

Lecture Notes in Electrical Engineering 902

Sourav Dhar

Dinh-Thuan Do

Samarendra Nath Sur

Howard Chuan-Ming Liu *Editors*

Advances in Communication, Devices and Networking

Proceedings of ICCDN 2021

 Springer

Lecture Notes in Electrical Engineering

Volume 902

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

Luca Oneto, Department of Informatics, Bioengineering, Robotics, University of Genova, Genova, Genova, Italy

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. ****

Sourav Dhar · Dinh-Thuan Do ·
Samarendra Nath Sur · Howard Chuan-Ming Liu
Editors

Advances in Communication, Devices and Networking

Proceedings of ICCDN 2021

 Springer

Editors

Sourav Dhar
Department of Electronics
and Communication Engineering
Sikkim Manipal Institute of Technology
Rangpo, Sikkim, India

Dinh-Thuan Do
Department of Computer Science
and Information Engineering
Asia University
Taichung City, Taiwan

Samarendra Nath Sur
Department of Electronics
and Communication Engineering
Sikkim Manipal Institute of Technology
Rangpo, Sikkim, India

Howard Chuan-Ming Liu
Department of Computer Science
and Information Engineering
National Taipei University of Technology
Taipei, Taiwan

ISSN 1876-1100

ISSN 1876-1119 (electronic)

Lecture Notes in Electrical Engineering

ISBN 978-981-19-2003-5

ISBN 978-981-19-2004-2 (eBook)

<https://doi.org/10.1007/978-981-19-2004-2>

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

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

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

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

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

Preface

ICCDN 2021, the 5th International Conference on Communication, Device and Networking, is held in the Department of Electronics and Communication Engineering (ECE) of Sikkim Manipal Institute of Technology (SMIT), Sikkim Manipal University (SMU), Sikkim, December 15–16, 2021. The publication of ICCDN is sponsored by LNEE, Springer (SCOPUS Indexed).

The aim of the conference is to provide a platform for researchers, engineers, academicians, and industry professionals to present their recent research works and to explore future trends in various areas of engineering. The conference also brings together both novice and experienced scientists and developers to explore newer scopes; collect new ideas; establish new cooperation between research groups; and exchange ideas, information, techniques, and applications in the field of electronics, communication, devices, and networking. The ICCDN-2021 committees rigorously invited submissions of manuscripts from researchers, scientists, engineers, students, and practitioners across the world related to the relevant themes and tracks of the conference. The call for papers of the conference was divided into six tracks as mentioned: Track 1: Electronics Devices and Nano-technology, Track 2: Signal Processing, Track 3: Microwave and Millimetre Wave Engineering, Track 4: AI and its Application, Track 5: Energy, Power and Control, Track 6: Communication and Networking. On behalf of the executive committee and the local organizing committee, we thank the authors and the participants for their original ideas, innovative applications, and practical achievements in the different fields related to the conference.

The conference is enriched with six keynote speeches each of 1 hour duration by eminent speakers from academia and industry. As speakers, we have with us Dr. Debdeep Sarkar from Indian Institute of Science (IISc), Bangalore, India; Dr. Tao Guo, PIESAT Information Technology Co, Ltd, China; Mr. Rampersad, EarthDaily Analytics, Vancouver, Canada; Dr. Rudra Sankar Dhar, NIT Mizoram, India; Dr. Vinayakumar Ravi, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia; and Dr. Angsuman Sarkar, Kalyani Government Engineering College, West Bengal, India.

A total of 126 papers have been received, out of which 58 papers (including 5 invited papers) have been accepted in the conference. Participants, from different parts of the country as well as across different countries, presented their paper through online platform. All these efforts undertaken by the organizing committees lead to a high-quality technical conference program, which featured high-impact presentations from keynote speakers and from paper presenters. On behalf of the ICCDN organizing committee, we would like to thank Springer for the kind cooperation. We would like to thank the chief patrons, patron, advisory board members, general chair, and program chair for their continuous support and encouragements. Finally, we would also like to express our appreciation to all contributors for submitting highly accomplished work, to all reviewers for their time and valuable comments, and to all members of the conference committees for giving generously of their time over the past year.

Rangpo, India
Taichung City, Taiwan
Rangpo, India
Taipei, Taiwan

Dr. Sourav Dhar
Dr. Dinh-Thuan Do
Dr. Samarendra Nath Sur
Dr. Howard Chuan-Ming Liu

Contents

Novel Area Effective Designs for Full Adder and Full Subtractor Using QCA	1
K. Bhagya Lakshmi, D. Ajitha, and Y. Sujatha	
Study of All-Optical Directional Coupler Based on Holes in Slab Photonic Crystal Structure	15
Anup Sharma, Haraprasad Mondal, and Kamanashis Goswami	
All-Optical Feynman Gate Using Frequency Encoding Scheme, Add/Drop Multiplexer and Reflective Semiconductor Optical Amplifier with Simulative Verification	25
Surajit Bosu and Baibaswata Bhattacharjee	
Effects of Dimensional Variations on Short Channel Parameters in 14 nm Channel Length TG–SOI FinFETs	35
Priyanka Saha, Swagat Nanda, Potaraju Yugender, and Rudra Sankar Dhar	
Dibit-Based 4-Bit Parity Generator Using Reflective Semiconductor Optical Amplifier and Frequency Encoding Scheme	45
Surajit Bosu and Baibaswata Bhattacharjee	
Design and Implementation of an Efficient QCA-Based Multilayer Multi-Bit Parallel Shift Register Using Reversible Level-Sensitive ‘D’ Flip-Flop	57
Rupsa Roy, Swarup Sarkar, and Sourav Dhar	
A 12-bit Low-power 50MS/s SAR ADC Optimized for Speed and Power in 45 nm CMOS Technology	71
M. S. Akshatha and M. Nagabushanam	
Decision Tree-Based Classification of sEMG and Accelerometer Data of Sign Language	87
Akhtar Ismail Nadaf and Sanjay A. Pardeshi	

Bibliometric Analysis of Published Literature on Mobile Healthcare in the Past One Decade	99
Saibal Kumar Saha, Sangita Saha, and Ajeya Jha	
Shetkari Mitra App—An Application to Maximize the Profit of Farmers	111
Niranjan Girhe, Divya Chaudhari, Prachi Channe, and Avinash Bhute	
Performance Analysis of Spectrum Sensing Algorithms	129
Jaya Sharma and Kirti Vyas	
Impact of Benefits on Preference for Traditional Detailing	137
Saibal Kumar Saha, Bibeth Sharma, Sonia Munjal, and Ajeya Jha	
Perceptual Difference Between Patient and Physician on Negative Aspects of Social Media Promotion	145
Samrat Kumar Mukherjee, Jitendra Kumar, Vivek Pandey, Jaya Rani Rani, Abhijit Sarkar, and Ajeya K. Jha	
Impact of Ethicality and Marginalized Group Convenience on Social Media Promotion of Branded Drugs	153
Samrat Kumar Mukherjee, Jitendra Kumar, Abhijit Sarkar, Bibeth Sharma, Jaya Rani Rani, and Ajeya K. Jha	
Tracking of Maximum Power of Solar PV Array Under Partial Shading Condition Using Grey Wolf Optimization Algorithm	161
Neha Cintury, Soumyarupa Saha, and Chitrangada Roy	
Yield Prediction of Indian Crops Based on Weather Data	173
P. Athulya and B. Mohammed Ismail	
Categorization of Diabetic Retinopathy Applying Ensemble Model	183
R. Shekhar and T. Sridhar	
The Utility of Simulink Subsystems in Handling and Processing of Biomedical Signals and Images	199
S. Sivaarunagirinathan, D. Jeyakumari, M. Sundar Prakash Balaji, P. Thanapal, and V. Elamaran	
Hum Noise and Breathing Interference Removal from an ECG Signal with Finite Word Length Effects	211
B. Ajith Bala, D. Jeyakumari, M. Sundar Prakash Balaji, G. Sasi, and V. Elamaran	
Online Affinity of Instructing Methods for Using Personal-Efficacy and Reading Capacious During Covid-19	223
V. PremaLatha, E. Sreedevi, S. Sivakumar, Soumya Ranjan Nayak, Akash Kumar Bhoi, and Uttam Ghosh	

Reliable Biometric Authentication with Privacy Protection 233
 Shilpi Barman Sharma, Ishika Dhall, Soumya Ranjan Nayak,
 and Pushpita Chatterjee

**Time Series Analysis of Cryptocurrency: Factors and Its
 Prospective** 251
 Sahil Sejwal, Kartik Aggarwal, Soumya Ranjan Nayak,
 and Joseph Bamidele Awotunde

**A Review on Internet of Things: Communication Protocols,
 Wireless Technologies, and Applications** 265
 Meenu Garg, Gurjinder Kaur, Gurmehr Singh, Gursharan Sandhu,
 Shefali Gupta, Soumya Ranjan Nayak, and Muhammad Fazal Ijaz

**Digitization of Inpatient Medical Records Using Electronic
 Writing Pads in a Teaching Hospital** 279
 Deepak Betadur, G. Somu, and P. Naveen Kumar

**Miniaturization of Dual Shaped Monopole Antenna for UWB
 Application** 289
 Ranjeet Kumar, Rashmi Sinha, Arvind Choubey,
 Santosh Kumar Mahto, Pravesh Pal, and Praveen Kumar

**Improvement of Isolation Between Co-polarization
 and Cross-polarization Radiation Using Circular Shorting
 Columns** 299
 Manoj Sarkar, L. Lolit Kumar Singh, Sudipta Chattopadhyay,
 and Abhijyoti Ghosh

**Band Rejection in Wideband Partial Ground Plane Antennas
 Using Defected Ground Structure** 309
 Boyapati Bharathidevi, Jayendra Kumar, and Narayana Rao Palepu

**Triple-Band Polarization Independent C-Band Metamaterial
 Absorber** 319
 Kashish Mahindroo, Vani Sadadiwala, Vimlesh Singh,
 Devender Sharma, and Sarthak Singhal

**A Study on the Ionospheric Propagation Using GNSS Receiver
 Over Hill Region** 327
 Sankha Subhra Debnath and Swastika Chakraborty

**Ka Band Tropospheric Scintillation Estimation Over North East
 Indian Region** 339
 Nirmal Rai, Swastika Chakravarty, Kapila Sharma, Gopal Thapa,
 and Rinkila Bhutia

**Compact Insert Fed Monopole Antenna with Four L-Slot Insertion
 Design for Wireless Applications** 347
 Y. Venkata Lakshmaiah, Bappadittya Roy, and R. Dewan

Deep Learning Approach for Wind Power Forecasting	355
Nishant Saxena, Rahul Kumar, Rachit Saxena, Sri Krishna Mishra, Ritu Jain, and Sujit Kumar Verma	
Driver Drowsiness Detection System	369
Pratik Mahapatra, Shivam Raj, and Amrita Biswas	
A Novel Deep Learning Approach for Non-invasive Blood Glucose Measurement from Photoplethysmography Signals	377
Gautham Reddy, Karthik K. Bhat, Umang Lunia, and Niranjana Krupa	
Machine Learning Approaches on Intrusion Detection System: A Holistic Review	387
Pinakshi De and Ira Nath	
Human Emotion Prediction Analysis on Post-COVID-19 Crisis in Digital Media Using Deep Learning	401
Nikita Agarwal and Ritam Dutta	
Role of Blockchain for Sustainability and Circular Economy	413
Anuj Kumar, Monika Arora, Kuldeep Bhalerao, and Meghna Chhabra	
Comparative Analysis of Machine Learning Models to Predict Stock Market Price	427
Shreya Sakshi, Arjun Kar, and Chitrapriya Ningthoujam	
Hybrid Features-Based Ensembled Residual Convolutional Neural Network for Bird Acoustic Identification	437
Hari Theivaprakasham, V. Sowmya, Vinayakumar Ravi, E. A. Gopalakrishnan, and K. P. Soman	
Lung Disease Prediction Using Deep Learning	447
Debasree Mitra, Pranati Rakshit, Anjali Jha, Dristi Dugar, and Kamran Iqbal	
Implementation of Different Classification and Prediction Models on Skin Cancer Using Deep Learning Techniques	461
Debasree Mitra and Pranati Rakshit	
Segmentation and Classification of Skin Cancer Using K-means Clustering and EfficientNetB0 Model	471
Vatsala Anand, Sheifali Gupta, Deepika Koundal, Soumya Ranjan Nayak, Jana Shafi, and Akash Kumar Bhoi	
Comparative Assessment of Performances of Various Machine Learning Algorithms in Detection of Liver Ailments	483
Dwaipayana Saha, Indrani Mukherjee, Jesmin Roy, and Pranati Rakshit	
Sentiment Analysis of Twitter Data Using Deep Learning	495
Pranati Rakshit, Pronit Sarkar, Debosmita Ghosh, Shubhankar Roy, Subhadip Talukder, and Partha Sarathi Chakraborty	

Stun Gun PCB Circuit with Arduino Power Shield 503
 Mehaboob Mujawar, D. Vijaya Saradhi, and K. Swetha

Comparative Study on Tuning PID and FOPID Using Genetic Algorithm for Heart Rate Control of a Pacemaker 511
 Ritika Saini, Ravi Kumar, Rijhi Dey, Ujjwal Mondal, and Rudra Sankar Dhar

Finite Element Method Based Determination of Non-linear Inductances of Three Phase Induction Motor 519
 Mohammed Nasir Ansari

Design and Comparison of RC, RC-PD and RC-PID Controller Architectures in a Servo-Motor System 529
 Litisha Mohapatra, Nishant Gupta, Rijhi Dey, Ujjwal Mondal, and Rudra Sankar Dhar

Effective Tone Control Circuit Using Proteus 535
 R. B. Dharanidaran and Bappaditya Roy

Tetra Bot—A Multi-purpose Robot 545
 B. Sricharan, K. Hima Bindu, A. Laasya Lata, and D. Ajitha

Getting Started with LPWAN: LoRa, Sigfox and NB-IoT 559
 Shridhar Sharma

Enabling Cognitive Radio in NOMA-Assisted Reconfigurable Intelligent Surfaces: Outage Performance Analysis 569
 Arjun Chakravarthi Pogaku, Nhan Duc Nguyen, Anh-Tu Le, and Dinh-Thuan Do

An Intelligent Vehicular Communication-Based Framework to Provide Seamless Connectivity in WBAN 583
 Koushik Karmakar, Sohail Saif, Suparna Biswas, and Sarmistha Neogy

A Novel RAW Slot Allocation Scheme for Improving the Performance of IEEE 802.11ah Multi-rate IoT Networks 593
 Badarla Sri Pavan and V. P. Harigovindan

Optimal Allocation of Micro-phasor Measurement Units in Distribution Network Considering Security Constraints 603
 Manam Ravindra, Donepudi Tata Rao, Rayapudi Srinivasa Rao, Adireddy Ramesh, and Karri Manoz Kumar Reddy

Home Automation Using Packet Tracer 611
 Pooja Reddy Bathula, Snigdha Pv, Laasya Lata Anumakonda, Mohammed Mahaboob Basha, and Pandya Vyomal Naishadhkumar

Interpretation of Wireless Communication Using OFDM Technology 625
Bipasha Chakrabarti, B. Roy, Priyanka S. Das, Prajit Paul,
and A. K. Bhattacharjee

Ergodic Capacity Analysis of RIS-aided System Relying on User Grouping and Fixed Power Allocation 637
Kaveti Umamaheswari, Fazal-E-Asim, and Dinh-Thuan Do

Author Index 647

About the Editors

Dr. Sourav Dhar (Senior Member, IEEE) is currently Professor and Head of the ECE Department, SMIT. His current research interests include IoT, WSN, remote sensing and microwave filter design. He is Member of IEEE, the IEEE-GRSS society and IEI, India. He has published more than 30 papers in SCI/Scopus indexed international journals and at conferences. He also serves as Reviewer for *Wireless Personal Communication*, *IEEE Transactions on Vehicular Technology* and several other journals and conferences.


Dr. Dinh-Thuan Do (Senior Member, IEEE) received the B.S., M. Eng. and Ph.D. degrees from Vietnam National University (VNU-HCMC), in 2003, 2007 and 2013, respectively, all in Communications Engineering. He was Visiting Ph.D. student with the Communications Engineering Institute, National Tsing Hua University, Taiwan, from 2009 to 2010. Prior to joining Ton Duc Thang University, he was Senior Engineer with VinaPhone Mobile Network, from 2003 to 2009. He was a founder of Wireless Communications Research Group funded by Ton Duc Thang University during period from 2016 to 2020. He was a supervisor for numerous Ph.D. students in the Ph.D. joint program managed by Ton Duc Thang University and Technical University of Ostrava (VSB), Czech Republic. He was a recipient of Golden Globe Award from Vietnam Ministry of Science and Technology in 2015 (Top 10 most excellent scientist nationwide). He currently works as Assistant Professor in the Department of Computer Science and Information Engineering, Asia University (Taiwan). His publications include 1 textbook, 7 book chapters, over 95 SCIE-indexed journal articles, over 100 other journal articles and international conference papers. According Stanford's list, he was recognized as a scholar of top 2% scientists in the world (2020). He is currently serving as Editor of *Computer Communications* (Elsevier), Associate Editor of *EURASIP Journal on Wireless Communications and Networking* (Springer), Associate Editor of *Electronics* and an Editor of *KSII Transactions on Internet and Information Systems*.

