be weak, normal and strong [15].

In Table 1, the 12 fuzzy sets are defined with the possible combinations. The inputs are LET, LRSS, ABW and RE, and the outcome shows the status of each node (NS). If all the metrics are strong, then the NS is strong. On the other hand, the NS will be weak if all the metric values are low. The other possible outcomes of these metric values are formed as fuzzy rules [16].

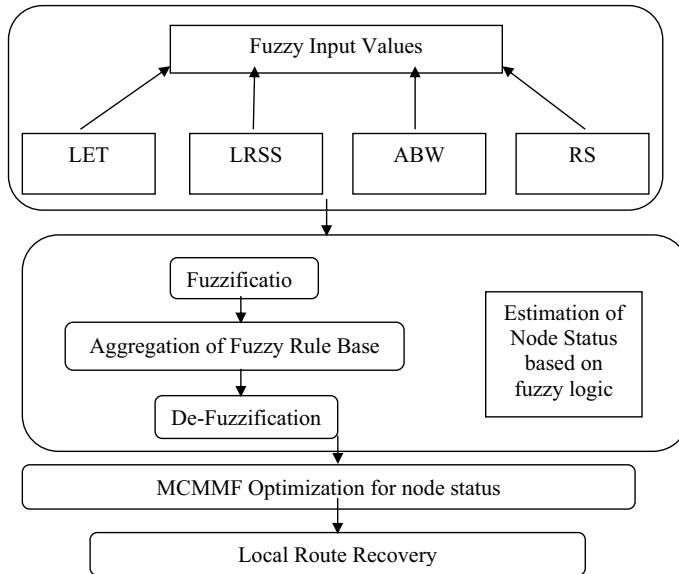


Fig. 1 Proposed process for estimation of node

Table 1 Fuzzy rules for determining output

S. No.	LET	LRSS	ABW	RE	Status of the node
1	Weak	Weak	Weak	Weak	Weak
2	Weak	Weak	Normal	Normal	Weak
3	Weak	Weak	Strong	Strong	Normal
4	Weak	Normal	Strong	Strong	Normal
5	Weak	Strong	Strong	Strong	Normal
6	Normal	Normal	Weak	Weak	Weak
7	Strong	Strong	Weak	Weak	Weak
8	Strong	Strong	Normal	Normal	Normal
9	Strong	Strong	Strong	Strong	Strong
10	Strong	Normal	Strong	Strong	Strong
11	Strong	Strong	Normal	Strong	Strong
12	Strong	Strong	Strong	Normal	Strong

De-fuzzification

De-fuzzification is the method by which the accurate input values are exfiltrated from a fuzzy set as a presentation value. We consider the centroid of area method for de-fuzzification while decision making process of the fuzzy. The below formula represents defuzzifier method.

$$\text{Fuzzy_ cost} = \frac{\sum_{i=1}^n x_i * \mu(x_i)}{\sum_{i=1}^n \mu(x_i)}$$

Here, x_i is the sample element, $\mu(x_i)$ indicates the function of membership for point x_i , and n stands for the count of elements in the sample [17].

4 Route Discovery and Route Recovery

The NS is finding to be either WEAK, NORMAL or STRONG after applying the fuzzy rules. Let us assume that the output of every node to be normal in the initial. Packets of data are transmitted from source to the destination node through nodes that are in the routing table. If a node wants to send a message, at first, a route request (RREQ) packet has been sent to the other nodes to maintain the routing table. Every node replies its status of node to the source node in the route reply (RREP) packet [18].

The status of a node can be checked in advance if a node transmits the packets to the following node. The packet is transmitted to the next node only if its status is either NORMAL or STRONG. If the successor node's status is WEAK, after which it sends route recovery warning (RRW) message to every neighbouring nodes. The neighbouring nodes after receiving the RRW packet are looking around to find the strong nodes. Local route recovery process can be initiated only when it finds a strong node by altering the route to the strong nodes. If they are unable to search for any strong nodes, then the normal nodes initiate the route recovery process [19]. Figure 2 depicts the above process.

Parameters Calculation

Here, LET, LRSS, ABW and RS are the parameters which need to be evaluated, and the computational is done using the following formulas.

The link expiry time (**LET**) is given as

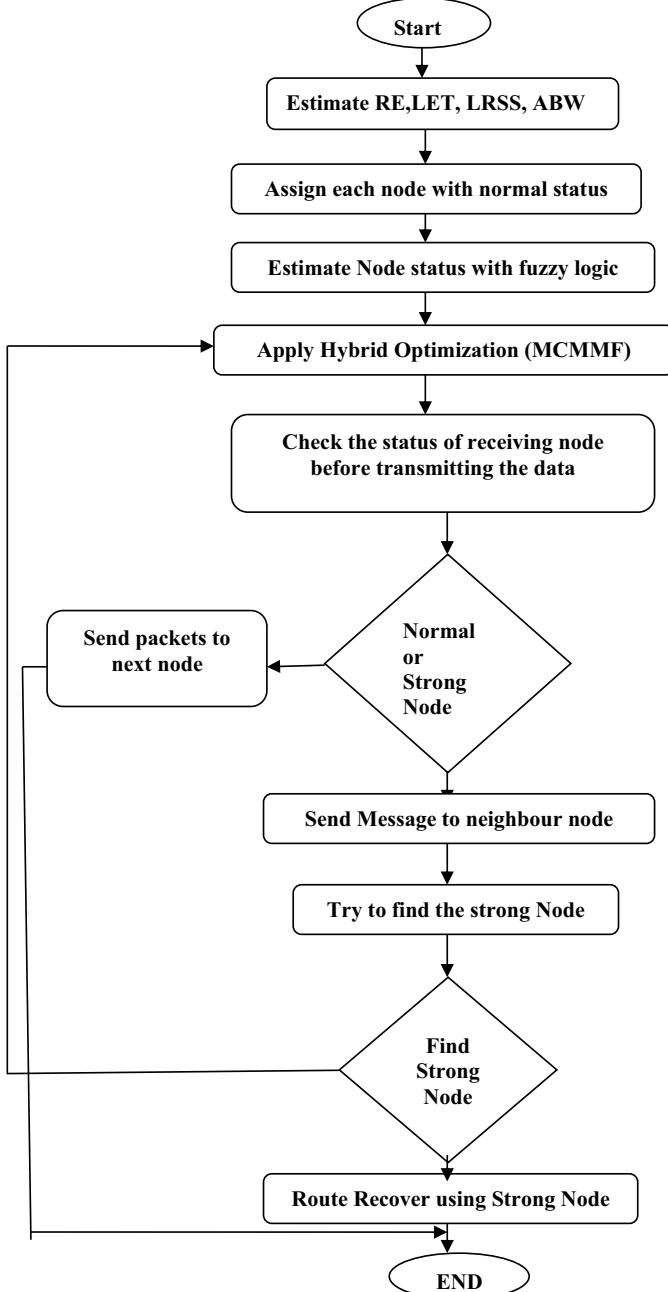
$$\text{LET} = \frac{-(pq + rs) + \sqrt{(p^2 + r^2)y^2 - (ps - qr)^2}}{(p^2 + r^2)} \quad (1)$$

where $p = s_1 \cos d_1 - s_2 \cos d_2$, $q = m_1 - m_2$, $r = s_1 \sin d_1 - s_2 \sin d_2$, $s = n_1 - n_2$.

Link received signal strength (**LRSS**) estimation is given as

$$\delta_r = \delta_t \times A_r \times A_t \times \frac{w^2}{(4 \times \pi \times v)^2} \quad (2)$$

where δ_r = received power, δ_t = transmitted power, A_t = antenna gain of the transmitter, A_r = antenna gain of the receiver, w = wavelength and v = distance between nodes.

Flowchart:**Fig. 2** Flow of proposed methodology

The available bandwidth (**ABW**) is calculated using

$$\beta_i = \max \{0, C_i - A_{R_{ij}}\} \quad (3)$$

The reception and transmission activities consume energy, they are under watch by every node in the network, and for every second, residual energy (**RE**) is calculated.

Let Δ_d = Energy dissipated for running the transmitter or receiver by node's circuit.

V^2 = Energy loss.

v = Distance between the nodes.

Δ_{loss} = Consumption of energy for transmitting the data packets.

The consumption of energy for transmitting the data packets of size h , a unit to a distance v , is given by

$$\Delta_t^{(h,v)} = \Delta_d \times h + \Delta_{\text{loss}} \times h \times v^2 \quad (4)$$

To get a packet which is of size h units, the energy consumed is

$$\Delta_R(h) = \Delta_d \times h \quad (5)$$

Hence, the total energy consumption to relay a packet is given by

$$\eta_{\text{tot}} = \Delta_t + \Delta_R + \Delta_o \quad (6)$$

where Δ_o is the energy consumed in overhearing activities.