Dr. Samarendra Nath Sur (Senior Member, IEEE) is Assistant Professor (SG) at the ECE Department, Sikkim Manipal Institute of Technology, India. His current research interests include broadband wireless communication, advanced digital signal processing and remote sensing. He is the recipient of the University Medal and Dr. S. C. Mukherjee Memorial Gold Centered Silver Medal from Jadavpur University in 2007. He is a member of IEEE, IEEE-IoT, IEEE-SPS, IEI, India, and IAENG. He has published more than 55 papers in SCI/Scopus indexed international journals and at conferences. He also serves as a reviewer for the *International Journal of Electronics*, *IET Communication*, *Ad Hoc Networks* and *IEEE Transactions on Signal Processing*. He is currently editing several books with Springer Nature, Elsevier, and Routledge and CRC Press. He is also serving as Guest Editor for special issues of the journal like Springer Nature and Hindawi.

Dr. Howard Chuan-Ming Liu is Professor in the Department of CSIE, National Taipei University of Technology, Taiwan, where he was the Department Chair from 2013 to 2017. Currently, he is Head of the Extension Education Center at the same school. He has published more than 80 papers in various prestigious journals and at international conferences. His current research interests include big data management and processing, uncertain data management, data science, spatial data processing, data streams, ad hoc and sensor networks and location-based services.

Novel Area Effective Designs for Full Adder and Full Subtractor Using QCA



K. Bhagya Lakshmi , D. Ajitha, and Y. Sujatha

Abstract Full adder and full subtractor are often a combinational logic circuit that performs three one-bit binary digits of addition and subtraction operations, respectively. Full adder is an important element for the development of different devices such as microprocessors and digital signal processors. The design of the area-efficient full adder is essential to build an area-efficient processor. In this paper, we have implemented an area-efficient full adder circuit and full subtractor circuit using quantum dot cellular automata (QCA). A better EX-OR gate configuration in terms of efficient area is chosen to create the full adder and full subtractor designs. The proposed full subtractor and full adder architecture requires 22 and 23 quantum cells, respectively, whose effective area is $0.0192 \mu\text{m}^2$ (micro-metre square) only. A tool called QCADesigner is used to implement the design and check its performance. The dissipation of energy is measured using the QCAD-E tool.

Keywords Complementary metal–oxide–semiconductor (CMOS) technology · Exclusive-OR (EX-OR) gate · Full adder (FA) · Full subtractor (FS) · Quantum dot cellular automata (QCA) · QCADesigner

1 Introduction

A remarkable development in complementary metal–oxide–semiconductor (CMOS) technology built IC has been noticed in the last few decades. Further scaling down of MOSFETs beyond its limit leads to various physical constraints [1]. Gordon Moore

K. Bhagya Lakshmi (✉)

Department of E.C.E, Sasi Institute of Technology and Engineering, Tadepalligudem, West Godavari District, Andhra Pradesh, India

D. Ajitha

School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology (VIT), Vellore, Tamil Nadu 632014, India
e-mail: ajitha.d@vit.ac.in

Y. Sujatha

Department of E.C.E, Sri Vasavi Engineering College, Andhra Pradesh, Tadepalligudem, West Godavari District 534101, India

from Intel Company predicted that transistors count fabricated on a chip doubles approximately for every 18 months. His forecast has been precise for over 40 years, and the name given to this concept is the Moore's law [2]. Many researchers are therefore trying to discover a technology that replaces CMOS technology. QCA is one of the new nanotechnologies that is a better substitute for CMOS technology and ideal for the design of nanoscale circuits. [3, 4].

During the last decade, significant focus is attained by QCA due to its advantages. Several combinational circuits like Exclusive-OR (XOR) gate structures are proposed with different structures [5–18]. Out of these, the five input majority voter gate, adder and subtractor designs gained much attention to achieve the layout structures in a single layer. Adder and subtractor are used in arithmetic and logic unit to compute addition, subtraction, multiplication and division. In this work, the proposed designs can be used in the devices that use increment or arithmetic process and also used in microcontroller and microprocessors.

The goal of this study is to use a novel XOR gate design in QCA to develop the structures of area-efficient full adder and full subtractor designs. Almost all digital signal processing (DSP) systems or very large-scale (VLSI) integrated circuits (IC) use multipliers, which are having inbuilt adders, for their computations. Therefore, the DSP systems or VLSI ICs will be area efficient if the adders are designed with less number of cells. The paper structure is as follows. Sections 2 and 3 brief the overview of QCA. The proposed coplanar structure of XOR gate, adder and subtractor is explained in parts 4, 5 and 6, respectively. Part 7 performance analysis and conclusion are presented in Sect. 8.

2 QCA Overview

Four quantum dots joined over quantum mechanical tunnelling barriers and organized in a square are called a QCA cell. Quantum dot is a place where electron can exist. The QCA cell contains only two electrons which oppose each other. Although electrons can tunnel between quantum dots, they are unable to leave the QCA cell. The transfer of data or processing in QCA circuits depends on the columbic interaction of electrons. Two polarizations are possible. They are '− 1' polarization and '+ 1' polarization. Depending upon the polarization, the binary data is encrypted [5].

Hence, $P = -1$ represents the logic '0' and $P = +1$ represents the logic '1'. The QCA cell is shown in Fig. 1b, and its polarization for logic '0' is shown in Fig. 1a. Binary calculation needs communication between the cells. A cell placed nearer to another cell, then due to columbic interaction among the cells, removes the degeneracy.

The electrostatic interactions among the electrons of QCA cells communicate the data from one (input) end to another (output) end in a QCA wire. The transmission of binary data in a normal QCA wire is shown in Fig. 2a. A 45°rotated QCA wire [6] is shown in Fig. 2b.

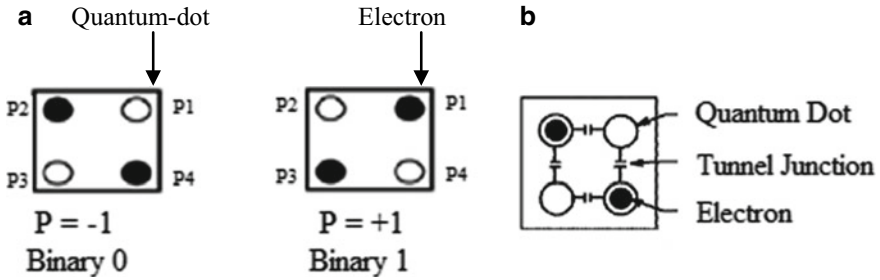


Fig. 1 a QCA cell polarization for logic ‘0’ and logic ‘1’ b QCA cell

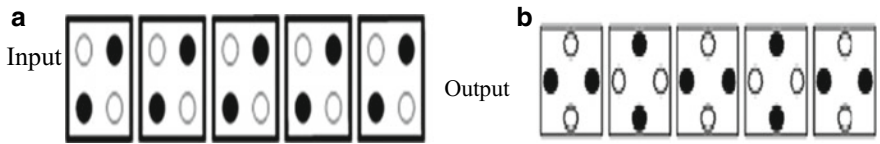


Fig. 2 Wire in QCA a normal cells b 45° rotated cells

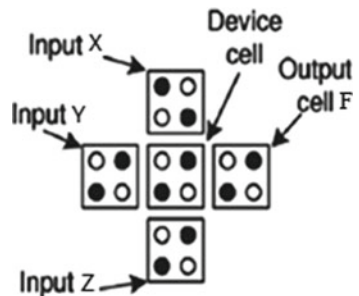
3 Basic QCA Gates

The basic QCA gates include majority voter gate (MVG) and NOT gate or inverter.

4 Majority Voter Gate (MVG)

MVG is a basic building block for construction of QCA digital logic circuits, as shown in Fig. 3. The output in MVG depends on the majority of the input logic. A three-input MVG consists of five QCA cells. Four terminal cells out of which three represent input terminals and one represents the output cell [7]. The middle cell is device cell. The logic function of MVG is explained in (1)

Fig. 3 Majority voter gate structure using QCA cells



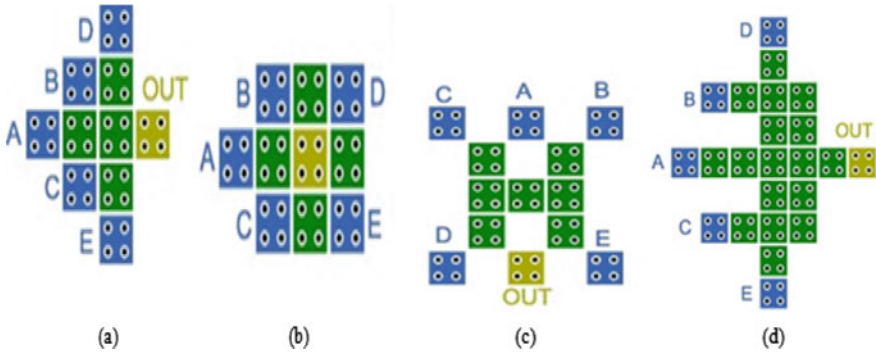


Fig. 4 Various designs of five-input majority voter gate

$$F = \text{Majority}(XYZ) = XZ + YZ + XY \tag{1}$$

Different five-input majority voter gates [8] are shown in Fig. 4. The working of five inputs MVG is similar to the working of three inputs MVG, and the only difference is that it takes five inputs.

5 Construction of AND and OR Gates

When one of the inputs of a three-input MVG is polarized to ‘- 1’, two-input AND gate is obtained. Similarly when one of the inputs of a three-input MVG is polarized to ‘+ 1’, two-input OR gate [5] is obtained as represented in Fig. 5a, b.

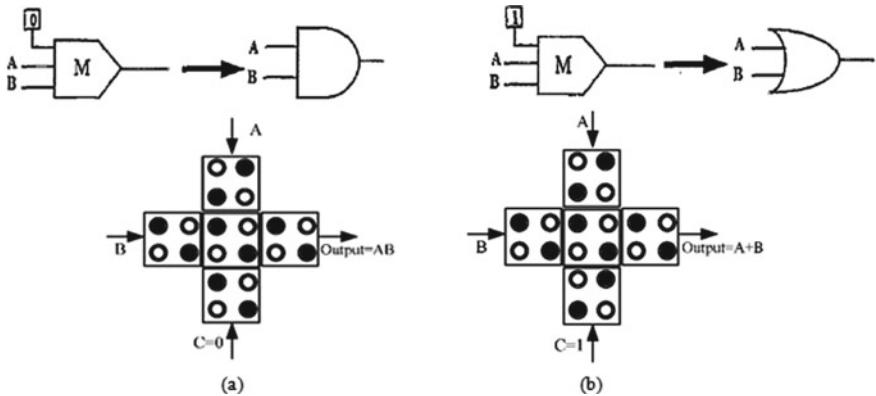


Fig. 5 a Two-input AND gate b two-input OR gate

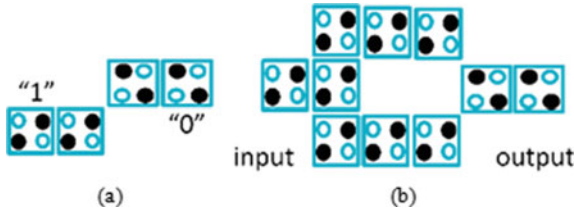


Fig. 6 Two different representations of QCA Inverter

6 QCA Inverter

The available QCA inverter designs are shown in Fig. 6. An inverter which is shown in Fig. 6a is designed by arranging cells in diagonal manner. The electrons in the diagonally oriented cells interact electrostatically, causing the input and output to always align in reverse polarization [9]. The inverter in Fig. 6b is the basic design. It is made to arrange the two electrons in the cell's opposing quantum wells in response to the input by diagonal QCA cells that make Columbic interactions.

7 Clocking Schemes in QCA

The important factor in implementing the QCA circuits is the usage of clock which coordinates, provides enough power to drive the circuit and also manages data flow [9]. The timing scheme arrangement depends on four non-overlapping clocks which manages the flow of data through cells. The four clock zones of QCA are shown in Fig. 7.

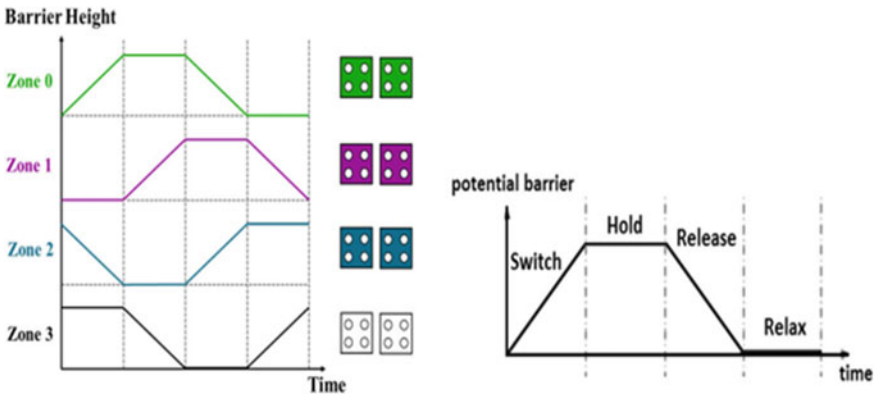
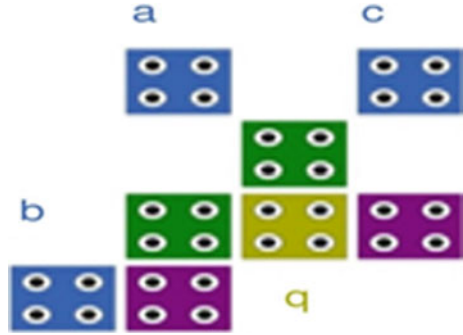


Fig. 7 QCA clocking

Fig. 8 Proposed three-input XOR gate



At hold phase zone, there is more tunnelling barrier that does not allow the state to change, however it will affect neighbouring zones. Finally, the release and the relax state limit the tunnelling barrier, thus the zone is not having impact on remaining zones. The shape of these zones can be uneven; however, their size needs to be in certain bounds forced by production and dissipation concerns. The placement of these zones should be proper, and thus it is difficult to design with efficiency. Thus, the clocking schemes in QCA make the design, which differ from CMOS circuits.

8 QCA EX-OR Gate

An effective three-input EX-OR gate [19] is designed by arranging QCA cells in an ordered manner which is shown in Fig. 8. This proposed design does not employ majority voter gate. This three-input EX-OR gate covers only eight quantum cells with effective area of $0.0016\mu\text{m}^2$. The number of clock zones used is 2, thus latency is just 0.5.

9 The Layout of Currently Designed Full Adder in QCA

A full adder (FA) is a binary combinational circuit that performs the addition operation among three input bits and gives the output of sum (S) and carry (C).

The three inputs a, b, c denote the first input, second input and previous carry, respectively. An expression for Sum and Carry is given by (2) and (3).

$$\text{Sum (S)} = c \text{ xor } b \text{ xor } a \quad (2)$$

$$\text{Carry (C)} = ab + ca + bc \quad (3)$$

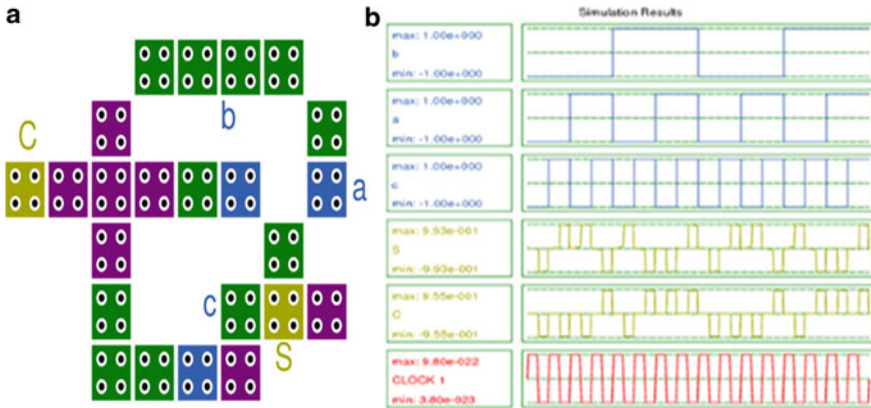


Fig. 9 a Proposed full adder b simulated results

A three-input XOR gate is used to generate Sum, and a three-input MVG is used to generate Carry. The construction of full adder using QCA is given below in Fig. 9a, and their simulated results are shown in Fig. 9b.

The Columbic interactions between the closely placed quantum cells and the clocking schemes used in the proposed structure result in the optimized structure of full adder by using the XOR gate as shown in Fig. 8. The same concept is extended to construct the full subtractor also, and the same is given in the following section.

10 The Layout of Proposed Full Subtractor in QCA

A digital combinational circuit that is known as a full subtractor (FS) to perform subtraction is constructed using QCA. It takes three input bits, namely subtrahend, minuend and previous borrow to generate the Difference (D) and Borrow (B) as output. Expressions for Difference and Borrow are given by (4) and (5) and the truth table in Table 2

$$\text{Difference } (D) = a \text{ xor } b \text{ xor } c \tag{4}$$

$$\text{Borrow } (B) = c'b + b'a + c'a \tag{5}$$

A three-input XOR gate is used to generate Difference, and a three-input MVG is used to generate Borrow. Construction of full subtractor using QCA is given below in Fig. 10a, and their simulated results are shown in Fig. 10b. Borrow is achieved by the majority voter gate. The difference is achieved by the XOR gate.