After receiving and transmitting a packet of h units, the newly calculated η_{new} is

$$\eta_{\text{new}} = \eta_{\text{old}} - \eta_{\text{tot}} \quad (7)$$

where η_{old} is the residual energy calculated up to the previous interval [20]

5 Hybrid Optimization (MCMMF)

Modified Cuckoo (MCuSA) Search Algorithm

In this algorithm, every egg in a nest is designated as a solution, and every cuckoo lays one egg (thus representing one solution). While regenerating new solutions $x_i^{(t+1)}$ for the i th bird, the local random walk is performed as follows.

$$X_i^{t+1} = X_i^t + \alpha_s \oplus H(P_a - \epsilon) \oplus (X_j^t - X_k^t) \quad (8)$$

Here, two different solutions X_j^t , X_k^t are selected randomly by random permutation, $H(u)$ is known as Heaviside function, ϵ is a random number taken from

uniform distribution, and s is the step size. Here, \oplus means the entry-wise product of two vectors.

And, the global random walk is carried out using Lévy flights

$$X_i^{t+1} = X_i^t + \alpha L(s, \lambda) \quad (9)$$

where

$$L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi \lambda / 2)}{\pi} \frac{1}{s^{1+\lambda}}, \quad (s \gg s_0 > 0) \quad (10)$$

Here, $\alpha > 0$ is the step size scaling factor.

The step size α and discovery probability P_a determine the next generation nests. The convergence rate may not be guaranteed for the large step size. And for small value of step size (α), it has a high speed convergence rate, but it may fail to reach global optimum. So, the search accuracy is increased for small value of P_a , and large value of P_a increases the diversity of the solutions and inhibits premature convergence.

Modified moth flame optimization (MMoFOA)

The main update mechanism is usage of logarithmic spiral of moths in this algorithm. Any types of spiral can be utilized here with the following conditions:

- Moth can be spiral's initial point.
- Spiral's final point is the position of the flame.
- Fluctuation of the range of spiral should not exceed from the search space.

Here, we multiply the search by which the search is performed twice. By doing such, the moth can identify every neighbour and reach the flame. The moth's position gets updated with respect to the flame. The distance from each moth is calculated, and the position is updated.

$$S(M_i, F_j) = D_i \cdot e^{bt} \cos(2\pi t) + F_j \quad (11)$$

where $b = 2 * \text{rand}$.

Here b is the constant for defining the shape of the logarithmic spiral.

$$\begin{aligned} D_i &= |F_j - M_i| \\ t &= (a - 1) * \text{rand} + 1 \end{aligned}$$

where D_i is the distance between the i th moth and j th flame, M_i is the i th moth, and F_j is the j th flame.

The spiral flying path of moths is calculated using above equation. The later position of a moth is decided with respect to a flame. The parameter t present in the spiral equation defines the closeness between the following position of the moth and to the flame ($t = -1$ is the nearest position to the flame, whereas $t = 1$ shows the

farthest). Hence, in all directions, around the flame, a hyper ellipse can be assumed, and the following position of the moth would be within this space. The moths occupy their positions around flames by considering the spiral movement which is a main component in this mechanism. The spiral equation permits a moth to float “around” a flame and not necessarily in the search space between them. So, there is a guarantee of exploration and exploitation of the search space.

In hybridization of two optimization techniques, the distance is calculated based on the modified cuckoo search algorithm, and this distance can be used in modified moth flame optimization. By using this kind of hybridization, the results obtained in identifying the best solution will be implemented.

6 Results

The performance is evaluated by considering the simulation parameters in Table 2 with the two functions; one is packet size, and other is speed. Depending on the packet size and node’s speed, we have calculated the average end-to-end delay, packet delivery ratio, packet drop, and routing overhead. The same has shown in figures also.

Case 1

Here, we consider packet size and varies from 100 to 1000 bytes. The performance results based on the above metrics have been shown by varying the packet size. Figures show the graphical representation of the result.

Figure 3 specifies when the packet size increased from 100 to 1000 bytes, corresponding to the average end-to-end delay. Since MCMMF predicts the node lifetime more accurately than MCuSA and MMoFOA, the delay is less for it.

MANET routing protocols employ least hop routing principle where the route with minimum number of intermediate nodes is selected for a destination. This leads to the selection of longer wireless links which makes the whole route the shortest in hop number. However, packet delivery ratio increases when the packet size is increasing.

It is observed from Fig. 4 that the proposed algorithm MCMMF is performing better than the other two algorithms MCuSA and MMoFOA by offering high packet delivery ratio.

The mechanism used for preventing link failures is simply heuristic and can fail because a link failure is not expected. If the prediction is wrong, the connection’s throughput will remain unchanged, as the source will continue to use the current path until a new alternative path is measured or the current path fails.

As shown in Fig. 5 in our assessments of results, the proposed algorithm MCMMF performs better than the other two algorithms MCuSA and MMoFOA by offering low routing overhead.

When the ad-hoc route is extended to several hops, the general packet loss depends on the wireless connection in the route. Therefore, the rise in the packet size of the data transmitted over an ad-hoc wireless route has a linear effect on the packet loss

Table 2 Simulation parameters

Parameter	Value
Network coverage	100 m × 100 m
Packet size	100–1000 bytes
Speed of nodes	10–100 m/s
MAC	802.11
Traffic source	CBR
Routing protocol	MCMMF
Number of nests	25
Number of iterations	1000
Number of nodes	100
Discovery probability (P_a)	Random
Step size scaling factor (α)	0.01
Lower bound	[0 0 0 0]
Upper bound	[1 1 1 1]
Number of moths (M_i)	30
Number of flames (F_i)	30
Number of iterations	1000
Number of variables	30
Constant for defining the shape of a logarithmic spiral, b	$2 * rand$
Transmit power	0.395 W
Receiving power	0.660 W
Idle power	0.035 W
Initial energy	10.1 J

percentage. As shown in Fig. 6, the MCMMF algorithm is performing better and offering the low packet drop, when comparing with the other two algorithms.

Case 2

Now, we consider node's speed to study the effect of node disconnections on link quality, which varies from 10 to 100 m/s, and performance metrics are calculated. The below figures show the result of the above metrics for MCuSA, MMoFOA and MCMMF for varying the speed. Then, figures show the graphical representation of the results.

A random network deployed in an area of 100 m × 100 m is considered. At first, 100 nodes are deployed in a square grid area by accommodating each node in a 50 × 50 grid cell. The sink is assumed to be located 100 m away from the above specified region.

The increasing speed of the motion will cause dynamic in topology more frequently, thereby increasing the likelihood of disrupted connections. Broken links can lead to a further phase of route recovery and route discovery.