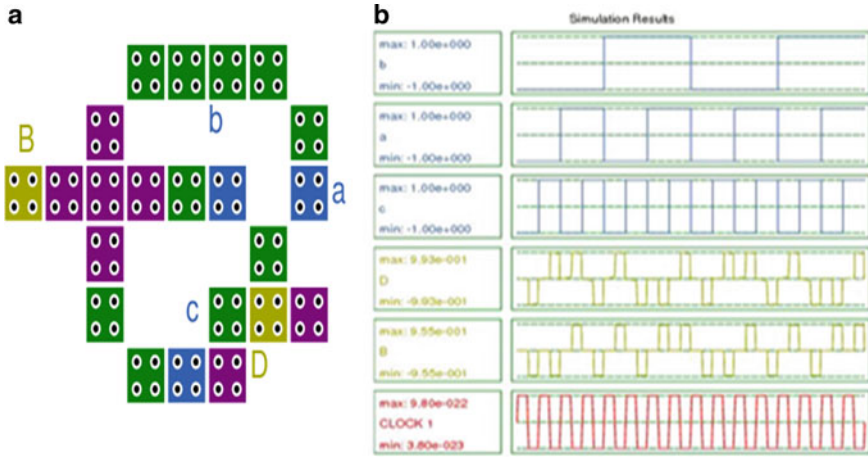


Fig. 10 a Proposed full subtractor b simulated results

11 Performance Analysis

For verifying the QCA circuit’s functionality, simulation is performed by a software tool QCA Designer version 2.0.3 [10]. The proposed full adder and full subtractor designs were compared to previously existing full subtractor and full adder designs as given in Table 1.

Table 1 Comparison analysis

Circuit	Cell count	Area (μm^2)	No. of clock zones	Latency
[11]	178	0.205	8	2
[12]	136	0.168	7	1.75
[13]	104	0.1043	7	1.75
[14]	32	0.028	2	0.5
[15]	94	0.13	3	0.75
[16]	83	0.10	4	1
Proposed (1-bit full subtractor)	22	0.0192	2	0.5
[20]	79	0.05	5	1.25
[21]	71	0.06	6	1.5
[22]	102	0.09	8	2
[23]	44	0.06	5	1.25
Proposed (1-bit full Adder)	23	0.0192	2	0.5

Table 2 Total energy dissipation of full adder using QCAD-E

$E_{\text{bath_total}}$	1.3899e – 003	1.4193e – 003	1.9474e – 003	1.1260e – 003	7.6523e – 004	2.0675e – 003	1.2960e – 003	1.7619e – 003	1.3899e – 003	1.4193e – 003	1.9474e – 003
$E_{\text{clk_total}}$	1.9852e – 003	1.5763e – 003	8.7431e – 004	2.2619e – 003	2.2619e – 003	1.6564e – 003	8.7431e – 004	1.5762e – 003	1.1822e – 003	1.5763e – 003	8.7431e – 004
$E_{\text{error_total}}$	– 1.4749e – 004	– 1.4451e – 004	– 2.0830e – 004	– 1.1384e – 004	– 7.1207e – 005	– 2.2166e – 004	– 1.2967e – 004	– 1.8514e – 004	– 1.4749e – 004	– 1.4451e – 004	– 2.0830e – 004

The construction of full subtractor is done by using nine MVG as seven AND gates and two OR gates along with six invertors [11]. Each individual term is implemented by the relevant gates following the cell count, and its complexity has been increased. The implementation of full subtractor is done by using nine MVG as six AND gates and three OR gates along with four invertors [12]. The design of full subtractor [13] implementation is done by designing a two-input XOR gate by using a five-input MVG, three-input majority gate and with two invertors. It requires two 2-input XOR gates and two 3-input MVG; this leads to more cells and occupies large area. A three-input XOR gate is implemented by using that Difference can be generated and Borrow is generated by using three-input majority gate and an NOT gate [14]. Nine three-input FNZ gates are used to implement full subtractor [15]. Further, a three-input XOR gate is used along with a 3-MVG to implement the full subtractor [16].

The full adder layout is accomplished by a three-input majority gate, five-input majority gate and a NOT gate. It has three layers so it is difficult to construct. A five-input majority gate, a three-input majority gate and two NOT gates are utilized in the existing design [21]. Three number of three-input majority gate and two NOT gates are used for the implementation of full adder [22]. Further the design of full adder is done by two five-input majority gates and a three-input majority gate and with four NOT gate [23].

The comparison analysis is based on cell count, effective area, latency and finally number of clock zones used. Observation from Table 1: the proposed designs used less number of quantum cells than previously proposed designs [11–16] and [20–23], which leads to less effective area.

Proposed full subtractor compared to [14] cell count has reduced to 31.2%, in terms of area it is reduced to 31.4%. When compared to [16], cell count has reduced to 73.5%, in terms of area it is reduced to 80.8%, and clock zones used are reduced up to two clock zones.

In Table 2, the dissipation of energy of the 1-bit FA is shown. $E_{\text{bath_total}}$ is the energy separated for each clock, and sum E_{bath} and average E_{bath} are calculated by considering the energy transferred to the neighbouring cells and clock [24–26]. Total dissipation of energy (Sum_Ebath) of the FA is $1.56e-002$ eV. The average energy dissipation per cycle (Avg_Ebath) is $1.42e-003$ eV. Similarly, the energy dissipation of the 1-bit FA is shown in Table 3. Total dissipation of energy (Sum_Ebath) of FS is $1.65e-002$ eV, and the average dissipation of energy per cycle (Avg_Ebath) is $1.50e-003$ eV. Tables 2 and 3 clearly display that the energy dissipation of the proposed designs is less with minimum error.

Because of the optimal placing of QCA cells and proper use of clock zones, the proposed designs are more compact and have less complexity than previous systems. By using the presented EX-OR gate, the proposed FA and FS are designed. Therefore, the projected designs are more efficient in the aspects of all the performance metrics considered in the QCA designs. From the comparison table, it is clearly evident that both the proposed designs (full subtractor and adder) achieve the similar improvement over the existing designs.

Table 3 Total energy dissipation of full subtractor using QCAD-E

$E_{\text{bath_total}}$	7.2306e 004	2.1021e 003	1.2553e 003	1.6602e 003	1.3233e 003	1.3751e 003	1.9781e 003	1.0703e 003	7.2306e 004	2.1021e 003	1.2553e 003
$E_{\text{clk_total}}$	1.9852e 003	8.2341e 004	1.6564e 003	1.6175e 003	1.6175e 003	1.6564e 003	8.2341e 004	1.9852e 003	1.9852e 003	8.2341e 004	1.6564e 003
$E_{\text{error_total}}$	-7.1660e -005	-2.1894e -004	-1.7207e -004	-1.3210e -004	-1.4323e -004	-2.0402e -004	-1.0655e -004	-7.1660e -005	-2.1894e -004	-2.1894e -004	-1.2991e -004

12 Conclusion

In this paper, a unique design for three-input EX-OR gate is implemented with minimum number of cells and area. By using the proposed EX-OR gate, full subtractor and full adder circuits are designed in QCA that has been presented. The design is found to outperform with the existing models in cell count, cost and area. The introduced designs are superior with respect to the performance metrics considered in QCA. These designs can be used for constructing compact subtraction circuits for arithmetic operations, DSP and networking-based systems.

References

1. Lent CS, Tougaw PD, Porod W, Bernstein GH (1993) Quantum cellular automata. *Nanotechnology* 4(1):49–57
2. Moore GE (2006) Cramming more components onto integrated circuits. Reprinted from *electronics*, vol 38, no 8, pp 114–117, April 1965 ff. *IEEE solid-state circuits society newsletter* vol 11, no 3, pp 33–35 (2006)
3. Lent CS, Tougaw PD (1997) A device architecture for computing with quantum dots. In: *Proceedings of the IEEE*, vol 85, no 4, pp 541–557
4. Tougaw PD, Lent CS (1994) Logical devices implemented using quantum cellular automata. *J Appl Phys* 75(3):1818–1825
5. Beigh MR, Mustafa M, Ahmad F (2013) Performance evaluation of efficient XOR structures in quantum-dot cellular automata (QCA). *Circ Syst* 4(2)
6. Akeela R, Wagh MD (2011) A five-input majority gate in QCA. In: *NSTI Nanotech*, vol 2, pp 978–981
7. Rai S (2008) Majority gate based design for combinational quantum cellular automata (QCA) circuits. In: *2008 40th Southeastern symposium on system theory (SSST)*, IEEE, pp. 222–224
8. Balali M et al (2017) A novel design of 5-input majority gate in QCA technology. In: *2017 IEEE symposium on computer applications & industrial electronics (ISCAIE)*, pp 13–16
9. Vamsi V, Ottavi M, Lombardi F (2006) Clocking and cell placement for QCA. In: *2006 sixth IEEE conference on nanotechnology*, vol 1, pp 343–346
10. Walus K (2005) QCADesigner: a rapid design and simulation tool for QCA. *Internet J Nanotech Appl* 2(1):1–7
11. Lakshmi S, Athisha G, Karthikeyan M, Ganesh C (2010) Design of subtractor using nanotechnology based QCA. In: *2010 international conference on communication control and computing technologies*. IEEE, pp 384–388
12. Dallaki H, Mehran M (2015) Novel subtractor design based on QCA (QCA) nanotechnology. *Int J Nanosci Nanotechnol* 11(4):257–262
13. Reshi JI, Banday MT (2016) Efficient design of nano scale adder and subtractor circuits using quantum dot cellular automata. In: *3rd international conference on electrical, electronics, engineering trends, communication, optimization and sciences*, pp 6
14. Bahar AN, Waheed S, Hossain N, Asaduzzaman M (2018) A novel 3-input XOR function implementation in quantum dot-cellular automata with energy dissipation analysis. *Alexandria Eng J* 57(2):729–738
15. Wani WH, Mohammad BT, Patel S (2016) Design of a half/full subtractor on QCA. *IJSRSET* 2(3):133–136
16. Ramachandran SS, Jegadish Kumar KJ (2017) Design of a 1-bit half and full subtractor using a quantum dot cellular automata (QCA). In: *2017 IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI)*. IEEE, pp 2324–2327

17. Danehdaran F, Khosroshahy MB, Navi K, Bagherzadeh N (2018) Design and power analysis of new coplanar one-bit full-adder cell in QCA. *J Low Power Electron* 14(1):38–48
18. Majeed AH, Zainal MSB, Alkaldy E, Nor DM (2020) Full adder circuit design with novel lower complexity XOR gate in QCA technology. *Trans Electr Electron Mater* 1(10):198–207
19. Lakshmi BK, Tejaswi S, Vamsi SC, Jeevanarani B (2020) A novel area efficient parity generator and checker circuits design using QCA. In: 2020 international conference on inventive computation technologies (ICICT), IEEE Xplore Part Number:CFP20F70-ART
20. Hashemi S, Tehrani M, Navi K (2012) An efficient QCA full-adder. *Sci Res Essays* 7(2):177–189
21. Hashemi S, Navi K (2015) A novel robust QCA full-adder. *Proc Mater Sci* 11:376–380
22. Hänninen I, Takala J (2010) Binary adders on QCA. *J Signal Process Syst* 58(1):87–103
23. Zoka S, Gholami M (2019) A novel efficient full adder–subtractor in QCA nanotechnology. *Int Nano Letters* 9(1):51–54
24. Timler J, Lent CS (2002) Power gain and dissipation in quantum-dot cellular automata. *J Appl Phys* 91:823–831
25. Lent CS, Mo L, Yuhui L (2006) Bennett clocking of quantum-dot cellular automata and the limits to binary logic scaling. *Nanotechnology* 17:40–42
26. Timler J, Lent CS (2003) Maxwell’s demon and quantum-dot cellular automata. *J Appl Phys* 94:1050–1060

Study of All-Optical Directional Coupler Based on Holes in Slab Photonic Crystal Structure



Anup Sharma, Haraprasad Mondal, and Kamanashis Goswami

Abstract In this paper, an investigation on all-optical direction couplers has been demonstrated. A two-dimensional, holes in slab, photonic crystal structure has been chosen for designing the directional coupler and its working principle is based on linear optics. Moreover, the plane wave expansion and finite difference time domain methods have been employed to analyze the photonic band gap (PBG) and electrical field distribution of the structure, respectively. Different designs of directional couplers have been analyzed like a single waveguide, double waveguide also with obstacle holes in between the waveguides. The outcome of the constructive/destructive interference due to phase difference has also been assessed. Moreover, the impact of variation of hole radius with respect to coupling length and coupled power has also been investigated.

Keywords Directional coupler · Photonics crystal · Linear optics · Hole in slab · Coupling length

1 Introduction

An all-optical integrated circuit is one of the upcoming frontiers to deal with the limitations faced by conventional electronics and communication circuits as it offers low power consumption and high throughput and lower latency [1]. The all-optical integrated circuits have been reported by various researchers all over the globe on various platforms such as semiconductor optical amplifier (SOA) [2–4], plasmonic [5–7], and photonic crystal (PhC) [8–10]. Out of which, photonic crystal (PhC) is one of the promising platforms for the development of photonic integrated circuit (PhC) due to its inherent ability to mold, switch and guide the propagation of light through the waveguide. Photonic crystal (PhC) was first introduced by Yablonovitch

A. Sharma · H. Mondal (✉)
Dibrugarh University, Dibrugarh, Assam, India
e-mail: mandal.haraprasad@gmail.com

K. Goswami
Indian Institute of Technology (ISM) Dhanbad, Dhanbad, Jharkhand, India

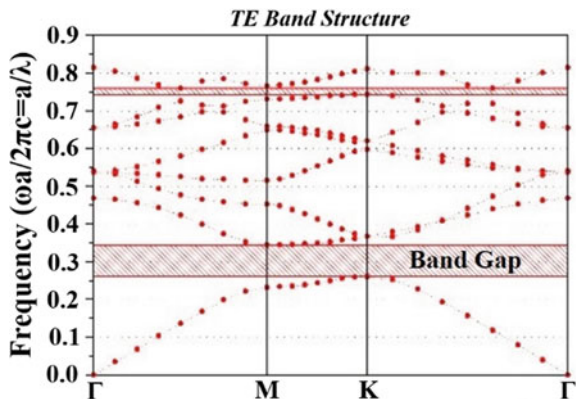
[11] and John [12] which can be fabricated in 1D, 2D and 3D that allows or forbids the propagation of light in bandgap at a very nanoscale range. By leveraging this extraordinary quality, lot of devices have been studied on photonic crystal platform such as demux [13, 14], Filter [15, 16], decoder [17–19], logic gates [20–22], polarizer [23], etc. Moreover, the optical coupler [24, 25] is one of the essential blocks in optical fiber communication, due to its ability to transfer signals from one waveguide to another. It can be achieved by placing two identical waveguides adjacent to each other, which is very simple to design for any optical application. Moreover, the optical coupler can transfer the signal in the same phase with higher directivity and isolation. Based on structural construction, optical couplers have been categorized as splitters, combiners, X-couplers, trees and stars [26]. Furthermore, the photonic integrated circuit (PIC) fabricated in the hole in slab demonstrate better mechanical stability as compare to rods in air structure [27]. In addition, the designs in linear mode offer low power loss and can be operated with a very low input power signal. Recently, Cuesta-Soto et al. [28] proposed an all-optical directional coupler based on the two-dimensional triangular lattice (rod in air) having the radius of $0.2a$, where $a = 511.5$ nm. In ΓK direction, two parallel rows of rods are removed to create a waveguide which is separated by a single row of defect rods with a reduced radius by $0.7r$. At 1.1 mm, and the full coupling has been observed at the normalized frequency of 0.3281 . The size of the PhC is found to be very small so might be very sensitive to tolerances in the fabrication processes. Similarly, Hojjat Sharifi et al. [29]. proposed an XNOR, NOR and AND logic gate by using the principle of the directional coupler. The nonlinear material Si nanocrystal (Si-nc) is used as nonlinear material in the Si-based triangular lattice having a radius of rods $0.2a$, where a is equal to 565 nm. The bit rate and response time are 1 Tbits/s and less than 1 ps, respectively. Besides being the nonlinear photonic crystal, the proposed system does not require high power values to operate. However, various parameters determine the efficiency of couplings like the wavelength of light, refractive indices of the material, input power intensity, the radius of the holes and the distance or hole between two waveguides. In this work, the impact of changing various parameters on optical directional coupler has been studied.

The remaining part of the paper is arranged in the following way: device description, theory and band diagram of the proposed structure are presented in Sect. 2. From Sect. 2.1 to 2.5, various arrangements have been done and explained such as single waveguide coupling, double waveguide coupling, Obstacles in the waveguide during coupling, constructive and destructive interference while coupling and effect of hole radius in coupling length. Finally, we present the conclusions in Sect. 3.

2 Device Design, Band Analysis and Coupling Theory

Figure 1 shows the schematic structure of the all-optical proposed directional coupler. The platform of the structure is considered as the hole in slab, and it is based on a two-dimensional photonic crystal platform. The dielectric slab has been organized in

Fig. 1 Photonic band gap of structure in TE mode



a hexagonal lattice structure. It composes of an array of 101×25 silicon slabs. The effective refractive index of the slab and the hole radius is considered to be 2.64 and $0.25a$, respectively, where ' a ' ($\equiv 450$ nm) is the lattice constant of the crystal. By using the plane wave expansion (PWE) method, the photonic band gap (PBG) has been obtained (as shown in Fig. 1) in transverse electric (TE) mode, which is ranging from $0.26168 (a/\lambda)$ to $0.34439(a/\lambda)$ that is between 1461 to 1932 nm in terms of wavelength.

When the light propagates through a medium, it has the ability to couple with another waveguide or optical cavities. The signal can travel back and forth from one waveguide to another. Moreover, the distance taken by a signal to be transferred from one waveguide to another at maximum is known as coupling length. Moreover, the power coupled is a crucial factor while designing the device, the monitor has been kept at coupling length (maximum coupling distance) to calculated the power received while performing the simulation. Sergey Kuchinsky et al. [30] reported a formula to calculate the coupling length.

$$L = 0.5 \frac{\Delta W}{W} V'_{\text{group}} \lambda_0$$

where v' group is the group velocity and $\Delta W/W$ is the change in the mode of frequency and λ_0 is the wavelength in vacuum.

2.1 Single Waveguide Coupling

Figure 2 shows the structure of a single waveguide coupler; the coupling has been observed back and forth between two waveguides. The array of 101×25 is considered to show the full effect of coupling, in which two waveguides have been created by creating a defect in the structure in the horizontal direction named as W1 and W2.

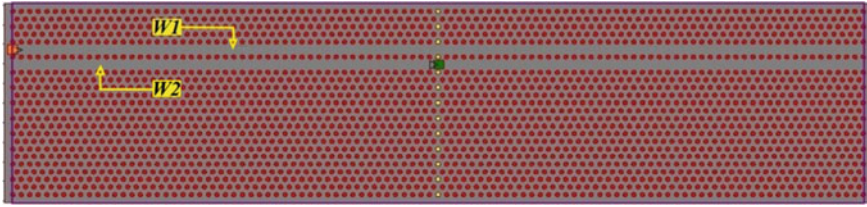


Fig. 2 Schematic structure of single waveguide directional coupler

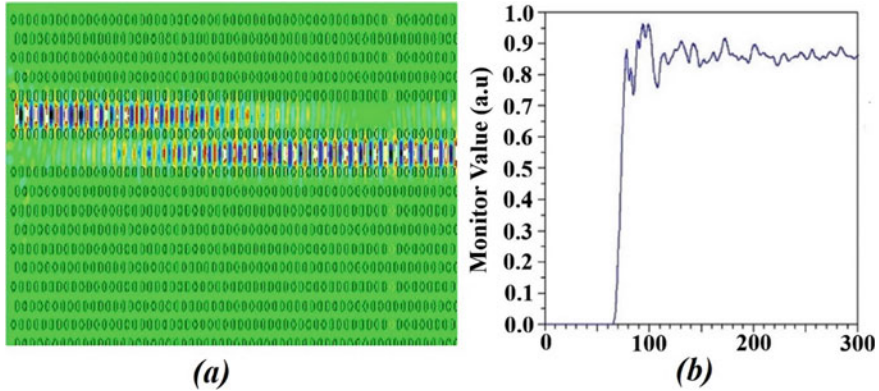


Fig. 3 **a** Electric field propagation through the single waveguide directional coupler. **b** Output power at $W2$ waveguide

The FDTD method has been used to obtain the electric field distribution within the waveguide. In Fig. 3a, the input signal of unity power has been applied in $W1$ waveguide; at the center of another waveguide ($W2$) the power coupled is maximum. Again, from the $W2$ waveguide the signal is transferred to waveguide 1. As per Fig. 3b, the power observed at monitor is 87%, some of the losses were observed due to evanescent field and the gap between the monitor width and adjacent top and bottom holes.

2.2 Double Waveguide Coupling

To study the double waveguide coupling, the size of the array has been increased by 50% in the horizontal direction, i.e., 151×25 as shown in Fig. 4. The signal power has been launched of the same intensity (as per Sect. 2.1). The signal is applied from $W1$ waveguide, and two monitors have been used to observe the power at $W2$ and $W3$ waveguides. One interesting phenomenon has been observed that when the signal is coupling from $W1$ to $W2$ then at the same time it started coupling from $W2$ to $W3$ ($W2$ achieve maximum) as shown in Fig. 5a. Due to this, the power observed

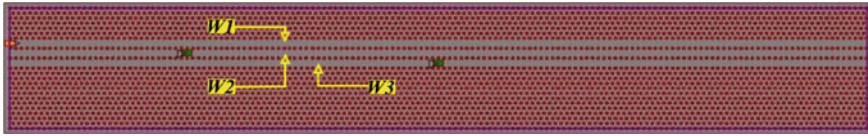


Fig. 4 Schematic structure of double waveguide directional coupler

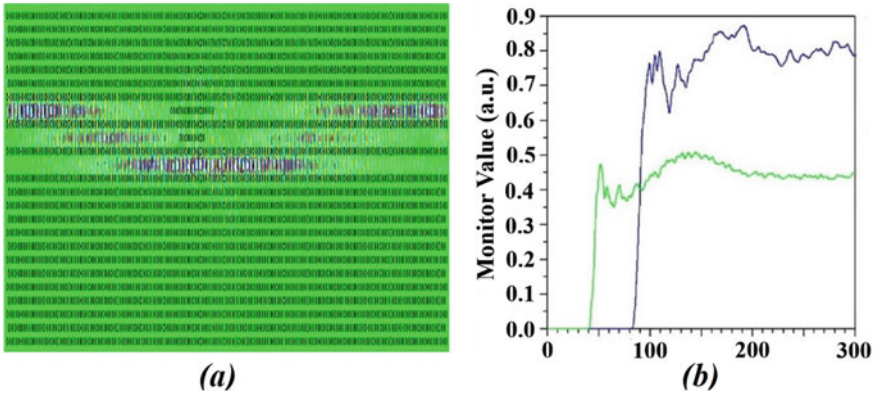


Fig. 5 **a** Electric field propagation through the double waveguides directional coupler. **b** Output power at W2 and W3 waveguides

at W2 is 44%, while at W3 it is 84% which is represented by green and blue lines, respectively, in the graph as shown in Fig. 5b. The length of the coupling has been found to be uneven in this case.

2.3 Obstacles in the Waveguide During Coupling

When the optical signal is transferred from one point to another, it is obvious to have some circuit (or any device) in between through which the signal shall not be passed. In this, we tried to study whether the obstacle affects the power coupled between two waveguides or not. As per Fig. 6, 15 and 9 holes of the same radius have been used

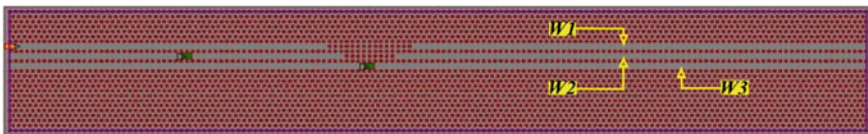


Fig. 6 Schematic structure of obstacle in the waveguide of directional coupler

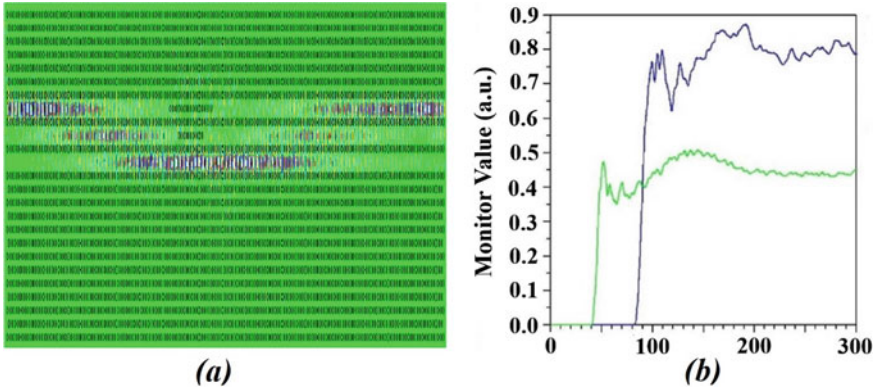


Fig. 7 a Electric field propagation through the obstacle based waveguides. b Output power at W2 and W3 waveguides

as an obstacle in W1 and W2, respectively. The simulation and the output power are shown in Fig. 7 which is almost the same as mentioned in Sect. 2.2.

2.4 Constructive and Destructive Interference While Coupling

In Fig. 8, constructive and destructive interference phenomenon while coupling is shown. In the first case, two signals have been launched of the same phase in W1 and W3 waveguides, due to constructive interference the signal gets coupled in W2 waveguide. However, in the second case, input signals have been launched from waveguide W4 and W6 waveguides at 0 and 180 phase angles respectively, which result in destructive interference at the W5 waveguide. The result is shown in Fig. 9.

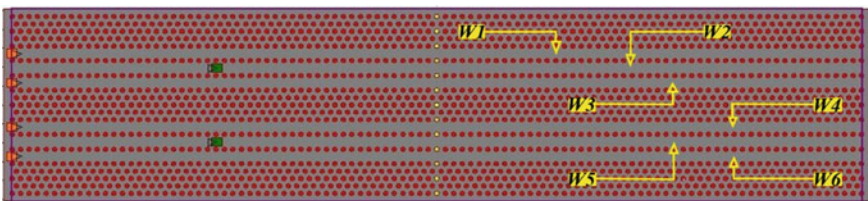


Fig. 8 Schematic structure of constructive and destructive interference while coupling

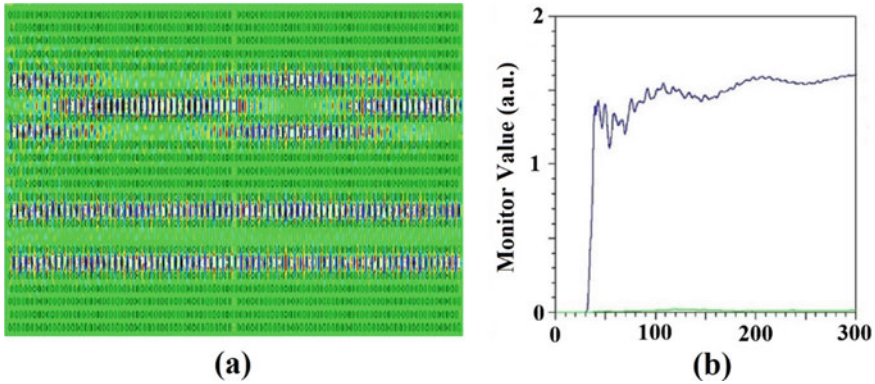


Fig. 9 **a** Electric field propagation through the waveguides when interference phenomenon is present. **b** Output power at W2 and W5

2.5 Effect of Hole Radius on Coupling Length and Power

In the given structure, the radii of the holes (between waveguides) have been varied from $0.20a$ to $0.30a$, where ‘ a ’ is the lattice constant of the structure, i.e., 450 nm . It is observed from Fig. 10 that the radii of the holes is directly proportional to coupling length up to some extent. The signal gets distorted or noisy when the radius is more or less than $0.30a$ and $0.20a$, respectively. Even so, the output power seems to be constant throughout the simulation, but at $0.30a$ it has been 45% and at $0.20a$ it has been 69% . It has been concluded that at $0.25a$ is the optimal radius where the coupling length is 22.95 nm with an output power of 87% .

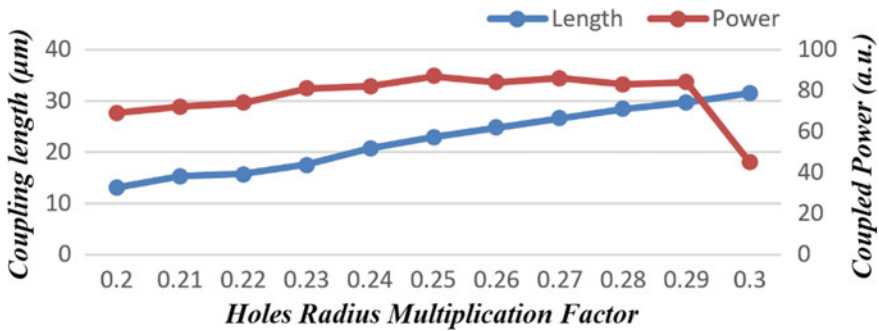


Fig. 10 Comparison between radius of the hole w.r.t coupling length and power

3 Conclusion

In this paper, coupling characteristics of the directional coupler have been investigated. The proposed structure may become a resourceful in optical communication which requires higher throughput and small coupling length. The proposed structure can easily be fabricated and integrated into exiting system. Result shows that device can be implemented with the coupling length is $22.95 \mu\text{m}$ and output power of 87%. Also, the device is able to provide consistent output throughout different case.

References

1. Saleh AAM, Simmons JM (2012) All-Optical Networking—evolution, benefits, challenges, and future vision. *Proc IEEE* 100(5):1105–1117
2. Connelly MJ (2001) Wideband semiconductor optical amplifier steady-state numerical model. *IEEE J Quantum Electron* 37(3):439–447
3. Liu Y, Tangdiongga E, Li Z, Zhang S, Waardt HD, Khoe GD, Dorren HJS (2006) Error-free all-optical wavelength conversion at 160 gb/s using a semiconductor optical amplifier and an optical bandpass filter. *J Lightwave Technol* 24(1):230
4. Dong J, Zhang X, Fu S, Xu J, Shum P, Huang D (2008) Ultrafast All-optical signal processing based on single semiconductor optical amplifier and optical filtering. *IEEE J Sel Top Quantum Electron* 14(3):770–778
5. Sorger VJ, Oulton RF, Ma RM, Zhang X (2012) Toward integrated plasmonic circuits. *MRS Bull* 37(8):728–738
6. Shimizu H, Zayets V (2018) Plasmonic isolator for photonic integrated circuits. *MRS Bull* 43(6):425–429
7. Yadav A, Gerislioglu B, Ahmadvand A, Kaushik A, Cheng GJ, Ouyang Z, Wang Q, Yadav VS, Mishra YK, Wu Y, Liu Y, RamaKrishna S (2021) Controlled self-assembly of plasmon-based photonic nanocrystals for high performance photonic technologies. *Nano Today* 37
8. Prakash C, Mondal HP, Goswami K, Sen M (2008) Investigation of optimum position of interface between strip waveguide and PhC Slot waveguide for maximum power coupling. In: 2018 3rd international conference on microwave and photonics (ICMAP), pp 1–2
9. Goswami K, Mondal H-P, Sen M (2020) Optimized design of 60° bend in optical waveguide for efficient power transfer. In: ICCDN 2019. Lecture notes in electrical engineering, vol 662
10. Goswami K, Mondal H-P, Das P, Thakuria A (2022) Realization of ultra-compact all-optical logic AND gate based on photonic crystal waveguide. In: ICCDN 2021. Lecture notes in electrical engineering, vol 776, pp 61–68
11. Yablonovitch E (1987) Inhibited spontaneous emission in solid-state physics and electronics. *Phys Rev Lett* 58:2059–2062
12. John S (1987) Strong localization of photons in certain disordered dielectric superlattices. *Phys Rev Lett* 58:2486–2489
13. Mondal H-P, Goswami K, Sen M, Prakash C (2018) An all-optical ultra-compact 4-channel wavelength de-multiplexer. In: IEEE international conference on microwave and photonics, Indian School of Mines, Dhanbad, Jharkhand
14. Gogoi D, Das K, Mondal H-P, Talukdar P, Hussain K (2016) Design of ultra-compact 3-channel wavelength de-multiplexer based on photonic crystal. In: IEEE international conference on automatic control and dynamic optimization technique (ICACDOT-2016), I2IT, Pune, Maharashtra
15. Mondal H-P, Sen M, Goswami K (2019) Design, and analysis of a 0.9Tb/s 6-channel WDM filter based on photonic crystal waveguides. *J Opt Soc Am B (OSA)* 36(11):3181–3188

16. Prakash C, Sen M, Mondal H-P, Goswami K (2018) Design and optimization of a TE-pass polarization filter based on a slotted-PhC waveguide. *J Opt Soc Am B* 791–1798
17. Mondal H-P, Sen M, Goswami K (2019) Design and analysis of all-optical 1-to-2 line decoder based on linear photonic crystal. *IET Optoelectron (IET)* 13(4):191–195
18. Mehdizadeh F, Soroosh M, Banaei H-A (2016) A novel proposal for optical decoder switch based on photonic crystal ring resonators. *Opt Quantum Electron* 48
19. Khosravi S, Zavvari M (2018) Design and analysis of integrated all-optical 2×4 decoder based on 2D photonic crystals. *Photonic Netw Commun* 35:122–128
20. Mondal H-P, Chanda S, Gogoi P (2016) Realization of all-optical logic AND gate using dual ring resonator. In: 2016 international conference on automatic control and dynamic optimization techniques (ICACDOT), pp 553–556
21. Mondal H-P, Chanda S, Sen M, Datta T (2015) All optical AND gate based on silicon photonic crystal. In: 2015 international conference on microwave and photonics (ICMAP), pp 1–2
22. Mondal H-P, Sen M, Prakash C, Goswami K, Sarma C-K (2018) Impedance matching theory to design an all-optical AND gate. *IET Optoelectron* 12(5):244–248
23. Ghosh RR, Datta T, Mondal H-P, Sen M (2015) Efficient TE-pass polarizer based on photonic crystal slot waveguide. In: International conference on opto-electronics and photonic materials, SASTRA University, Thanjavur, Tamil Nadu
24. Sahu PP (2012) All-optical switch using optically controlled two mode interference coupler. *Appl Opt* 51(14)
25. Danaie M, Kaatuzian H (2011) Improvement of power coupling in a nonlinear photonic crystal directional coupler switch. *Photonics Nanostruct Fundam Appl* 9(1):70–81
26. <https://www.fiberoptics4sale.com/blogs/archive-posts/95047750-optical-fiber-couplers>. Last accessed on 8 Oct 2021
27. Goswami K, Mondal H-P, Sen M (2021) A review on all-optical logic adder: Heading towards next-generation processor. *Opt Commun* 483
28. Cuesta-Soto F, Martínez A, García J, Ramos F, Sanchis P, Blasco J, Martí J (2004) All-optical switching structure based on a photonic crystal directional coupler. *Opt Express* 12(1):161–167
29. Sharifi H, Hamidi SH, Navi K (2017) All-optical photonic crystal logic gates using nonlinear directional coupler. *Photonics Nanostruct Fundam Appl* 27: 55–63
30. Kuchinsky S, Golyatin VY, Kutikov AY, Pearsall TR, Nedelikovic D (2002) Coupling between photonic crystal waveguides. *IEEE J Quantum Electron* 38(10):1349–1352

All-Optical Feynman Gate Using Frequency Encoding Scheme, Add/Drop Multiplexer and Reflective Semiconductor Optical Amplifier with Simulative Verification



Surajit Bosu and Baibaswata Bhattacharjee

Abstract In the recent emerging research field, the reversible gate is very popular to researchers and they use resources to develop the reversible gate. It yields a huge impact on optical nanotechnology, quantum, and DNA computing. In the all-optical domain, reversible gates like Feynman gate, Fredkin gate, Peres gate, etc., are very demanding. Therefore, a design of frequency encoded Feynman gate using add/drop multiplexer (ADM) and reflective semiconductor optical amplifier (RSOA) is proposed. In long-range propagation, frequency encoding gives benefit in respect to the other encoding techniques. This encoding technique may improve the bit error problems. Due to the high-gain and low-noise property of RSOA, the proposed design performs the operation like computation, data processing, etc., at ultra-high speed. Here, to verify the proposed design, we have used the MATLAB Simulink (R2018a) software.