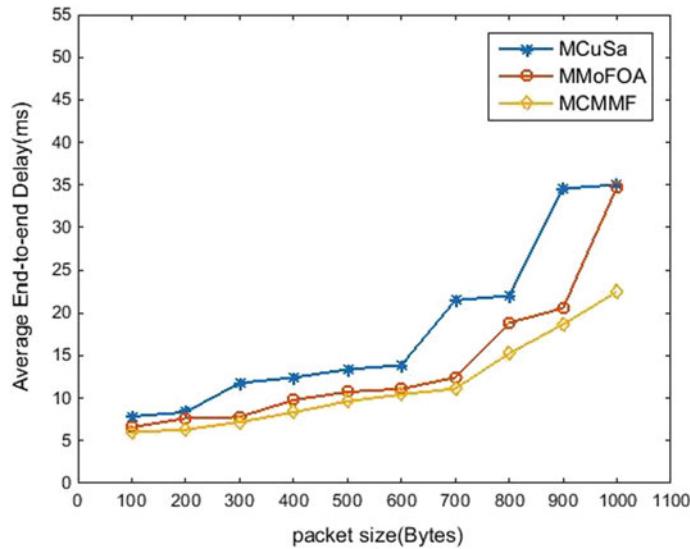


Fig. 3 Packet size vs. average end-to-end delay

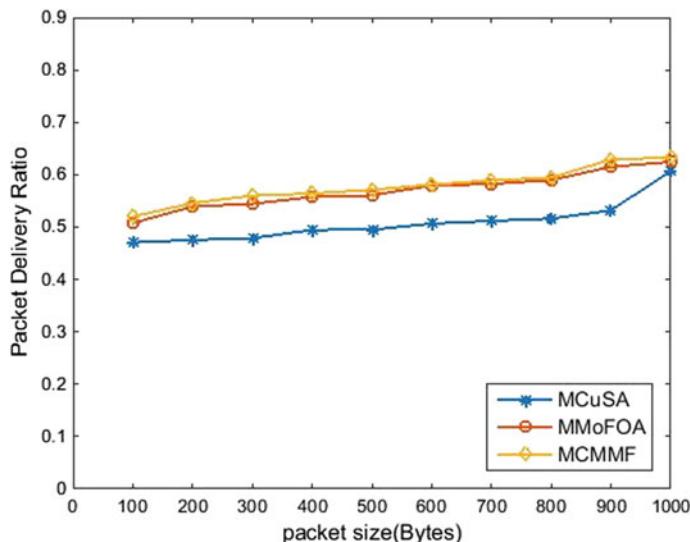


Fig. 4 Packet size versus packet delivery ratio

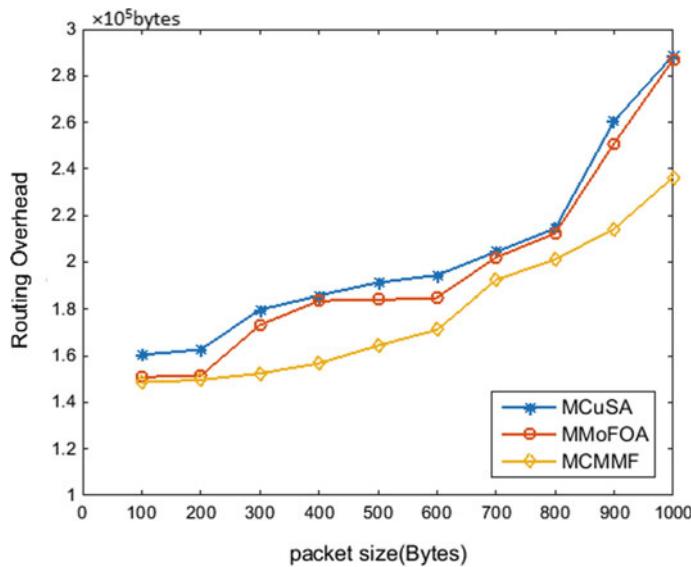


Fig. 5 Packet size versus routing overhead

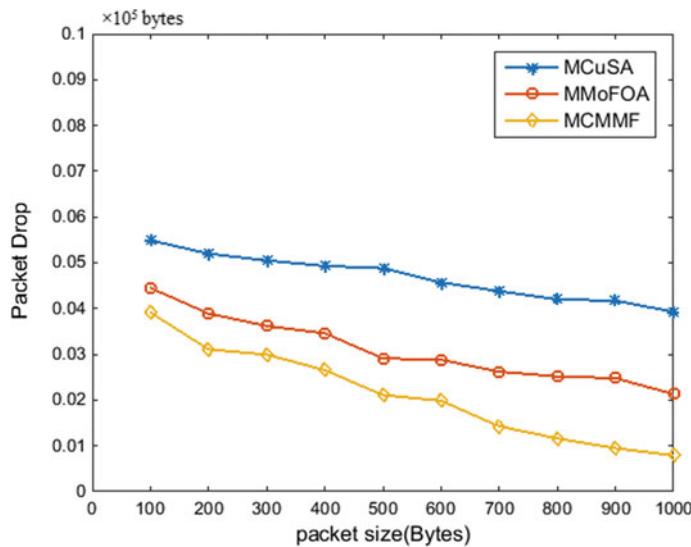


Fig. 6 Packet size versus packet drop

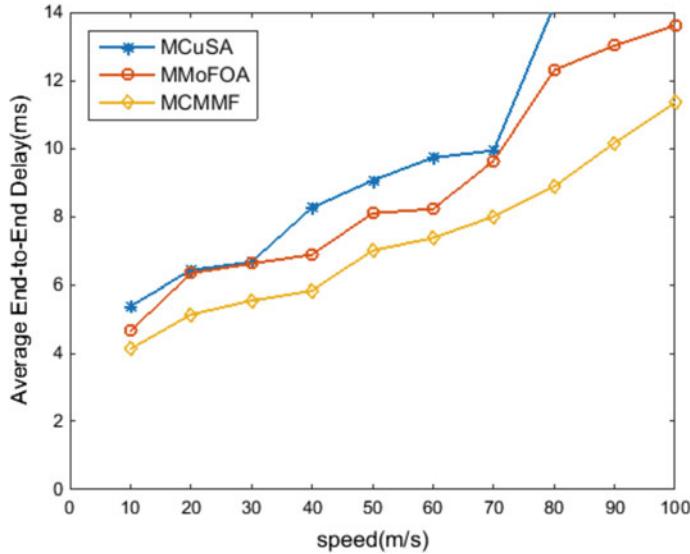


Fig. 7 Speed versus average end-to-end delay

As the node speed increases, it increases the average packet delay. It is observed from Fig. 7 that the proposed algorithm MCMMF is performing better than the other two algorithms by offering low end-to-end delay.

The packet delivery ratio decreases with increasing speed. This is because connection breaking will happen more frequently at higher speeds, thus increasing the packet loss fraction. In addition, as mobility speed increases, more updates are created with events which lead to a further reduction of the packet delivery fraction.

From Fig. 8, it is observed that the proposed algorithm MCMMF performs better than the other two algorithms by offering high packet delivery ratio.

Routing overhead increases at high speeds for all evaluated methods. In particular, the scenarios for high mobility due to the movement of the nodes lead to path breaks. The route error (RERR) message then floods the network to raise the overhead routing load. This is due to the frequent connection failures between the nodes. It is observed from Fig. 9 that the proposed algorithm MCMMF performs better than the other two algorithms by offering low routing overhead.

With a node speed increasing, the topology changes more often and induces a dynamic in routing direction, thus resulting in a higher packet loss rate because of the unavailability of the route to the forwarding nodes. It is observed from Fig. 10 that the MCMMF algorithm is performing better and offering the low packet drop when compared with the other two algorithms.