Keywords Optical communication · Frequency encoding · Reflective semiconductor optical amplifier · Feynman gate

1 Introduction

Signal processing with an electronics system has speed limitations. This limitation can be overcome using all-optical devices where the photon acts as an information transporter. Photon has important properties like high speed, a high degree of parallelism, storage capacity, and low noise. Based on these properties, all-optical devices play a sound role in the field of computation and data processing. In long transmission, encoding of the data signal is needful. Therefore, a number of encoding methods are developed by scientists in optical communication. Among all the encod-

S. Bosu (✉)

Department of Physics, Bankura Sammilani College, Bankura, West Bengal, India
e-mail: surajitbosu7@gmail.com

B. Bhattacharjee

Department of Physics, Ramananda College, Bishnupur, Bankura, West Bengal, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
S. Dhar et al. (eds.), *Advances in Communication, Devices and Networking*, Lecture Notes in Electrical Engineering 902, https://doi.org/10.1007/978-981-19-2004-2_3

ing techniques [1–4], frequency encoding [5–9] has achieved much popularity due to its beneficial consequences and by the virtue of frequency never changes when it goes through some physical phenomenon like absorption, refraction, reflection, etc. In this modern era, low-dissipation loss-based devices like reversible gates are very attractive to the research community. As a result, classical gate like AND, OR, and Ex-OR gate, etc., is inferior with respect to reversible gate. Dissipation of energy is occurred due to bit loss by the classical irreversible gate. Landauer[10] introduced that bit information loss creates heat production in the circuit. Single bit loss can converted into heat energy of $KT\ln(2)$ Joule, where T is the absolute temperature and K is Boltzmann constant. Today's high-speed communication work with a very large number of information data. Therefore, a large number of processors handled these large number of bit operations. As a result, a huge number of bit information is lost. These generate heat in the device. Therefore, scientists have developed different types of reversible gates. After the literature survey, reversible logic gates such as Feynman gate, Fredkin gate, Pauli gate, and Toffoli gate are devised using different types of nonlinear materials [2, 11–16]. We can say that a frequency encoded Feynman gate using ADM and RSOA is not reported to date. So we have proposed a design that gives fast switching time and also have used frequency encoding techniques that decreases the probability of bit error. The remaining of this paper is organized as the following:

Feynman gate is described in Sect. 3. Section 4 describes the switching action of RSOA and the working function of ADM. The operational scheme of the design of the Feynman gate is described in Sect. 5. In Sect. 6, the simulation model with the results of the proposed design is discussed. The results and discussion are in Sect. 7. Last but not the least, the conclusion with future works is given in Sect. 8.

2 Related Work

In this regard, some of the outstanding works are discussed in this section. Ghazali et al. [12] designed an all-optical Feynman gate using optimized SOANOLN. Ghosh et al. [17] implemented a Feynman gate which is based on single-electron threshold logic and simulate the model. Mukherjee et al. [14] proposed the design of the Feynman gate utilizing the counter propagation scheme of RSOA and use soliton pulse to analyze the different parameters. They use the MATLAB software to simulate the model. Kumar et al. [13] have introduced the design of an MZI-based Feynman gate using the electro-optic effect of Lithium niobate. MATLAB software is used for simulation. Maity et al. [16] devised the design of the Feynman gate which is based on terahertz optical asymmetric demultiplexer (TOAD). Biswas et al. [18] proposed the Feynman gate by the analysis of energy dissipation in the field of quantum-dot cellular automata. Khan [19] proposed a method of realization of the Feynman gate by using single-electron transistor. This transistor is a nano-electronic quantum-tunneling device. Bordoloi et al. [20] has introduced an all-optical reversible Feynman gate by using a Mach-Zehnder interferometer. In the next paragraph, we discuss the Feynman gate.

3 Feynman Gate

Feynman gate (FG) [21] has two inputs and two outputs. It is also called controlled-NOT gate. It is a reversible gate which obeys the mapping (A, B) to $(P=A, Q=A \oplus B)$. In FG gate, controlling and controlled inputs are A and B, respectively. The outputs of the FG are represented by P and Q. The block diagram of FG gate is given in Fig. 1, and its truth table is given in Table 1.

4 Operational Principle of ADM and RSOA

This design is based on the working principle of ADM and RSOA. So in this section, we discussed the switching action of RSOA and the working principle of ADM.

In today’s communication wave, division multiplexing (WDM) is very crucial that increases the channel capacity and also increases the data management system. All-optical (A-O) wavelength converter plays a crucial role in the A-O WDM network. In this paper, we have proposed an all-optical frequency encoded Feynman gate where RSOA is used as an A-O wavelength or frequency converter. The operation of RSOA is explained below.

The nonlinear property like cross gain modulation (XGM) is used to create wavelength conversion in RSOA. A probe signal (low intense) and a pump signal (high intense) are injected into the input ports of the SOA, but the output gives probe signal frequency with a higher power. The low-intense probe signal achieves high intensity because of the XGM nonlinear property of SOA. This SOA is called a reflective semiconductor optical amplifier (RSOA) [5, 22–24]. The co-propagating scheme of RSOA’s is given in Fig. 2a. A high-reflective (HR) and an anti-reflective (AR) coating are placed on the opposite faces of RSOA. It shows a high gain. Here, we introduced

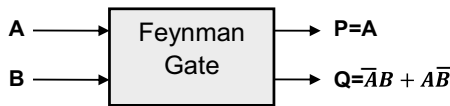


Fig. 1 The block diagram of Feynman gate

Table 1 Feynman gate’s truth table

Input		Output	
A	B	P	Q
0	0	0	0
0	1	0	1
1	0	1	1
1	1	1	0

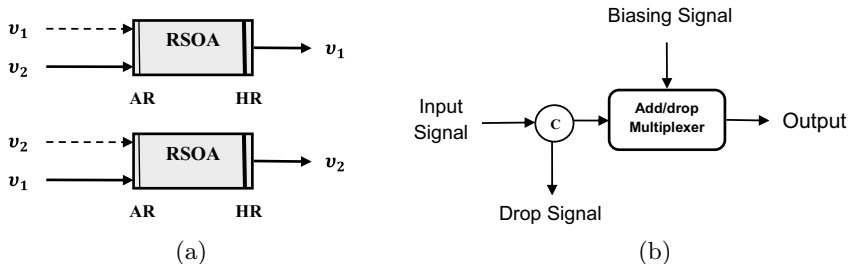


Fig. 2 a RSOA's co-propagating scheme. b Operational diagram of add/drop multiplexer

C-band (1536–1570nm) signals because RSOAs are worked properly in the C-band. 5–20 dBm saturation power may be used to optimize the RSOA [9]. A frequency selector device like ADM is introduced in this work. If we consider the frequency, ν_1 into the input port and frequency, ν_2 into biasing port then ADM selects the input frequency, ν_1 at the output port whereas we cannot get any signal from the drop port. But when we consider the same frequency, ν_2 (or ν_1) into the input and biasing port then ADM reflects the frequency, ν_2 (or ν_1) at the drop port whereas we cannot get any signal at the output[7]. The operational figure of ADM is given in Fig. 2b.

5 The Proposed Scheme of Operation of Frequency Encoded Feynman Gate

The realization of frequency encoded RSOA-based Feynman gate is as follows: Input signals are frequency encoded so two different frequencies imply two different digital logic states. Here, ν_1 and ν_2 are imply “0” and “1” digital logic states, respectively. Feynman gate's input signals “A” and “B” are related to the output signals with Eqs. 1 and 2.

$$P = A \quad (1)$$

$$Q = \bar{A}B + A\bar{B} \quad (2)$$

Case-1: In this case, $A = \nu_1$ and $B = \nu_1$ are injected to the inputs of the proposed Feynman gate. The input signal (A) frequency ν_1 reaches the input port of ADM-1, and this signal reaches also to the biasing port of ADM-2. Similarly, input signal (B) of frequency ν_1 reaches the biasing port of ADM-1, and it reaches also the input of ADM-2. Since ADM-2 and ADM-1 receive the same biasing and input frequencies, then ν_1 is obtained at the drop port of ADM-1 and ADM-2. These ν_1 frequencies act as the pump signal of RSOA-2 and RSOA-1, respectively. Here, a fixed frequency ν_1 works as probe signal of RSOA-2 and RSOA-1. As a result, the outputs of RSOA-

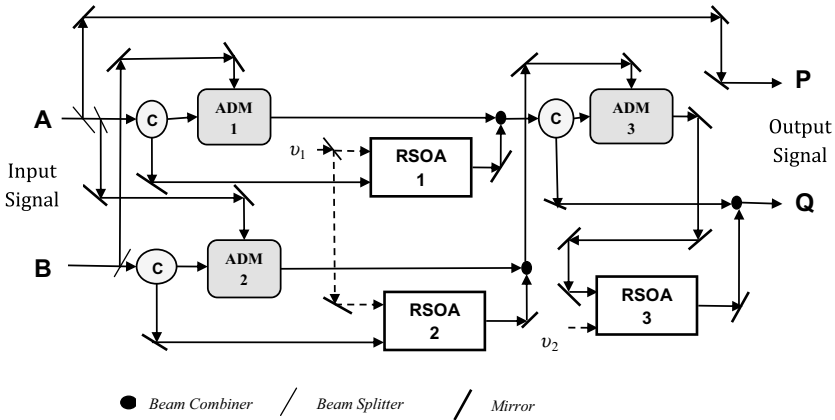


Fig. 3 Operational diagram of devised Feynman gate

Table 2 Proposed Feynman Gate’s truth table in terms of frequency encoding

Input				Output			
A	Digital state	B	Digital state	P	Digital state	Q	Digital state
ν_1	0	ν_1	0	ν_1	0	ν_1	0
ν_1	0	ν_2	1	ν_1	0	ν_2	1
ν_2	1	ν_1	0	ν_2	1	ν_2	1
ν_2	1	ν_2	1	ν_2	1	ν_1	0

1 and RSOA-2 are frequencies ν_1 . These output frequencies (ν_1) are connected to ADM-3 as biasing and input frequency. Therefore, it yields ν_1 at the drop port because the input and biasing frequency are the same. Here, RSOA-3 is not working because of missing probe signal. Finally, at the output (Q), ν_1 is obtained, and at the output (P), ν_1 is obtained because the output (P) is connected with the input (A).

Case-2: In this case, $A = \nu_1$ and $B = \nu_2$ are applied to the input ports of the proposed Feynman gate which are given in Fig. 3. The input signal (A) with frequency ν_1 is injected into the biasing and input port of ADM-2 and ADM-1, respectively. The input signal (B) with frequency ν_2 is also connected with the biasing port of ADM-1 and the input port of ADM-2. Therefore, ν_1 is obtained at the output port of ADM-1, and ν_2 is obtained at the output of ADM-2. In this case, RSOAs (1 and 2) do not work because of missing pump signal. ADM-1’s output, i.e., ν_1 arrives at the input port of the ADM-3. The output of ADM-2, i.e., ν_2 reaches the biasing port of ADM-3. As a result, the ADM-3 yields the frequency ν_1 at the output. ADM-3’s output acts as a pump signal of RSOA-3, and ν_2 acts as a probe signal. Finally, RSOA-3 yielded ν_2 at the output (Q) and at the output (P), ν_1 is obtained because output (P) is directly connected with the input (A) (Fig. ??).

Case-3: In this case, $A = \nu_2$ and $B = \nu_1$ are applied to the input ports of this design. The input signal (A) with frequency ν_2 is applied to the biasing port of ADM-2 and input of ADM-1. The input signal (B) with frequency ν_1 is also connected with the biasing port of ADM-1 and input port of ADM-2. Therefore, ν_2 is obtained at the output port of ADM-1, and ν_1 is obtained at the output of ADM-2. In this case, both the RSOAs (1 and 2) do not work because of missing pump signal. ADM-1's output, i.e., ν_2 reaches the input port of the ADM-3, and the output from ADM-2, i.e., ν_1 reaches the biasing port of ADM-3. As a result, the frequency ν_2 is obtained at the output port of ADM-3. The ADM-3's output acts as a pump signal of RSOA-3, and ν_2 acts as probe signal. Finally, RSOA-3's output, i.e., ν_2 is obtained at the output (Q), and at the output (P), ν_1 is obtained because output (P) is directly connected with the input (A).

Case-4: In this case, $A = \nu_2$ and $B = \nu_2$ are injected into the inputs of the proposed Feynman gate. The input signal (A) of frequency ν_2 reaches the input port of ADM-1 and reaches also the biasing port of ADM-2. Similarly, the input signal (B) of frequency ν_2 reaches the biasing port of ADM-1 and reaches also the input port of ADM-2. Since ADM's (1 and 2) input and biasing frequencies are the same so it is drop port can reflect ν_2 . These ν_2 frequencies act as the pump signal of RSOA-2 and RSOA-1, respectively. Here, a fixed frequency ν_1 works as probe signal of RSOA-2 and RSOA-1. So the outputs of RSOA's (1 and 2) outputs are the frequency ν_1 . These output frequencies ν_1 are connected as biasing and the input frequency of ADM-3. Therefore, a frequency ν_1 is obtained at the drop port of ADM-3 because biasing and input frequencies are the same. Here, RSOA-3 is not working due to the missing of probe signal. So, finally at the output (P), ν_2 is obtained because output (P) is directly connected with the input (A), but at the output (Q), ν_1 is obtained.

6 Simulation Results of Proposed Feynman Gate

In the previous section, the operational scheme of the proposed design is explained theoretically. Now, we discuss the simulation model of the proposed design. Using MATLAB Simulink (R2018a) software, the proposed design is simulated. RSOAs and ADMs are programmed based on their characteristics using MATLAB language. If frequencies, $\nu_1 = 193.5$ THz (wavelength = 1550 nm) and $\nu_2 = 194.1$ THz (wavelength = 1545 nm) are considered as the probe signal and pump signal, then 193.5 THz is obtained at the output port, whereas 194.1 THz is obtained at the output when $\nu_1 = 193.5$ THz and $\nu_2 = 194.1$ THz are considered as the pump and probe signal. If $\nu_1 = 193.5$ THz and $\nu_2 = 194.1$ THz are considered as the input and biasing signal frequencies, then the input frequency 193.5 THz is passed through the ADM, and it is obtained at the output. When biasing and the input signal frequencies are matched, then input signal is passed through the ADM, and it is obtained at the drop port. Simulation of the proposed design is based on these above considerations.

Case-1: Here, simulation is done by the 193.5 THz frequencies at both the inputs, then 193.5 THz is obtained at the outputs P and Q.