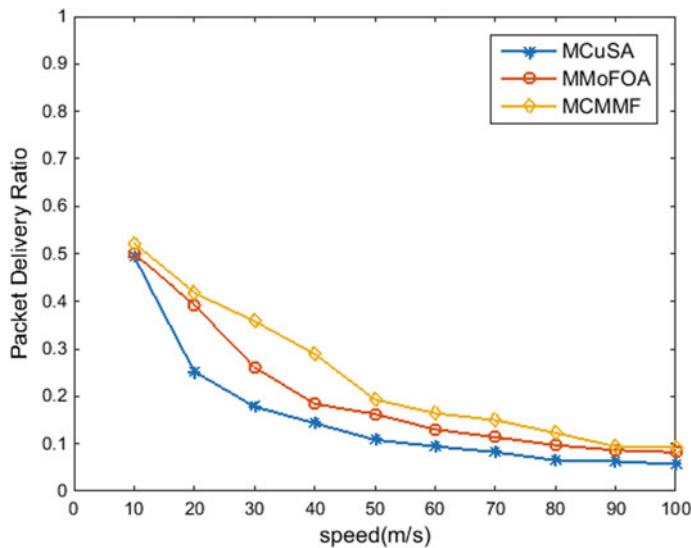


Fig. 8 Speed versus packet delivery ratio

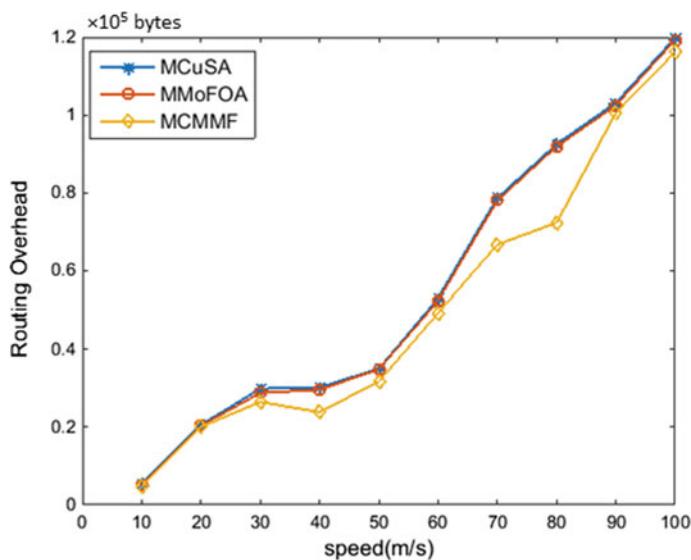


Fig. 9 Speed versus routing overhead

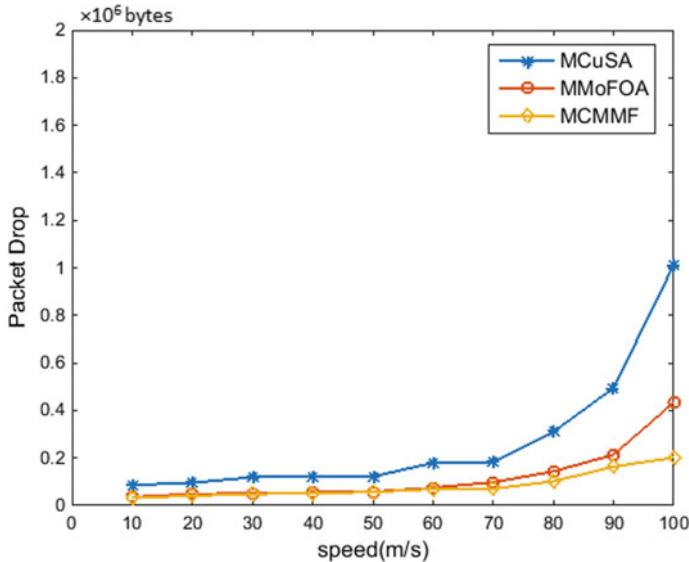


Fig. 10 Speed versus packet drop

7 Conclusion

In this paper, a hybrid optimization algorithm, i.e. modified cuckoo-modified moth flame (MCMMF) optimization which predicts the lifetime of a node to recovery the route in mobile ad-hoc network, has been proposed. Using the predictions to decide a node status, the parameters are fuzzified, and fuzzy rules have been formulated. The proposed method is compared with modified cuckoo optimization, modified moth flame optimization. The information is made to pass on among all the nodes. Thus, every node's status is verified before the transmission of data. The performance of a route recovery is formed in a way that the corresponding routes are redirected to the strong nodes or normal nodes in the case of a weak node. The simulation results show that proposed MCMMF algorithm achieves good outputs.

References

1. Kok P, Loh K (2006) A scalable, efficient and reliable routing protocol for wireless sensor networks. Springer, Germany (LNCS, 4159), pp 409–418
2. Ki-Il K, Sang-Ha K (2006) Effectiveness of reliable routing protocols in mobile ad hoc networks. *Wirel Pers Commun J* 38(3):377–390
3. Mobile Ad Hoc Network, en.wikipedia.org/wiki/Mobile_ad_hoc_network.
4. Marco Conti, Body, Personal (2003) Local ad hoc wireless networks. In: The handbook of ad hoc wireless networks. CRC Press LLC

5. Raghav RS (2021) Reinforced Cuckoo Search based fugitive landfill methane emission estimation. *Environ Technol Innov* 21, Article number 101207
6. Ahmed I, Tepe KE, Singh BK (2010) Reliable coverage area based link expiration time (LET) routing metric for mobile ad hoc networks, in *Ad Hoc Networks*. Springer, Berlin Heidelberg, pp 466–476
7. Zhang XM, En BW, Jing JX, Dan KS (2011) An estimated distance-based routing protocol for mobile ad hoc networks. *IEEE Trans Vehicular Tech* 60(7):3473–3484
8. Wei G, Ling Y, Guo B, Xiao B, Vasilakos AV (2011) Prediction-based data aggregation in wireless sensor networks: combining grey model and Kalman filter. *Comput Commun* 34:793–802
9. Kaleem SK (2018) Modeling and implementation of MPPT based wind energy system using Cuckoo search. *Int J Eng Technol (UAE)* 7:671–675
10. A Vasilakos, C Ricudis, K Anagnostakis, W Pedryca, W Pedryca, Evolutionary-fuzzy prediction for strategic QoS routing in broadband networks. In: The 1998 IEEE international conference on fuzzy systems proceedings, Anchorage, AK, 4–9 May 1998, vol 2, pp 1488–1493
11. Narendra BY, Suresh Babu CH (2013) Route recovery schemes for efficient MANET. *Int J Emerg Trends Eng Dev* 1:3
12. Banchhor C (2020) Data Knowl Eng 127, Article number 101788
13. Patnala SK, Nageswara Rao K (2020) An effective fuzzy based route recovery technique using moth flame optimization algorithm. *J Adv Res Dynam Control Syst* 12(02)
14. Reddy MSK (2020) Cuckoo search optimization based MPPT for Integrated DFIG-wind energy system. In: 2020 International conference on decision aid sciences and application (DASA 2020); Virtual, Sakheer, Bahrain; 7 Nov 2020 through 9 Nov 2020. Category number CFP20Y89-ART; Code 166557
15. Shuchita U, Charu G (2010) Node disjoint multipath routing considering link and node stability protocol: a characteristic evaluation. *Int J Com Sci Issues* 7(1):2
16. Thirugnanasambandam K (2019) Reinforced cuckoo search algorithm-based multimodal optimization. <https://doi.org/10.1007/s10489-018-1355-3>
17. Vamsidhar E (2018) A novel approach for feature selection and classifier optimization compressed medical retrieval using hybrid cuckoo search. ISSN: 20893272. <https://doi.org/10.11591/ijeei.v6i4.584>
18. Mahender K (2019) Asian Research Publishing Network (ARPN) 2019. ISSN: 18196608
19. Rajkamal K, Immadi G (2018) *Int J Eng Technol (UAE)* 7(2.20):367–383 (Special Issue 20)
20. Anguraj DK (2021) Enriched cluster head selection using augmented bifold cuckoo search algorithm for edge-based internet of medical things. *Int J Commun Syst* 34(9), Article number e4817

Analysis of Password Protected Documents Using Statistical Approaches on High Performance Computing



Ajeet Singh, Vikas Tiwari, Allu Swamy Naidu, Appala Naidu Tentu, K. Surjan Raju, and Ashutosh Saxena

Abstract Password-based validation frameworks are as yet the most regularly utilized components for ensuring the data regardless of being helpless against dictionary reference-based attacks. Password breaking is the way towards speculating or recuperating a secret key from put away areas or from an information transmission framework. Best state-of-the-art password analysing methods like HashCat, John the Ripper and rainbow crack empower clients to check billions of passwords each second against the secret key hashes. This paper discusses various techniques including traditional, probabilistic and statistical methods for cracking the password protected files. Further, experimental evaluation, rationale and performance analysis on some sample password protected files are presented in this paper. The findings in this paper will also help understanding of both password-composition policies and metrics for quantifying password security.

Keywords Privacy · Password cracking · Dictionary attack · Brute-force attack · Personally identifiable information · Learned patterns

1 Introduction

Despite significant progress in attackers' abilities to crack passwords, text-based passwords remain the most used validation approach in computer-based systems. There are a wide range of approaches to validate clients of a framework; for example, a client can introduce a physical article like a key card, demonstrate character utilizing an individual trademark like a fingerprint or use something that solitary the client

A. Singh (✉) · V. Tiwari · A. S. Naidu · A. N. Tentu · A. Saxena

C.R. Rao Advanced Institute of Mathematics Statistics and Computer Science, University of Hyderabad Campus, Prof. CR Rao Road, Hyderabad 500046, Telangana, India

e-mail: ajeetc@uohyd.ac.in

K. S. Raju · A. Saxena
CMR Technical Campus, Hyderabad 501401, India
e-mail: hod.cse@cmrte.ac.in

knows. Passwords [1, 2] are one of the methods designed to provide authentication [3–5]. As opposed to different methodologies recorded, an essential advantage of utilizing confirmation through a password [6] is that if your password becomes compromised it very well may be effortlessly changed. In this paper, we discuss about general view of password cracking, methodologies for password cracking [7] at the point when an attacker can sign in to the framework by giving a client name and password pair, methods [8, 9] when an attacker approaches how passwords are put away on the framework. Figure 1 represents some scenarios attempts with which password cracking can occur. Figure 2 gives an overview of password hash salting flow.

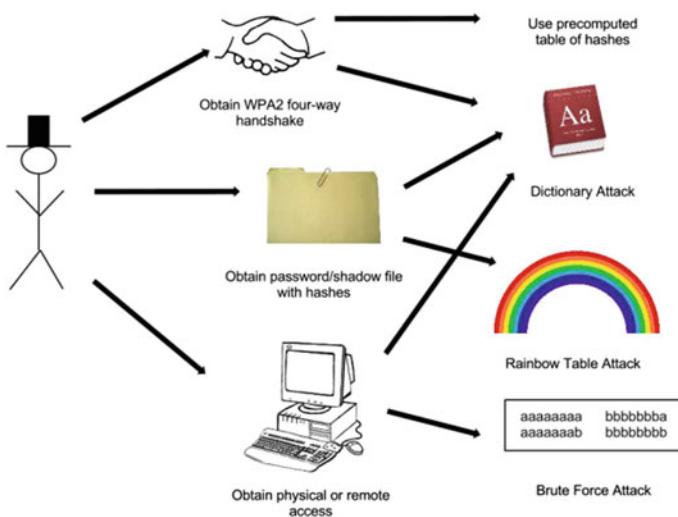


Fig. 1 Various techniques for password cracking

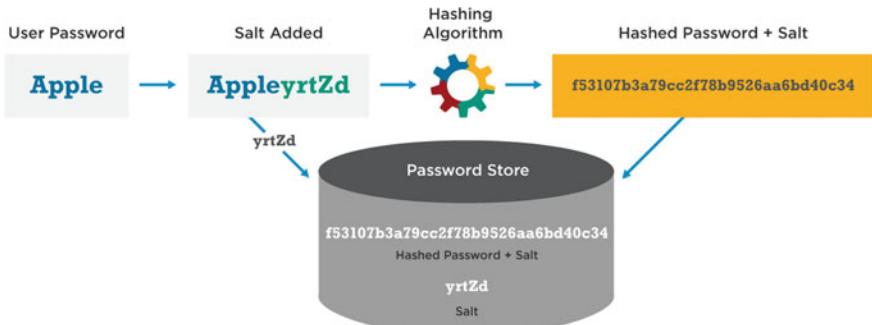


Fig. 2 Password hash salting flow

1.1 Contribution Highlights

Our contribution in this paper is summarized as follows:

- Various techniques including traditional, probabilistic and statistical methods for cracking the password protected files are discussed.
- Experimental evaluation, rationale, computational features and performance analysis on some sample password protected files are presented in this paper.
- Dictionary wordlist creation is performed based on statistical patterns observation by state-of-the-art tools such as *Flex*, *pydicator* and *crunch*.
- Integrated our generated dictionaries with Elcomsoft software tool [10]. Further, tested the functionality for password protected MS-Word input files up to 6-characters length.