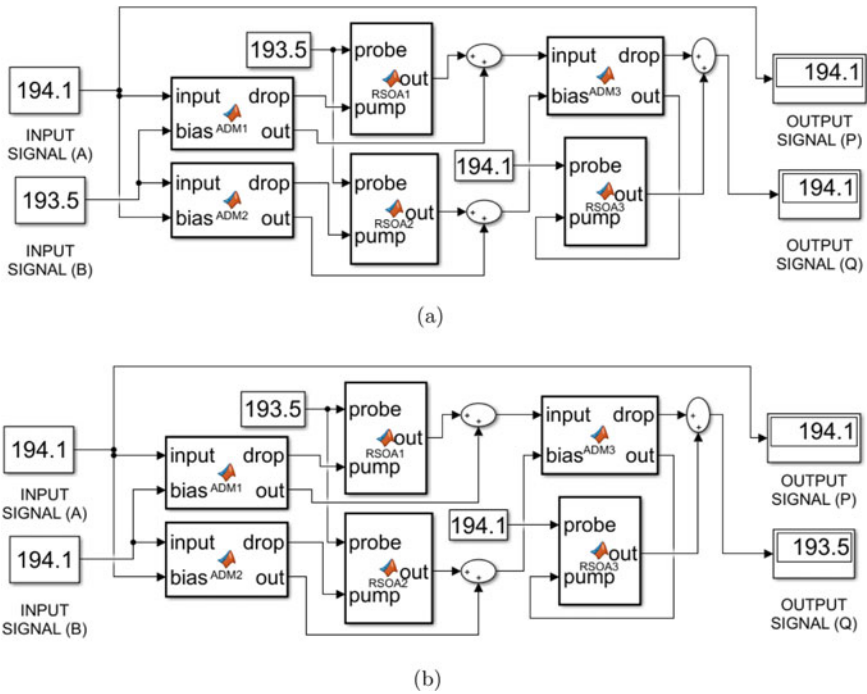


Fig. 4 **a** Simulation model for the input $A=194.1$ THz and $B=193.5$ THz that indicate the logic states $A=1$ and $B=0$. **b** Simulation model for the input $A=194.1$ THz and $B=194.1$ THz that indicate the logic states $A=1$ and $B=1$.

Case-2: In this case, the 193.5 THz and 194.1 THz frequencies are injected into the A and B, respectively. 193.5 THz and 194.1 THz are yielded at the outputs P and Q, respectively.

Case-3: In this case, the 194.1 THz and 193.5 THz frequencies are injected into the A and B, respectively. 194.1 THz and 194.1 THz are obtained at the outputs P and Q, respectively. This simulation model is shown in Fig. 4a.

Case-4: Here, the devised model is simulated with the 194.1 THz frequencies at both the inputs, then 194.1 THz and 193.5 THz are obtained at the outputs, P and Q, respectively. This simulation model is shown in Fig. 4b.

7 Results and Discussion

In this section, the theoretical interpretation and the simulation results are discussed. Input signals are introduced in the design with a time interval of 50 ps. These input and output signal patterns are shown in Figs. 5a, b, and also the simulation models are given in Figs. 4a, b. When this design is simulated with input waveforms, then output

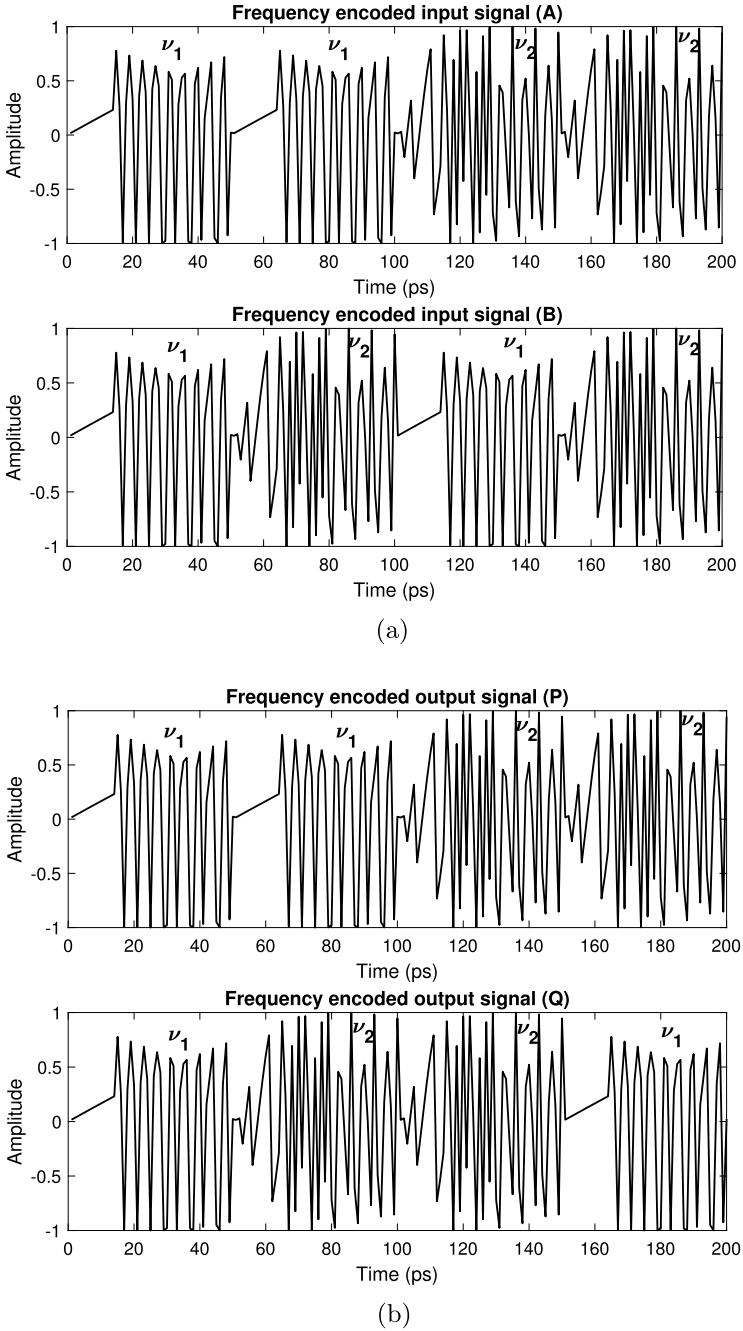


Fig. 5 **a** Input signal waveforms for inputs *A* and *B*. **b** Simulated output waveforms for outputs *P* and *Q*

signal waveforms are yielded. After verification of the simulation results, graphical results, and truth table (Tables 1 and 2), it may be concluded that the Feynman gate's truth table is perfectly matched with the simulated output results. In the first 50 ps, we introduced the input signals A and B with frequency ν_1 , which imply $A=0$ and $B=0$. So after the completion of simulation frequencies, ν_1 are obtained at the output terminal " P " and " Q " that indicates $P=0$ and $Q=0$. In the next 50 ps, we introduced the input signals A and B with frequencies ν_1 and ν_2 which imply $A=0$ and $B=1$. So after the completion of simulation frequencies, ν_1 and ν_2 are obtained at the output terminal " P " and " Q " that indicate $P=0$ and $Q=1$. In the next 50 ps, we introduced the input signals A and B with frequencies ν_2 and ν_1 which imply $A=1$ and $B=0$. So after the completion of simulation frequencies, ν_2 is obtained at the output terminal " P " and " Q " that indicates $P=1$ and $Q=1$. In last 50 ps, we introduced the input signals A and B with frequencies ν_2 which imply $A=1$ and $B=1$. So after the completion of simulation frequencies, ν_2 and ν_1 are obtained at the output terminal " P " and " Q " that indicates $P=1$ and $Q=0$. These results are matched with the graphical results that are given in Fig. 5a, b.

8 Conclusion

In this article, a frequency encoded Feynman gate using ADM and RSOA is designed. For the verification of the truth table, MATLAB Simulink (R2018a) software is used. Using the benefits of frequency encoding, this design may decrease the bit error problems. RSOA operates at low power as well as ultra-high speed so our proposed design may also obey these properties. The author's future intention is to design the multiplexer, demultiplexer, adder, subtractor using the proposed design.

Acknowledgements We are thankful to the Department of Physics, Bankura University, Bankura, Pin-722155, West Bengal, India for support to conduct this research work.

References

1. Nishchal NK (2019) Optical cryptosystems. IOP Publishing, UK
2. Dey S, De P, Mukhopadhyay S (2019) An all-optical implementation of Fredkin gate using Kerr effect. *Optoelectro Lett* 15(4):317–320
3. Dey S, Mukhopadhyay S (2019) All-optical integrated square root of Pauli-Z (SRZ) gates using polarization and phase encoding. *J Opt* 48(4):520–526
4. Raja A, Mukherjee K, Roy JN (2021) Analysis of new all optical polarization-encoded dual SOA-based ternary NOT & XOR gate with simulation. *Photonic Netw Commun* 41(3):242–251
5. Saha S, Mukhopadhyay S et al (2020) All optical frequency encoded quaternary memory unit using symmetric configuration of MZI-SOA. *Opt Laser Technol* 131:106386
6. Bosu S, Bhattacharjee B (2022) All-optical frequency encoded dibit-based parity generator using reflective semiconductor optical amplifier with simulative verification. *FACTA Universitatis. Series: Electronics and Energetics* 35(1):029–41

7. Bosu S, Bhattacharjee B (2021) A novel design of frequency encoded multiplexer and demultiplexer systems using reflected semiconductor optical amplifier with simulative verification. *J Opt* 50(3):361–370
8. Mukherjee Kousik (2021) Design and analysis of all optical frequency encoded X-OR and X-NOR gate using quantum dot semiconductor optical amplifier-Mach Zehnder interferometer. *Opt Laser Technol* 140:107043
9. Bosu S, Bhattacharjee B (2021) A design of frequency encoded dibit-based comparator using reflective semiconductor optical amplifier with simulative verification. In: 2021 devices for integrated circuit (DevIC). IEEE, pp 175–179
10. Landauer Rolf (1961) Irreversibility and heat generation in the computing process. *IBM J Res Dev* 5(3):183–191
11. Garai SK (2014) A novel method of developing all optical frequency encoded Fredkin gates. *Optics Commun* 313:441–447
12. Ghazali NF, Wahid MHA, Ahmad Hambali NAM, Juhari N, Shahimin MM (2018) Characterization of all-optical Feynman and Fredkin gates utilizing optimized Soanolm. In: AIP conference proceedings, vol 2045. AIP Publishing LLC, pp 020079
13. Kumar S, Raghuvanshi SK et al (2016) Design of optical reversible logic gates using electro-optic effect of lithium niobate based Mach–Zehnder interferometers. *Appl Opt* 55(21):5693–5701
14. Mukherjee K, Maji K, Raja A (2019) All-optical feynman gate using reflective semiconductor optical amplifiers and binary to gray code converter. *Adv Appl Math Sci* 19(9):919–992
15. Mandal D, Mandal S, Mandal MK, Garai SK (2018) Alternative approach of developing optical binary adder using reversible peres gates. *Int J Opt* 2018
16. Maity GK, Maity SP, Roy JN (2012) Toad-based Feynman and Toffoli gate. In: 2012 second international conference on advanced computing & communication technologies. IEEE, pp 343–349
17. Ghosh A, Jain A, Singh NB, Sarkar SK (2016) Single electron threshold logic based feynman gate implementation. In: 2016 second international conference on research in computational intelligence and communication networks (ICRCICN). IEEE, pp 266–268
18. Biswas PK, Bahar AN, Habib MA, Abdullah-Al-Shafi M (2017) Efficient design of Feynman and Toffoli gate in quantum dot cellular automata (QCA) with energy dissipation analysis. *Nanosci Nanotechnol* 7(2):27–33
19. Khan MHA (2015) Single-electron transistor based implementation of not, Feynman, and Toffoli gates. In: 2015 IEEE international symposium on multiple-valued logic. IEEE, pp 66–71
20. Bordoloi K, Theresal T, Prince S (2014) Design of all optical reversible logic gates. In: 2014 international conference on communication and signal processing. IEEE, pp 1583–1588
21. Feynman RP (1985) Quantum mechanical computers. *Opt News* 11(2):11–20
22. Kotb Amer, Guo Chunlei (2020) Reflective semiconductor optical amplifiers-based all-optical NOR and XNOR logic gates at 120 Gb/s. *J Mod Opt* 67(18):1424–1435
23. Kotb A, Zoiros KE, Guo C (2018) Performance investigation of 120 Gb/s all-optical logic XOR gate using dual-reflective semiconductor optical amplifier-based scheme. *J Comput Electron* 17(4):1640–1649
24. Maji K, Mukherjee K, Raja A, Roy JN (2020) Numerical simulations of an all-optical parity generator and checker utilizing a reflective semiconductor optical amplifier at 200 GPBS. *J Comput Electron* 1–15

Effects of Dimensional Variations on Short Channel Parameters in 14 nm Channel Length TG–SOI FinFETs



Priyanka Saha , Swagat Nanda , Potaraju Yugender, and Rudra Sankar Dhar 

Abstract In the current semiconductor era, MOSFET gradually loses its performance to the short channel effect (SCE) due to decreasing channel length. Advanced FinFET technology introduces reduction in the SCEs. A tri-gate FinFET, where the range of the gate is 14 nm, is accomplished and presented here using Silvaco TCAD. The geometrical dimensions of the device are characterized based on the width and the height of the fin to attain an aspect ratio of 0.5, 1, and 2. It can be reformed in two ways where firstly width is constant with varying heights to get aspect ratio and vice versa. The electrical parameters such as I_{on} current, threshold voltage, sub threshold current, drain-induced barrier lowering, and I_{on}/I_{off} current ratio are compared for the different aspect ratios to attain an optimized aspect ratio and also from the I_D-V_{GS} graph, the threshold voltage is determined with respect to V_{DS} . The logarithmic I_D-V_{GS} graph depicts the leakage current. Thereby, the development of the TG FinFET structure is proposed, where the standardized schematic is at fixed height, width, and length of 10 nm.

Keywords Aspect ratio (AR) · DIBL · Silvaco TCAD · Subthreshold swing · Tri-gate NanoFETs

1 Introduction

The semiconductor industries are facing serious problems as research goes ahead in manufacturing fast and compact devices. Decreasing the channel length day by day in MOSFET fabrication. The most important demands are the limiting control over the channel by gate terminal which in turn increases the short channel effects (SCE) such as hot electron effect, impact ionization, DIBL [1], and leakage currents [2]. In 1965, Moore predicated that the numbers of transistors per square inch will double every two years [3] which remains true. Due to shrinking device dimensions, more number of gates was added to increase device performance [4]. In California at the

P. Saha · S. Nanda · P. Yugender · R. S. Dhar (✉)

Department of Electronics and Communication Engineering, National Institute of Technology Mizoram, Aizawl, Mizoram, India
e-mail: rdhar@uwaterloo.ca

University of Berkley, the first fin field effect transistor was developed by Chenming Hu and his research team [5] which is a type of multi-gate MOSFET. In this FinFET, it integrates one or more gate into a single device. When we view the structure it looks like a set of fins so the name is given as FinFET and the thickness of the device will be the length of the channel [6]. This FinFET technology is the most commonly used technology in integrated circuits as these 3D fins provide more density for the same size.

The features of these FinFETs makes a better replacement for bulk MOS at nanoscale and as the fabrication technology is almost the same with conventional MOS transistor [7], the FinFETs substitute the conventional MOSFETs. The short channel effects are reduced as the two gates are electrically coupled with the thin silicon body. These FinFETs provide higher noise tolerance and low power [8]. The 14 nm FinFETs are the MOSFET technology which are the replacement of 22 nm.

In this paper, the inspiration is mainly the study of the effect of geometrical variations in the fin structure of a 14 nm gate length TG SOI FinFET device. The dimensions are varied by first keeping a constant width of the fin and varying the height to achieve an aspect ratio of 0.5, 1, and 2. Then the height is kept constant and the width of the fin is varied to achieve an aspect ratio of 0.5, 1, and 2. The electrical traits of the different structures are deliberated.

2 Device Implementation and Principle

The TG FinFET is designed with a silicon only channel structure of 14 nm length and is schematically shown in Fig. 1. The dimensional parameters of the structure, which are characterized by the height (H_{FIN}), width (W_{FIN}) of the fin, and gate oxide thickness are listed in Table 1. The initial dimensions of the device are fixed at 10 nm height, width, and length, having a compact area, a square device, and an aspect ratio of 1. The design aspect ratio, which is the H_{FIN}/W_{FIN} ratio, is then varied from 0.5 to 2. The aspect ratio of the device can be modified in two ways. Since the aspect ratio is prescribed as the ratio of H_{Fin} to W_{FIN} , either height can be changed keeping the width constant or the width can be varied keeping a constant height. The structure is displayed using similar doping levels of the various regions of the device.

The 3D TG FinFET device with electrical parameters is reproduced by using SILVACO TCAD Atlas tools [9]. For better modeling, some models like fixed lifetimes model with Shockley-Read-Hall (SRH), band gap narrowing and auger recombination are comprised in this device simulation with 14 nm gate length FinFET.

The length and width of the source as well as drain regions range are adjusted at 10 nm along with the 14 nm gate range, while the widths and heights are different to change the height/width ratio keeping the gate oxide at 2 nm of SiO_2 .

Fig. 1 Schematic representation of the SOI NanoFET

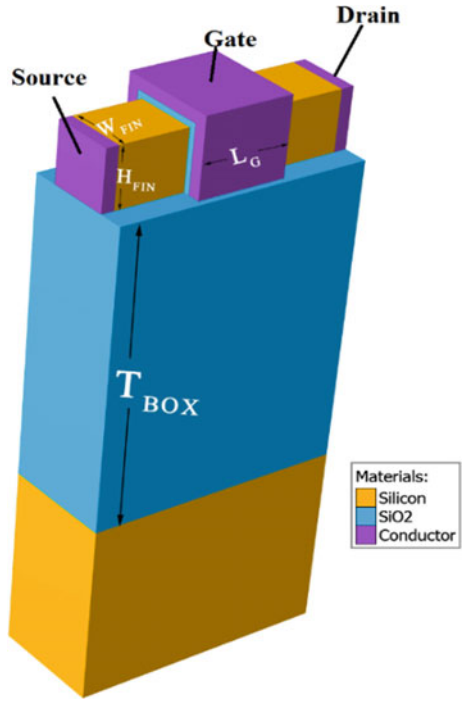


Table 1 TG FinFET feature chart of 14 nm device

Category	Measurement
Drain/source length	10 nm
Channel range (L_G)	14 nm
Thickness of lateral oxide (SiO_2)	2 nm
Silicon fin thickness (W_{FIN})	5, 10, 20 nm
Silicon fin height (H_{FIN})	5, 10, 20 nm
Channel doping (N_A)	10^{15} cm^{-3}
Drain/source doping (N_D)	10^{18} cm^{-3}

3 Results with Outcomes

Silvaco TCAD is used for device design. The standard reproduction of the structure is first edited at a fixed height, width, and length of 10 nm. The dimensions are then varied by first keeping the length and height at 10 nm and varying the width as 5 nm, 10 nm, and 20 nm changing the aspect ratio as 2, 1, and 0.5, respectively. The dimensions are further varied by keeping the length and width at 10 nm and varying the height as 5 nm, 10 nm, and 20 nm. This modifies the aspect ratio as 0.5, 1, and 2, respectively.

The graph of the drain current—gate voltage characteristics in linear scale is displayed in Fig. 2. The threshold voltages and the maximum I_{on} current is settled from the linear transfer characteristics graph. The changes of the threshold voltage is determined for the various aspect ratios are shown in Fig. 3. When the height was fixed at 10 nm, and the width was varied as 5 nm, 10 nm, and 20 nm, the V_{TH} of the devices are set up with 0.22 V, 0.2 V, and 0.17 V gradually where the V_{DS} is 0.1 V. Similarly, when the width was constant at 10 nm, and the height was varied as 5 nm, 10 nm, and 20 nm, the threshold voltage of the devices are ascertained as

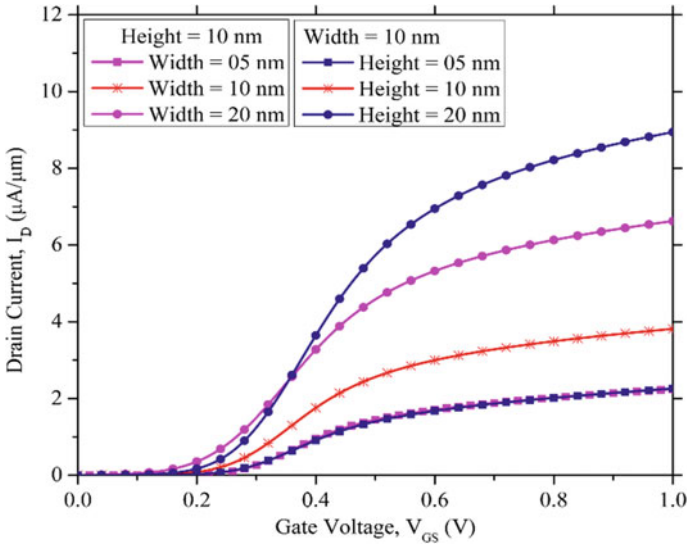


Fig. 2 Drain current—gate voltage graph for aspect ratios of 0.5, 1, and 2

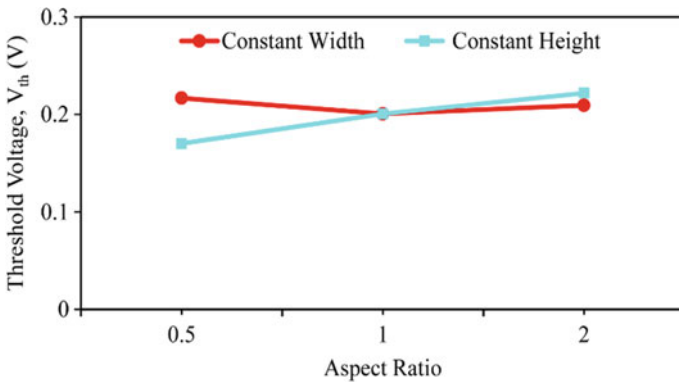


Fig. 3 Changes of V_{th} for several aspect ratios

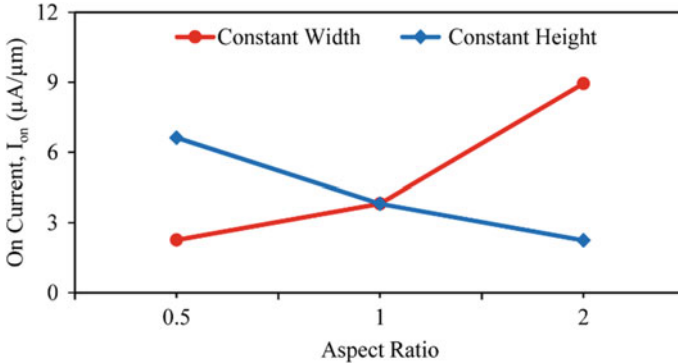


Fig. 4 Changes of ON current for several aspect ratios

0.216 V, 0.2 V, and 0.209 V, respectively, at $V_{DS} = 0.1$ V. The threshold voltages for this change is plotted in Fig. 3.

When V_{DS} is 0.1 V, the I_{on} current of a SOI device, with a fixed width of 10 nm and aspect ratios of 0.5, 1, and 2 are 2.26 $\mu A/\mu m$, 2.82 $\mu A/\mu m$, and 8.94 $\mu A/\mu m$, respectively, while the I_{on} is determined to be 6.62 $\mu A/\mu m$, 3.82 $\mu A/\mu m$, and 2.25 $\mu A/\mu m$ for the SOI device with a fixed width of 10 nm and aspect ratios of 0.5, 1, and 2. Thus, the device with constant width shows higher I_{on} current with increase in aspect ratio while the I_{on} current shows a decreasing trend in the device with constant height and increasing aspect ratios as shown in Fig. 4.

The drain current—gate voltage transfer characteristics, in logarithmic plot, as shown in Fig. 5, illustrates the leakage current, the subthreshold swing, and DIBL. The I_{off} can also be fixed by Eq. (1) [10] as:

$$I_{off}(\text{nA}) = 100 \frac{W}{L} 10^{-\frac{V_{TH}}{SS}} \tag{1}$$

where W = width, L = length, V_{TH} = device threshold voltage, and SS = subthreshold swing of the respective device.

The I_{off} shows a wide variation from 29 pA/ μm with aspect ratio of 2 and 5 nA/ μm for aspect ratio of 0.5 for the device with constant height. For the device where the width is kept constant, the I_{off} currents are 62 pA/ μm for an aspect ratio of 0.5 and increases to 0.8 nA/ μm where the aspect ratio is also incremented to 2, as shown in Fig. 6. This directly affects the I_{on}/I_{off} current ratio. A higher ratio of the ON to OFF current ratio leads to faster switching of the device. When the aspect ratio is 0.5, the I_{on}/I_{off} current ratio for constant width device is 3.6×10^4 and constant height device is 1.3×10^3 . When the aspect ratio is increased to 2, the I_{on}/I_{off} current ratio for constant width device is 1.1×10^4 and constant height device is 7.8×10^4 . This shows a higher degree of change in the I_{on}/I_{off} ratio, but for an aspect ratio of 1, this ratio is 1.2×10^4 , which is an acceptable trade off as shown in Fig. 7.

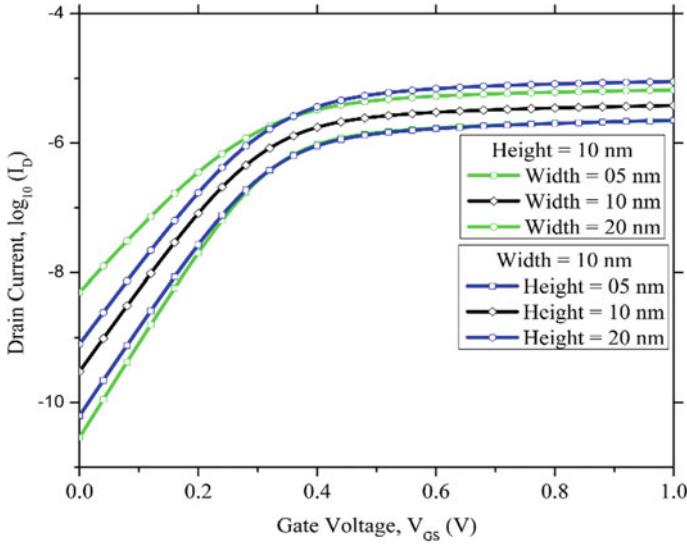


Fig. 5 Logarithmic plot for drain current— gate voltage features for several aspect ratios at $V_{DS} = 0.1$ V

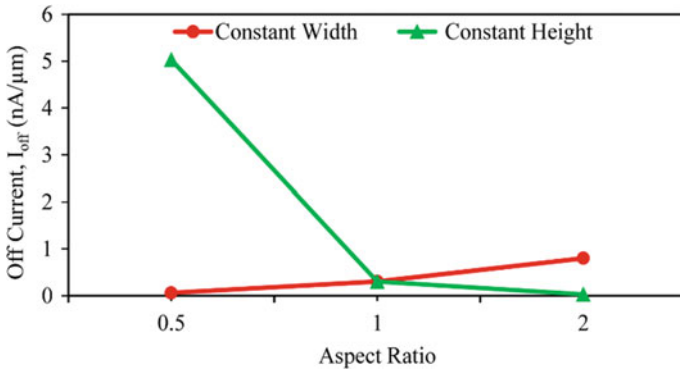


Fig. 6 Changes for I_{off} at several aspect ratios

The subthreshold swing for SOI device for an aspect ratio of 1 is seen to be around 78 mV/decade. However, the SS fluctuates from 73 to 82 mV/decade, when the AR is changed from 0.5 to 2, respectively, for a constant width device and decreases from 100 to 68 mV/decade for constant height device, as depicted from Fig. 8. This shows opposite trends for SS with increasing aspect ratios for constant height and constant width devices. Hence, it can be inferred that the device with an AR of 1 is the ideal device for consideration.

The DIBL is referred as the modification in threshold voltage for a large drain voltage variation. The DIBL can be equated using Eq. (2) as [11]:

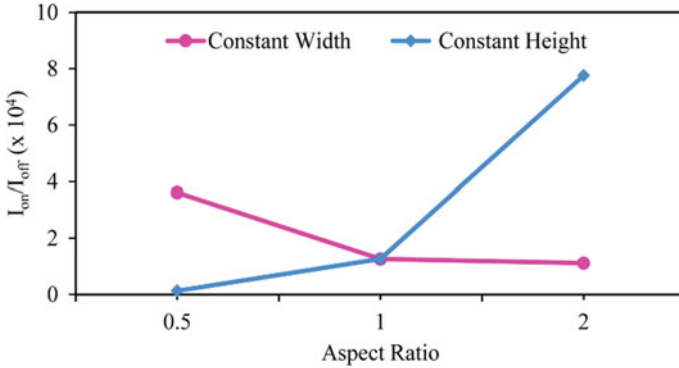


Fig. 7 Disparity of I_{on}/I_{off} ratio at several aspect ratios

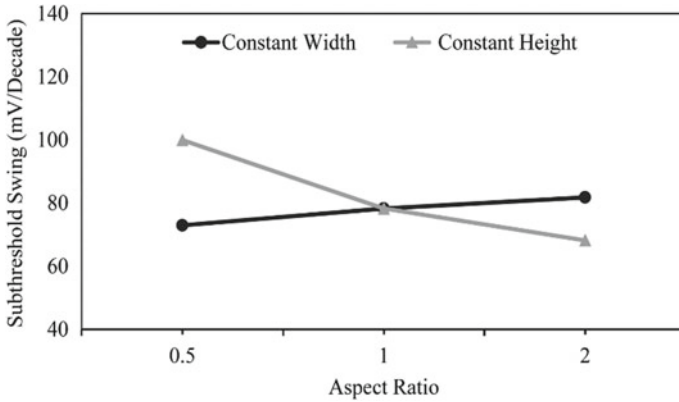


Fig. 8 SS parameter comparison at several aspect ratios

$$DIBL \left(\frac{mV}{V} \right) = \frac{\nabla V_{TH}}{\nabla V_{DS}} \tag{2}$$

where ∇V_{TH} =threshold voltage change and ∇V_{DS} =drain to source voltage change. And it should be minimum. The comparison of DIBL shows a higher DIBL of 168 mV/V for the device with constant height and an aspect ratio of 0.5. The DIBL decreases to 97 mV/V for aspect ratio of 1 and 45 mV/V for aspect ratio of 2. This trend becomes opposite for the device with constant height with the DIBL being 71 mV/V for aspect ratio of 0.5 and increases to 97 mV/V and 125 mV/V for ratio of 1 and 2, respectively, as depicted in Fig. 9. This shows that the aspect ratio of 1 is the better trade off with respect to DIBL.

Figure 10 shows the output characteristics plot for TG n-FinFET devices with various heights and widths, for a V_{GS} of 0.4 V. The graph shows the comparison of two different device dimensions. When the height is kept constant at 10 nm and the

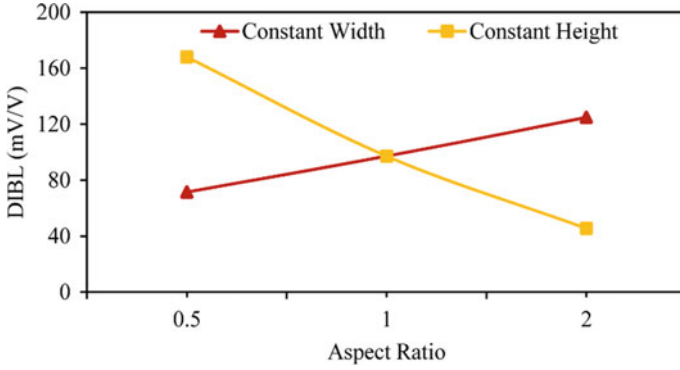


Fig. 9 DIBL at various calculated aspect ratios

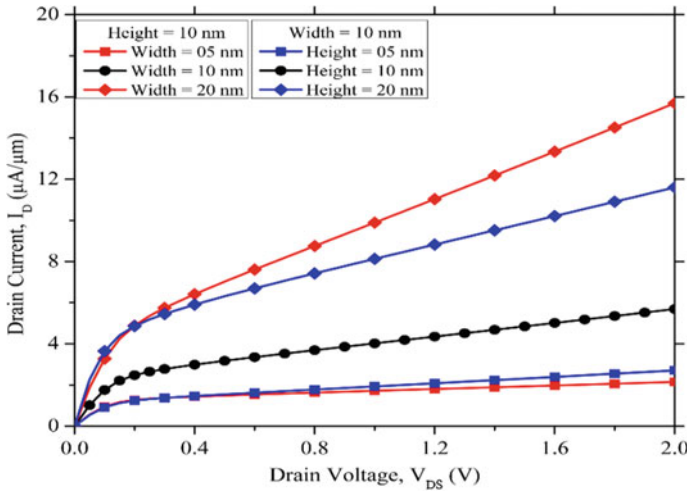


Fig. 10 Drain current versus drain to source voltage plot for several aspect ratios with $V_{GS} = 0.4$ V

widths are varied from 5 to 20 nm, the maximum drain current is observed for a width of 20 nm, thus for an aspect ratio of 0.5. Similarly, when the width is kept constant at 10 nm and the heights are varied from 5 to 20 nm, the maximum drain current is observed for a height of 20 nm, thus for an aspect ratio of 2. So it is inferred that the device with a height and width of 10 nm, having an aspect ratio of 1, portrays satisfactory performance, when compared to other electrical characteristics and can be device of choice among all the considered devices.

4 Conclusion

This article describes the upliftment of a FinFET which is n-channelled Tri-gate where the gate range is 14 nm by Silvaco TCAD. Initially, the feature is advanced with the source/drain length, width, and height maintained at 10 nm, to keep an aspect ratio of 1 for characterizing the performance at small device area. The SiO₂ is the oxide layer of gate which is 2 nm in the devices. The height and width of the device are then varied to obtain an aspect ratio of 0.5 and 2. The electrical characteristics and the short channel effects of the device with different aspect ratios are then compared. For the device, where the aspect ratio is changed by keeping a fixed height of the fin, it is observed that although the short channel effects improve by increasing the aspect ratio, the I_{on} current reduces threefold, which is not acceptable. For the device, where the aspect ratio is changed by keeping a fixed width of the fin, it is observed that increasing the aspect ratio from 0.5 to 2 increases the I_{on} current fourfold but decreases the I_{on}/I_{off} current ratio by three times and also degrades the SS and DIBL which is also not acceptable. Thus a trade-off is achieved by maintaining an aspect ratio of 1 which provides acceptable range of I_{on} , I_{off} , and other SCEs.

Acknowledgements The authors are indebted to Dept. of Electronics and Communication Engineering, National Institute of Technology Mizoram for providing flexible platform, facilities and support to grow skill for carry out this work.