1.2 Background and Related Work

This section covers various methods and frameworks developed for password guessing over past years. Kedem et al. [11] discussed the strategy of brute force attack on the Unix passwords in the perspective of SIMD architecture computer. Hitaj et al. [12] given a deep learning methodology for password guessing. Chou [13] given a framework and strategy for password breaking dependent on learned designs from unveiled passwords. Dell et al. [14] and Drmuth et al. [15] proposed a faster platform of password guessing exploiting an ordered Markov enumerator. Juels et al. [16] and Ma et al. [17] given the most occurring passwords and an investigation of probabilistic secret word models. Yampolskiy et al. [18] given an strategy for the dissecting user password selection behaviour for the reduction of certain password space. Weir et al. [19] discussed the attack procedure for password cracking exploiting probabilistic context-free grammars. Kelley et al. [20] discussed techniques for measuring passwords strength measure by resembling password cracking procedures. Bonneau et al. [21] discussed strong passwords choosing strategies, furthermore the development of flawed confirmation over the years.

Wang [22], in his thesis, discussed the key issues in password security. Ma et al. [23] performed the study of probabilistic password models. Narayanan et al. [24] surveyed fast dictionary-oriented attacks on passwords exploiting time-space trade-off technique. Weir et al. [25] presented framework for password cracking exploiting probabilistic context-free grammars. Veras et al. [26] discussed semantic examples of passwords and their corresponding security impact. Melicher et al. [27] proposed a fast and efficient procedure for modelling password guessability using neural networks. Aggarwal et al. [28] did the survey of different modern technologies in password cracking techniques. Tirado et al. [29] presented another dispersed brute force secret phrase breaking method. Hitaj et al. [30] presented a deep learning-oriented methodology for password guessing. Ji et al. [31] given a massive-scale empirical concentrate on the crackability, relationship and security of passwords. Li et al. [32]

given a huge scope observational investigation of Chinese Web passwords. Yampolskiy [33], in their work, analysed client secret phrase determination conduct for decrease of password space. Gong-Shen et al. [34] performed the password weakness evaluation and recuperation dependent on rules mined from enormous data.

Das et al. [35] identified a couple of basic strategies clients frequently utilize to change an essential password between destinations which can be utilized by an aggressor to make password speculating endlessly simpler. Li et al. [36] given an investigation of individual data in human-picked passwords and their security suggestions. Merhav et al. [37] performed attacks. Lu et al. [38] given an estimation investigation of validation rate-restricting components of present day sites. Pal et al. [39] presented a password similarity model using neural networks. Guri et al. [40] analysed the individual data spillage during secret word recuperation of Internet providers. Bailey et al. [41] given the insights on password re-use and versatile strength for monetary records. Emin Islam [42] presented an strategy to crack more password hashes with specific patterns. Stobert et al. [43] talk about the general password life cycle and client conduct in overseeing passwords. Kelley et al. [44] performed measuring password strength by mimicking secret key breaking calculations. Shay et al. [45] discussed the effect of direction and criticism on password creation conduct. Wang et al. [46] given a framework to analyse the passwords of specific Chinese Web end-users.

1.3 *Organization of the Paper*

Section 2 covers a brief summary on password storage. Various methods of password cracking and other findings are discussed in Sect. 3. Section 4 covers the theoretical password cracking model using Elcomsoft attack tool. Experimental evaluation, computational features, results discussion and performance analysis are presented in Sect. 5, and conclusions are given in Sect. 6.

2 Password Storage

This section provides a short discussion about how the passwords [47, 48] are stored in the system. Putting away client names and comparing passwords in plaintext is not any more an adequate arrangement. Endeavouring to shroud passwords put away as plaintext (e.g. putting the secret phrase record somewhere down in a tangled registry chain of importance) would add up to security through obscurity which would likewise be inadmissible. The Unix arrangement of record the executives, on the other hand, is superior: one of the agreements. Beginning adaptations of Multics (the forerunner to Unix) put away the secret phrase record in clear content, yet just perceptible with superuser consents. This arrangement likewise attacked when a

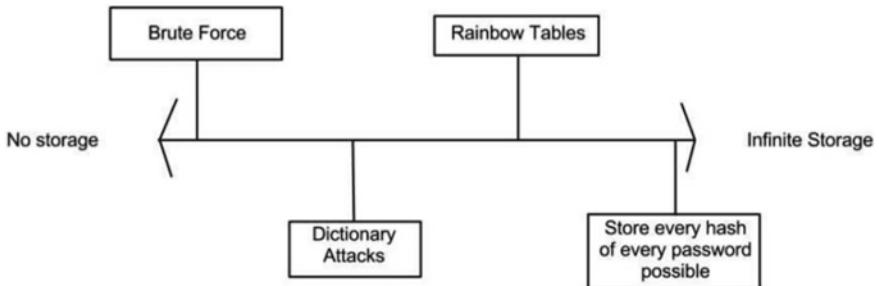


Fig. 3 Variation with respect to storage

bug exchanged some impermanent documents around and the secret key record (in plaintext) was printed for each client upon login [49].

Unix, rather, stores the hashed estimation of passwords in the secret key record rather than the genuine passwords [50, 51]. At that point when a client inputs their secret key, the framework can essentially take the hash of the info and contrast it with the put away hash esteem [52, 53]. Variety regarding capacity is appeared as Fig. 3.

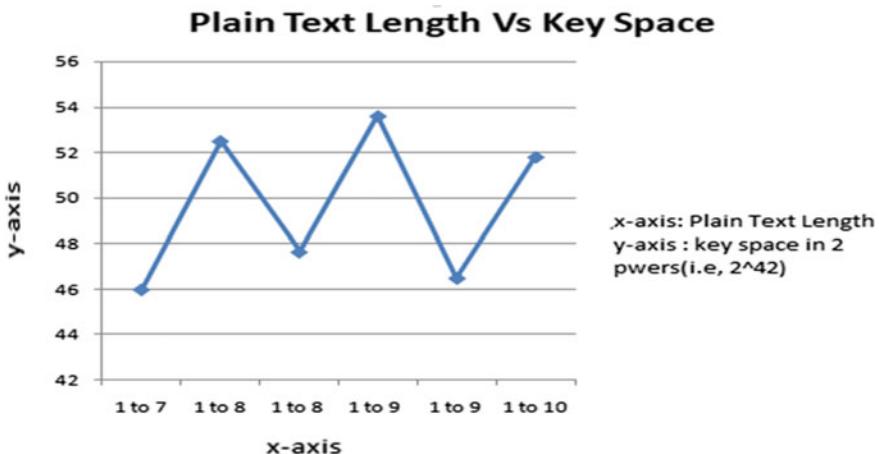
- In a large portion of the Unix-based record frameworks, the secret key document is situated at */and so on/password*. Each line in this record contains data around one record on the framework. The record itself is lucid by all clients, however, is just writable with superuser benefits.
- The secret key document for Windows, known as the security accounts manager (SAM) record, contains seven colon delimited fields: the client name, client number, scrambled secret phrase, hashed secret word, hashed secret phrase under an alternate calculation, complete name of client and lastly home registry. Rather than the Unix secret key record, the Windows SAM document isn't meaningful once the working framework has booted.
- Many sites and online administrations expect clients to sign in with a common secret word plot. This requires the capacity of secret word data. Be that as it may, online administrations normally store passwords for their framework in a non-normalized way, and these frameworks are not generally planned by engineers with foundations in protection or security.

3 Various Methodologies and Findings

This section discusses various methodologies for cracking the password protected files along with some interesting findings.

Table 1 Size of rainbow table and extent of key space covered

Char set	Plaintext length	Key space	Table size (GB)
ascii-32-95	1–7	2^{46}	52
ascii-32-95	1–8	$2^{52.5}$	460
Mix-alpha-numeric	1–8	$2^{47.65}$	127
Mix-alpha-numeric	1–9	$2^{53.6}$	690
Lower-alpha-numeric	1–9	$2^{46.5}$	65
Lower-alpha-numeric	1–10	$2^{51.8}$	316

**Fig. 4** Variation of plaintext length versus key space covered

3.1 Rainbow Tools for Password Recovery

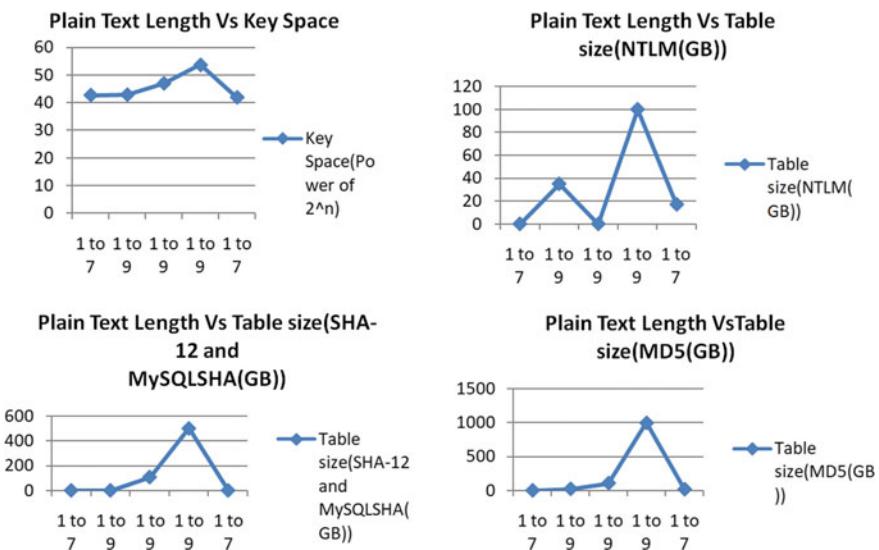
Here, the strategy is straightforward, i.e. by precomputing some part to the issue (regularly either by explaining subproblems or by discovering normal arrangements), the time cost of taking care of the issue all in all is diminished, while the space necessities are considerably not as much as what might be expected to completely precompute the solution.

3.1.1 RainbowCrack

RainbowCrack [54] is a computer program that was developed by Zhu Shuanglei. It cracks hashes exploiting rainbow tables and recovers the plaintext. Table 1 and graph (shown in Fig. 4) represent the size of rainbow table and extent of key space covered for different character types having different lengths.

Table 2 Sizes of rainbow tables for different Hash algorithms

Char set	Plaintext length	Key space	Table size		
			NTLM (GB)	SHA-1 and MySQLSHA1 (GB)	MD5 (GB)
All-space	1–7	$2^{42.7}$	–	–	–
Alpha-space	1–9	$2^{42.86}$	35	–	23
Lower-alpha-numeric-space	1–9	2^{47}	–	108	108
Mix-alpha-numeric	1–9	$2^{53.7}$	1000	504	1000
Mix-alpha-numeric-space	1–7	$2^{41.9}$	17	–	17

**Fig. 5** Comparison between the plaintext lengths and key space, NTLM, MD5, SHA-1 and MySQLSHA1

3.1.2 Rainbow Tables Generation

Rainbow tables generation is a free software application currently available in English, and it was last updated on 2010 and downloadable. Table 2 gives the sizes of rainbow tables for different hash algorithms.

The graph (shown in Fig. 5) gives comparison between the plain text lengths and key space, NTLM, MD5, SHA-1 and MySQLSHA1.

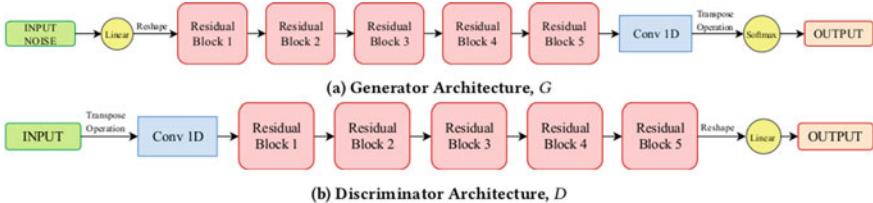


Fig. 6 PassGAN’s architecture

Rcracki_mt can be used to perform a rainbow table attack on password hashes. *GUIRainbowCrack* is extended version of RainbowCrack in GUI. GUI RainbowCrack is a new way to crack password using pre-computed password hash table (PCPH). It substantially improves the speed of standard password cracking.

3.2 PassGAN: Generative Adversarial Networks Architecture Based Password Guessing

PassGAN [12] replaces human-generated password rules with theory-grounded machine learning algorithms. PassGAN architecture view is shown in Fig. 6.

3.2.1 Architecture Overview

Figure 6 represents PassGAN’s architecture. There are two components named as generator and discriminator.

It utilizes—*matplotlib*—2.1.1, *numpy*—1.13.3, *Tensorflow*—1.4.1, *Tensorflow-gpu*—1.4.1.

4 Password Cracking Using Elcomsoft Attack Tool

The rationale and various attacks provided by Elcomsoft password recovery tool [10] are summarized as follows.

4.1 Rationale

- It supports all versions of Microsoft Office 2.0 to Microsoft Office 2019 and PDF form as password protected input file.

- Dictionary, such as, exploitation of rockyou.txt [55] and brute force attacks with user-defined masks and templates. User can add his own created dictionary of any size in the tool then execute the attack.
- Supports multiple language dictionaries.
- Supports GPU acceleration with state-of-the-art NVIDIA cards.
- Hardware acceleration reduces password recovery time by a factor of ≈ 50 .
- Supports 64 CPUs and up to 8 GPUs.

5 Experimental Evaluation and Performance Analysis

This section presents experimental set-up, simulation environment, obtained results in various test case scenarios and performance analysis in terms of CPU and GPU utilization.

5.1 Set-Up and Simulation Environment

Our high performance computing workstation set-up and simulation environment consist of following software and hardware specifications: Operating system as Windows 10, Intel Core i7-8750H processor, RAM size 32 Gb, NVIDIA Quadro P600 graphics card having 384 NVIDIA CUDA Cores and Python 3.7 installed.

5.2 Obtained Results

The input in the experiments is in the form of password protected word documents. The obtained results in various test case scenarios are given in Table 3. The time to recover password is judged based on attack scenarios, i.e. password character length, chars combination, search space and attack type. Various experimental test cases of 4, 5, 6 characters are taken into consideration.

Performance analysis is done based on CPU and GPU utilization (representation is given as Figs. 7 and 8).

6 Conclusive Discussion

This paper discusses various techniques including traditional, probabilistic and statistical methods for cracking the password protected files. Experimental evaluation, rationale, computational features and performance analysis on some sample password protected files are also presented in this paper.

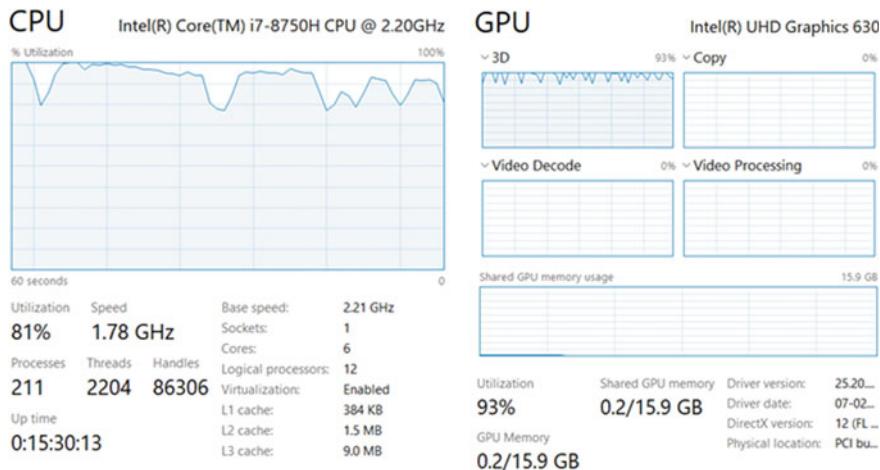


Fig. 7 CPU and GPU utilization

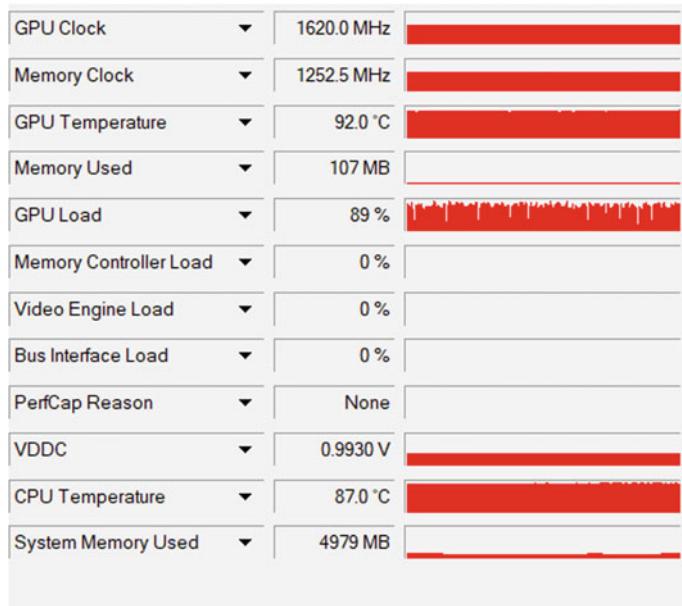


Fig. 8 Sensors view of NVIDIA Quadro P600 graphics card

Table 3 Tabular representation of the obtained results in various test scenarios

Char len	Combination	Search space	Attack type	Time to recover password
4	$< a...z >< 0...9 >$	36^4	Brute force attack	≈ 65 s
5	$< a...z >< 0...9 >$	36^5	Dictionary attack	≈ 110 s
6	$< a...z >< A...Z >< 0...9 >$	62^6	Dictionary attack	≈ 154 s
6	$< a...z >< A...Z >< 0...9 >$	62^6	Dictionary attack	≈ 1500 s
6	$< a...z >< A...Z >< 0...9 >$	62^6	Dictionary attack	≈ 1140 s
6	$< a...z >< A...Z >< 0...9 >$	62^6	Dictionary attack	≈ 240 s
6	$< a...z >< A...Z >< 0...9 >$	62^6	Dictionary attack	≈ 618 s
6	$< a...z >< A...Z >< 0...9 >$	62^6	Dictionary attack	≈ 1862 s

References

- Narayanan M, Shmatikov V (2005) Fast dictionary attacks on passwords using timespace trade-off. In: Proceedings of the 12th ACM conference on computer and communications security, pp 364–372
- John the Ripper password cracker. Openwall Project. <http://www.openwall.com/john>
- Ur B, Noma F, Bees J, Segreti SM, Shay R, Bauer L, Christin N, Cranor LF (2015) I added ‘!’ at the end to make it secure: Observing password creation in the lab. In: Eleventh symposium on usable privacy and security (SOUPS 2015), pp 123–140
- Karapanos N, Marforio C, Soriente C, Capkun S (2015) Sound-proof: usable two-factor authentication based on ambient sound. In: 24th USENIX security symposium (USENIX security 15), pp 483–498
- Wang D, Wang P (2016) Two birds with one stone: two-factor authentication with security beyond conventional bound. IEEE Trans Depend Secure Comput 15(4):708–722
- Wang D, Cheng H, Wang P, Huang X, Jian G (2017) Zipf’s law in passwords. IEEE Trans Inf Forens Secur 12(11):2776–2791
- Morris R, Thompson K (1979) Password security: a case history. Commun ACM 22(11):594–597
- Hellman ME (1980) A cryptanalytic time-memory trade-off. IEEE Trans Inf Theory 26(4):401–406
- Troy Hunt. <http://www.troyhunt.com/2011/06/brief-sony-password-analysis.html>
- <https://www.elcomsoft.com/aopr.html>
- Kedem G, Ishihara Y (1999) Brute force attack on UNIX passwords with SIMD computer. In: Proceedings of the 8th conference on USENIX security symposium, vol 8, Berkeley, CA, USA, 1999. USENIX Association, p 8
- Hitaj B, Gasti P, Ateniese G, Perez-Cruz F (2019) PassGAN: a deep learning approach for password guessing. In: International conference on applied cryptography and network security. Springer, Cham
- Chou HC, Lee HC, Yu HJ, Lai FP, Huang KH, Hsueh CW (2013) Password cracking based on learned patterns from disclosed passwords. Int J Innov Comput Inf Control 9(2)

14. DellAmico M, Michiardi P, Roudier Y (2010) Password strength: an empirical analysis. In: Proceedings of the IEEE INFOCOM. IEEE, pp 1–9
15. Drmuth M, Angelstorf F, Castelluccia C, Perito D, Abdelberi C (2015) OMEN: faster password guessing using an ordered Markov enumerator. In: ESSoS. Springer, pp 119–132
16. Juels A, Rivest RL (2013) Honeywords: making password-cracking detectable. In: Proceedings of the 2013 ACM SIGSAC conference on computer & communications security. ACM, pp 145–160
17. Ma J, Yang W, Luo M, Li N (2014) A study of probabilistic password models. In: IEEE symposium on security and privacy (SP). IEEE, pp 689–704
18. Yampolskiy RV (2006) Analyzing user password selection behavior for reduction of password space. In: Proceedings of the IEEE international Carnahan conferences on security technology, pp 109–115
19. Weir M, Aggarwal S, de Medeiros B, Glodek B (2009) Password cracking using probabilistic context-free grammars. In: Proceedings of the 30th IEEE symposium on security and privacy, pp 391–405
20. Kelley PG, Komanduri S, Mazurek ML, Shay R, Vidas T, Bauer L, Christin N, Cranor LF, Lopez J (2012) Guess again (and again and again): measuring password strength by Simulating Password-Cracking Algorithms. In: 2012 IEEE symposium on security and privacy, San Francisco, CA, pp 523–537
21. Bonneau J, Herley C, Van Oorschot PC, Stajano F (2015) Passwords and the evolution of imperfect authentication. ACM 58(7):78–87
22. Wang D (2017) Research on key issues in password security. PhD Dissertation, Peking University. <http://wangdingg.weebly.com/uploads/2/0/3/6/20366987/phd-thesis0103.pdf>
23. Ma J, Yang W, Luo M, Li N (2014) A study of probabilistic password models. In: 2014 IEEE symposium on security and privacy. IEEE, pp 689–704
24. Narayanan A, Shmatikov V (2005) Fast dictionary attacks on passwords using time-space trade-off. In: Proceedings of the 12th ACM conference on computer and communications security, pp 364–372
25. Weir M, Aggarwal S, De Medeiros B, Glodek B (2009) Password cracking using probabilistic context-free grammars. In: 2009 30th IEEE symposium on security and privacy. IEEE, pp 391–405
26. Veras R, Collins C, Thorpe J (2014) On semantic patterns of passwords and their security impact. In: NDSS
27. Melicher W, Ur B, Segreti SM, Komanduri S, Bauer L, Christin N, Cranor LF (2016) Fast, lean, and accurate: modeling password guessability using neural networks. In: 25th USENIX security symposium, pp 175–191
28. Aggarwal S, Houshmand S, Weir M (2018) New technologies in password cracking techniques. In: Cyber security: power and technology, pp 179–198
29. Tirado E, Turpin B, Beltz C, Roshon P, Judge R, Gagneja K (2018) A new distributed brute-force password cracking technique. In: International conference on future network systems and security. Springer, pp 117–127
30. Hitaj B, Gasti P, Ateniese G, Perez-Cruz F (2019) Passgan: a deep learning approach for password guessing. In: International conference on applied cryptography and network security. Springer, pp 217–237
31. Ji S, Yang S, Hu X, Han W, Li Z, Beyah R (2015) Zero-sum password cracking game: a large-scale empirical study on the crackability, correlation, and security of passwords. IEEE Trans Depend Secure Comput 14(5):550–564
32. Li Z, Han W, Xu W (2014) A large-scale empirical analysis of Chinese web passwords. In: 23rd USENIX security symposium, pp 559–574
33. Yampolskiy RV (2006) Analyzing user password selection behavior for reduction of password space. In: Proceedings 40th annual 2006 international Carnahan conference on security technology. IEEE, pp 109–115
34. Gong-Shen MKL, Wei-Dong Q, Jian-Hua L (2016) Password vulnerability assessment and recovery based on rules mined from large-scale real data. Chin J Comput 39(3):454–467

35. Das A, Bonneau J, Caesar M, Borisov N, Wang X (2014) The tangled web of password reuse. In: NDSS symposium 2014, p 7
36. Li Y, Wang H, Sun K (2016) A study of personal information in human-chosen passwords and its security implications. In: IEEE INFOCOM 2016—the 35th annual IEEE international conference on computer communications. IEEE, pp 1–9
37. Merhav N, Cohen A (2020) Universal randomized guessing with application to asynchronous decentralized brute force attacks. *IEEE Trans Inf Theory* 66(1):114–129
38. Lu B, Zhang X, Ling Z, Zhang Y, Lin Z (2018) A measurement study of authentication rate-limiting mechanisms of modern websites. In: Proceedings of the 34th annual computer security applications conference, pp 89–100
39. Pal B, Daniel T, Chatterjee R, Ristenpart T (2019) Beyond credential stuffing: password similarity models using neural networks. In: 2019 IEEE symposium on security and privacy (SP). IEEE, pp 417–434
40. Guri M, Shemer E, Shirtz D, Elovici Y (2016) Personal information leakage during password recovery of internet services. In: 2016 European intelligence and security informatics conference (EISIC). IEEE, pp 136–139
41. Bailey DV, Dürmuth M, Paar C (2014) Statistics on password re-use and adaptive strength for financial accounts. In: International conference on security and cryptography for networks. Springer, pp 218–235
42. Emin Islam T (2015) Cracking more password hashes with patterns. *IEEE Trans Inf Forens Secur* 10(8):1656–1665
43. Stobert E, Biddle R (2014) The password life cycle: user behaviour in managing passwords. In: 10th symposium on usable privacy and security (SOUPS 2014), pp 243–255
44. Kelley PG, Komanduri S, Mazurek ML, Shay R, Vidas T, Bauer L, Christin N, Cranor LF, Lopez J (2012) Guess again (and again and again): measuring password strength by simulating password-cracking algorithms. In: 2012 IEEE symposium on security and privacy. IEEE, pp 523–537
45. Shay R, Bauer L, Christin N, Cranor LF, Forget A, Komanduri S, Mazurek ML, Melicher W, Segreli SM, Ur B (2015) A spoonful of sugar? The impact of guidance and feedback on password-creation behavior. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp 2903–2912
46. Wang D, Wang P, He D, Tian Y (2016) Birthday, name and bifacial-security: understanding passwords of Chinese web users. In: 28th USENIX security symposium (USENIX security 19), pp 1537–1555
47. Wang D, Zhang Z, Wang P, Yan J, Huang X (2016) Targeted online password guessing: an underestimated threat. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp 1242–1254
48. Wang KC, Reiter MK (2019) How to end password reuse on the web. In: Proceedings of ACM CCS
49. Grassi PA, Fenton JL, Newton EM, Perlner RA, Regenscheid AR, Burr WE, Richer JP, Lefkovitz NB, Danker JM, Choong Y-Y et al (2017) NIST special publication 800-63b: Digital identity guidelines. Enrollment and identity proofing requirements. <https://pages.nist.gov/800-63-3/sp800-63b.html>
50. Jaggard AD, Syverson P (2018) Oft target. In: Proceedings of the PET
51. Adams A, Sasse MA (1999) Users are not the enemy. *Commun ACM* 42(12):40–46
52. Bonneau J (2012) The science of guessing: analyzing an anonymized corpus of 70 million passwords. In: 2012 IEEE symposium on security and privacy. IEEE, pp 538–552
53. Mazurek ML, Komanduri S, Vidas T, Bauer L, Christin N, Cranor LF, Kelley PG, Shay R, Ur B (2013) Measuring password guessability for an entire university. In: Proceedings of the 2013 ACM SIGSAC conference on computer & communications security, pp 173–186
54. <https://web.archive.org/web/20080705140750/http://www.antsight.com/zsl>
55. <https://www.kaggle.com/wjburns/common-password-list-rockyou.txt>

Author Index

A

- Abdul Subhani Shaik, 507
Abinash Sahoo, 315
Aditya Kumar, 399
Ajay James, 89
Ajay Kumar, 377
Ajay Kumar Dharmireddy, 113
Ajeet Singh, 533
Allu Swamy Naidu, 533
Aman Panwar, 357
Amita Rani, 365, 377
Amar Kumar Kathwas, 211
Amit Malik, 365
Amit Prakash Singh, 1
Amiya Kumar Rath, 249
Ananthi, K., 79
Anjaiah, T., 11
Anjali Devi, S., 517
Anju Kulkarni, 297
Anshul Kumar Mishra, 143
Anuradha, T., 157
Anushka Dikshit, 203
Anushka Dixit, 67
Anushree, 67
Appala Naidu Tentu, 533
Appala Srinuvasu Muttipati, 51
Arav Pandey, 391
Arjun Raghav, 357
Aseem Chandel, 143
Ashruti Rai, 203
Ashutosh Saxena, 533
Athira, B., 337

B

- Baidyanath Kumar, 315
Balachandra Pattanaik, 59
Bandlapalli Sudharani, 169
Bhagyasree, P. V., 89
Bharath Kumar, K., 487
Bhawna Singh, 457
Bisma, N. D., 89
Bivasa Ranjan Parida, 249

C

- Challa Ram, G., 419
Cheboyina Sindhu NagaDurga, 157
Chilakala Sudhamani, 487

D

- Deepti Kakkar, 41
Deivasigamani, D., 121
Dhanvanth Kumar, B., 273
Dilshad, S. K., 475
Dinesh Prasad, 19

G

- Gayathri, K. V., 113
Geetha, G., 273
Girish Kumar, D., 419
Govinda Kumar, E., 121
Gunapati Sreenivasulu, 169

H

- Hari Prasad, G. V., 487

I

Ijas Ahamed, A., 79

J

Jagadeeswara Rao, E., 11

Janani, I., 305

Jasdeep Kaur, 67

Jayalakshmi Nema, 51

Jayant Mani Tripathi, 143

John Blessy Mandru, 447

Jyothi Gutala, 465

K

Kalidindi Lakshmi Divya, 429

Kalyani, G., 465

Karunakar Reddy Santhamgari, 131

Kavitha Rani, B., 497

Krishna Chaitanya, V. V., 273

Kshitij Karnawat, 297

P

Partha Sarathi Pattnayak, 325

Parveen, 407

Pavan Kumar, Y., 191

Pooja, 103

Prabhu Surya, R., 79

Priyadharsini, S., 79

Priya Geda, 285

R

Radha Dharavathu, 51

Raj Chaurasiya, 391

Rakesh Saur, 211, 225

Ramatulasi Tammineni, 437

Ravi Payal, 1

Ravi Raja, S. Y., 191

Ravi Sankar Sangam, 29

Rekha Madhuri, M., 259

Renuka, N., 487

Rishabh Gupta, 399

RishiPal Singh, 407

L

Likhitha, N., 181

Lydia D. Isaac, 305

S

Sagar Bafna, 357

Sahil Sobhani, 347

Sahithi, K., 113

Sai Krishna Santosh, G., 475

Sajal Saxena, 347

Sampath, D., 273

Sandeep Kumar Panda, 237

Sandeep Samantaray, 315

Sandeep, Y., 259

Sangeeta Viswanadham, 51

Saroj Kumar Nanda, 237

Sasibhushana Rao Pappu, 517

Satish Kumar Patnala, 517

Sejal Balasaheb Shirsale, 347

Shivam Yadav, 143

Shushrutha, K. S., 103

Sirisha, G. N. V. G., 437

Sitharam Sai Kiran, K., 181

Somya Goyal, 347, 357, 391, 399

Sravani Potula, 131

Sreenivasa Rao Ijjada, 113, 131

Sridhar, B., 181, 191

Srilatha, K., 113

Sriya Datta, T., 259

Srujan Raju, K., 497, 533

Subba Reddy Chavva, 29

Subha Mastan Rao, T., 517

Sudheer Kumar Terlapu, 419

Sukhwinder Singh, 41

N

Nabin K. Gantayet, 315

Nagendra Babu, P., 191

Naivedhya Sharma, 399

Nandini Singh, 67

Navya Sri Kalepalli, 465

Nihar Asare, 297

Nihar Ranjan Mohanta, 315

Nikhil, K. V., 181

Nilotpal Pathak, 457

Nithisha, M., 259

Sumam Mary Idicula, [337](#)
Sumit Anilkumar Singh, [297](#)
Sumitra Motade, [285](#)
Suresh Chandra Satapathy, [237](#)
Sushil Kumar, [407](#)
Sushma, M., [113](#)
Suwarna Gothane, [497](#)
Swasti Singhal, [457](#)

T

Tejendra Kumar Singh, [41](#)

U

Ujjawal K. Singh, [315](#)
Umayia Mushtaq, [19](#)

V

Vardhan, K. P. V. A., [181](#)
Vartika Paharia, [347](#)
Venkata Subbarao, M., [419](#)
Venubabu Rachapudi, [517](#)
Vijaya Baskar, V., [429](#)
Vikas Tiwari, [533](#)
Vikrant Bhateja, [203](#)
Viney Rangasamy, B., [79](#)
Vinoth Kumar, R., [121](#)
Vipin Kumar, K. S., [89](#)
Virendra Singh Rathore, [225](#)

Y

Yohannes Bekuma Bakare, [59](#)
Yuvraj Singh, [391](#)