References

1. Pei G, Kedzierski J, Oldiges P, Jeong M, Kan EC (2002) FinFET design considerations based on 3-D simulation and analytical modeling. *IEEE Trans. on Electron Devices* 49(8):1411–1419
2. Roy K, Mukhopadhyay S, Mahmoodi-Meimand H (2003) Leakage current mechanisms and leakage reduction techniques in deep submicrometer CMOS circuits. *Proc IEEE* 91(2):305–327
3. Schaller RR (1997) Moore's law: past, present and future. *IEEE Spectr* 34(6):52–59
4. Colinge JP (2008) *FinFETs and Other Multi Gate Transistors*. Springer, New York, USA
5. Hisamoto D, Lee WC, Kedzierski J, Takeuchi H, Asano K, Kuo C, Anderson E, King TJ, Bokor J, Hu C (2000) FinFET—a self-aligned double-gate MOSFET scalable to 20 nm. *IEEE Trans Electron Devices* 47(12):2320–2325
6. Kajal, Sharma VK (2020) FinFET: a beginning of non-planar transistor era. *Nanoscale VLSI. Energy systems in EE*. Springer, Singapore
7. Bhattacharya D, Jha NK (2014) FinFETs: from devices to architectures. *Adv Electron* 365689:21
8. Vora PH, Lad R (2017) A review paper on CMOS, SOI and FinFET technology. *Des Reuse Ind Articles* 1–10. <https://www.design-reuse.com/articles/41330/cmos-soi-finfet-technology-review-paper.html>
9. S. International (2016) Atlas user's manual device simulation software. Silvaco Int., Santa Clara
10. Mohapatra SK, Pradhan KP, Sahu PK (2013) Some device design considerations to enhance the performance of DG-MOSFETs. *Trans Electr Electron Mat* 14(6):291–294
11. Tosaka Y, Suzuki K, Sugii T (1999) Scaling-parameter-dependent model for subthreshold swing S in double-gate SOI MOSFET's. *IEEE Electron Dev Lett* 15(11):466–468

Dibit-Based 4-Bit Parity Generator Using Reflective Semiconductor Optical Amplifier and Frequency Encoding Scheme



Surajit Bosu and Baibaswata Bhattacharjee

Abstract Future generation communication and computation system are dealing with ultra-high-speed data transfer, systematic management as well as lower cost. So the researchers are motivated to implement the different logic gates, digital devices, combinational circuits, optical computers. In optical computing systems, the parity generator is familiar as an essential part. In the case of all-optical devices design, reflective semiconductor optical amplifiers (RSOAs) appear as a promising candidate. Due to the versatile gain medium of RSOA, it has various important applications in passive optical networks. In comparison with semiconductor optical amplifiers (SOAs), RSOAs exhibit better gain performance because of their double pass property. Therefore, a frequency encoded dibit-based 4-bit parity generator is designed using low-cost RSOA. It shows also better switching properties. This design operates at a high speed. This device is benefited with dibit-based logic as reduces bit error problem. MATLAB Simulink (R2018a) software is used to verify the proposed design.

Keywords Dibit-based logic system · Reflective semiconductor optical amplifier · Add/drop multiplexer · Parity generator · Frequency encoding

1 Introduction

Nowadays, optics-based devices add some significant values in the area of high-speed communication. So the photon becomes more popular for information transmission. Photons can carry information at an ultra-fast speed rather than electron, so it is more attractive to the researchers rather than the electron. The data signal can be transmitted in long-range using different types of encoding techniques [1, 2]. Among

S. Bosu (✉)

Department of Physics, Bankura Sammilani College, Bankura, West Bengal, India
e-mail: surajitbosu7@gmail.com

B. Bhattacharjee

Department of Physics, Ramananda College, Bishnupur, Bankura, West Bengal, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
S. Dhar et al. (eds.), *Advances in Communication, Devices and Networking*, Lecture Notes in Electrical Engineering 902, https://doi.org/10.1007/978-981-19-2004-2_5

all of them, frequency encoding [3–5] is more reliable in signal propagation. In optical communication, optical devices like the adder, subtractor, comparator, parity generator are essential for arithmetic, logic units, decision-making circuits, and memory units. These are the basic building blocks of optical data processors. Here, digital logic states “0” and “1” are represented by two different frequencies ν_1 and ν_2 , respectively. In communication and data storage systems, the parity checker and generator are very essential devices. In this communication, an all-optical frequency encoded even and odd parity generator using RSOA are proposed, and also we have used add/drop multiplexer (ADM) for frequency routing. This devised design is based on dibit logic system. Therefore, an alternative approach is taken to design this parity generator. Our proposed design generates simultaneously even and odd parity without using an extra device and extra control terminal. Therefore, frequency encoding and dibit-based systems reduce the bit error problems. Since RSOA is an ultra-fast switching device with a low noise [17]. Since RSOA is the key component of our devised design so this design also can operate at high speed and with low noise. The organization of the paper is as follows: Related works are briefly described in Sect. 2. The function of ADM and RSOA is described in Sect. 3. Section 4 describes the operation scheme of the proposed parity operator. The simulation experiment of the proposed model is explained in Sect. 5. The results and discussion are given in Sect. 6. Finally, the conclusion is given in Sect. 7.

2 Related Work

In last few years, scientists have progressed the research work for parity generators and parity checkers. Some of these legendary work are discussed below. All-optical parity generator and checker based on micro-ring-resonator (MRR) have been reported by Rakshit et al. [6], and their model is also verified using numerical simulation. Wang et al. [7] have reported all-optical parity checker and the works implemented in the nanoscale-integrated chip. Nair et al. [8] have introduced a 3-bit all-optical parity generator and checker using SOA-MZI-based tree architecture. Bhattacharyya et al. [9] have reported a 4-bit parity generator using an SOA-assisted Sagnac switch, and the design is verified through numerical simulation. Chowdhury et al. [10] have introduced a design of 4-bit parity generator and checker using nonlinear material-based switches. Dimitriadou et al. [11] have proposed a 4-bit parity generator, checker using quantum-Dot-SOA-based MZI switch and verified through numerical simulation. Ghosh [12] has reported a parity generator and parity checker in a modified trinary number system. Here, the key components of the device are the Savart plate and spatial light modulator. Kumar et al. [13] have reported MZI-based parity checker utilizing the electro-optic effect. Their work are analyzed mathematically and with the help of MATLAB and optiBPM software their design are verified. Mehra et al. [14] have proposed a 7-bit parity generator and checker circuit using an SOA-MZI-based switch. Kaur et al. [15] have proposed a 3-bit parity generator and checker using SOA-MZI-based switch, and the design is simulated in a time-domain

simulation model based on the transfer matrix method (TMM). Singh et al. [16] have reported an even and odd parity generator using plasmonic metal-insulator-metal (MIM)-based waveguide, and the model is simulated in MATLAB. Maji et al. [17] have proposed a design of a 4-bit parity generator using a RSOA.

3 The Function of RSOA and ADM

In today’s communication wave, division multiplexing (WDM) is very crucial that increase the channel capacity and also increases the data management system. All-optical (A-O) wavelength converter plays a crucial role in A-O WDM network. In this paper, we have proposed dibit-based 4-bit parity generator where RSOA used as an A-O wavelength or frequency converter. The operation of RSOA is explained below.

The non-linear property of RSOA, like cross gain modulation (XGM), is used to create wavelength conversion. A probe signal (low intense) and a pump signal (high intense) are injected into the input port of SOA, but the output gives a high-power signal with probe signal frequency. The low-intense probe signal achieves high intensity because of the XGM non-linear property of SOA. This SOA is called a reflective semiconductor optical amplifier (RSOA) [17–19]. Figure 1a shows the co-propagating scheme of RSOA. A high-reflective (HR) and an anti-reflective (AR) coating are placed on the opposite faces of RSOA. Here, we have introduced C-band (1536–1570 nm) region for the pump and probe signal because RSOAs are worked properly in the C-band region. 5–20 dBm [4, 20] saturation power can be used to get the optimum result.

A frequency selector device like add/drop multiplexer (ADM) is introduced in this work. If we consider the frequency, ν_1 into the input port and frequency, ν_2 into biasing port, then ADM selects the input frequency, ν_1 at the output port whereas we cannot get any signal from drop port. But when we consider the same frequency, ν_2

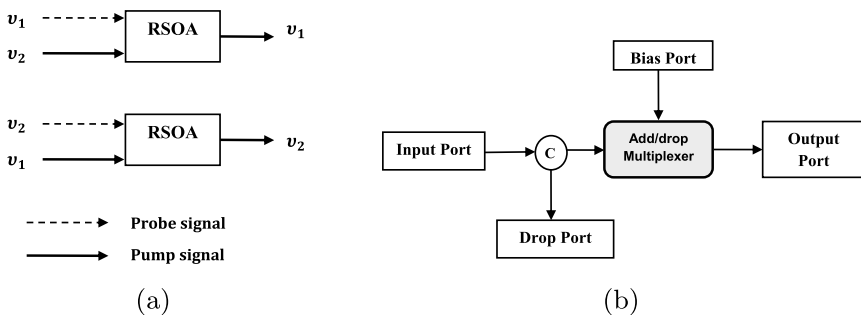


Fig. 1 a RSOA’s co-propagating scheme. b Operational diagram of add/drop multiplexer

(or ν_1) into the input and biasing port, then ADM reflects the frequency, ν_2 (or ν_1) at the drop port whereas we cannot get any signal at the output [5, 20]. The operational figure of ADM is given in Fig. 1b.

4 The Proposed Scheme of Operation of 4-Bit Generator

In this section, a 4-bit parity generator using RSOA and ADM is explained. Here, A, B, C, and D are the frequency encoded dibit-based inputs whose parity will be generated. The Boolean expression of even and odd parity generators is explained by the following equations.

$$Y_{\text{even}} = A \oplus B \oplus C \oplus D \quad (1)$$

$$Y_{\text{odd}} = \overline{A \oplus B \oplus C \oplus D} \quad (2)$$

The diagram of this design is given in Fig. 2. Here, dibit-based logic and frequency encoding techniques are used. In dibit logic [4, 20], a digit represented by two consecutive bit positions. These “0” and “1” are represented by the dibits “01” and “10”, respectively. Since the proposed design is frequency encoding, so two different frequencies ν_1 and ν_2 when placed side by side as “ $\nu_2 \nu_1$ ” indicates the logic state “1” and “ $\nu_1 \nu_2$ ” indicates the digital logic state “0”. The operation of the proposed design is based on Eqs. 1 and 2. Now, the operation scheme of the proposed frequency encoded dibit-based 4-bit parity generator describes in the following cases (Fig. 2).

4.1 Case-1: (Inputs Are the Same)

In this case, $A'=\nu_1$, $A''=\nu_2$, $B'=\nu_1$, $B''=\nu_2$, $C'=\nu_1$, $C''=\nu_2$, $D'=\nu_1$, $D''=\nu_2$ frequencies are injected into the input terminals. Therefore, ν_1 frequencies are obtained from the RSOA-4, RSOA-3, RSOA-2, and RSOA-1. The outputs of RSOA-2 and RSOA-1 are injected into A5 (ADM-5) as input and bias signals. Since both the frequencies of A5 are the same, then ν_1 is obtained at the drop port. This ν_1 acts as a pump signal of RSOA-6 and a probe signal frequency, ν_1 is fixed with it so the output frequency of RSOA-6 is ν_1 . This ν_1 acts as an biasing signal of A7, and its input signal is ν_1 which is the output frequency of RSOA-8. Since both the frequencies of A7 are the same, then A7 reflects the input signal, ν_1 at the drop port. This ν_1 works as a pump signal of RSOA-10. Therefore, the output of RSOA-10 is frequency, ν_1 . One part of this signal directly gives the dibit output Y'_{even} and another part acts as an input of A8 which is biased by the signal of frequency, ν_2 . Therefore, A8 selects the input signal ν_1 to the output and this output frequency, ν_1 works as the pump signal of RSOA-11. Since ν_2 is the probe frequency of RSOA-11 so it yields ν_2 at the dibit output terminal,

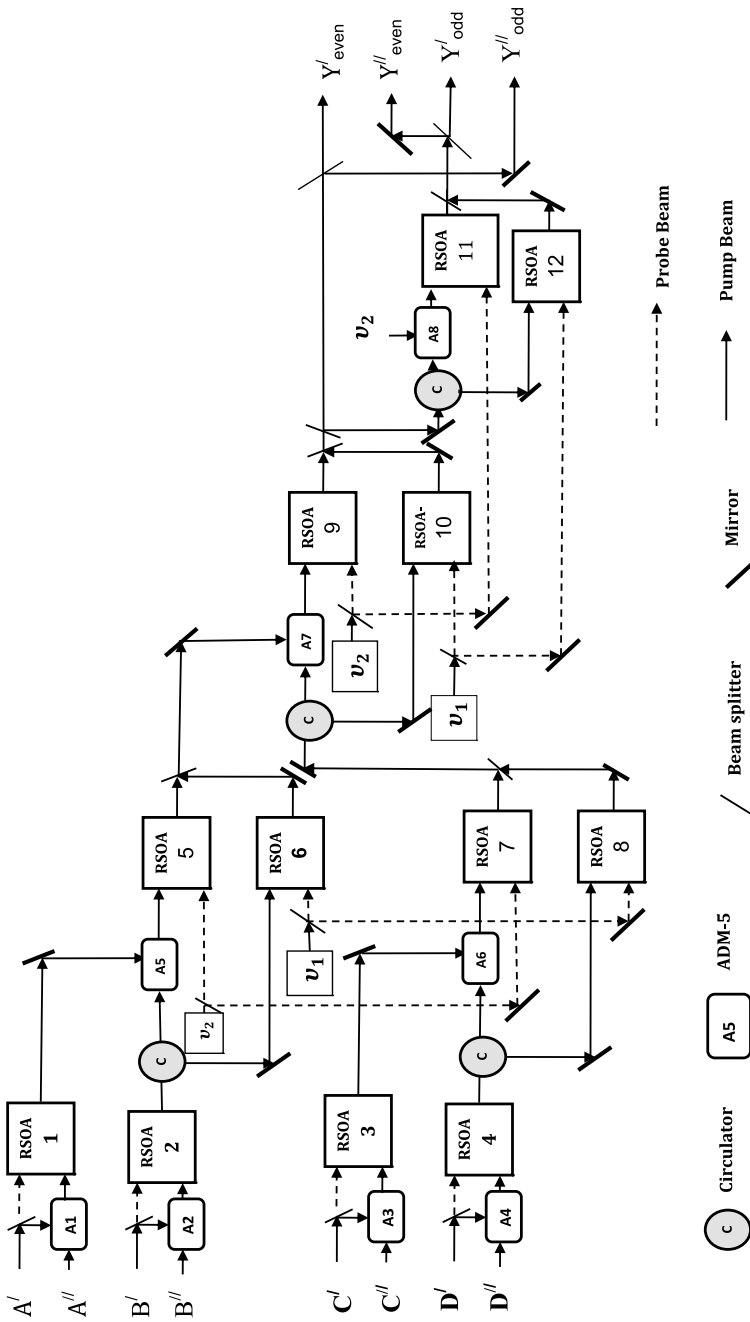


Fig. 2 4-bit parity generator's block diagram

Y''_{even} . Finally, the dibit outputs v_1 and v_2 are obtained at the output terminals, Y'_{even} and Y''_{even} respectively, that indicate the digital logic state "0". The dibit outputs v_2 , v_1 are obtained at the dibit output terminals, Y'_{odd} and Y''_{odd} that altogether indicates the digital logic state "1". Therefore, when inputs, A=0, B=0, C=0, and D=0, then the outputs show even parity, $Y_{\text{even}}=0$ and odd parity, $Y_{\text{odd}}=1$. In a similar way, other outputs can be obtained.

4.2 Case-II: (Inputs Are Different)

Here, $A'=v_2$, $A''=v_1$, $B'=v_1$, $B''=v_2$, $C'=v_2$, $C''=v_1$, $D'=v_2$, $D''=v_1$ are applied as the input signals of the device. Therefore, frequencies, v_2 , v_2 , v_1 , and v_2 are yielded by the RSOA-4, RSOA-3, RSOA-2, and RSOA-1, respectively. The outputs of RSOA-2 and RSOA-1 are inserted into the A5 as input and biasing signals, respectively. Since both the frequencies of A5 are not the same then the input signal is passed through the A5. This signal (v_1) works as a pump signal of RSOA-5, and its probe signal is v_2 so the output of RSOA-5 is v_2 . This signal of frequency, v_2 works as biasing frequency of A7. The output frequencies of RSOA-3 and RSOA-4 are v_2 which works as biasing and input frequencies of A6. Since both the frequencies of A6 are the same, then A6 reflects the input signal of frequency, v_2 at the drop port. This frequency, v_2 , works as a pump signal of RSOA-8. So the RSOA-8's output is the frequency, v_1 , because the probe signal is v_1 . This v_1 acts as the input signal of A7, and its biasing signal is v_2 which is the output frequency of RSOA-5. Since both the frequencies of A7 are not the same, then A7 selects the input signal, v_1 at the output. This v_1 works as a pump signal of RSOA-9. Therefore, the output of RSOA-9 is the frequency, v_2 . One part of this frequency directly shows the dibit output Y'_{even} and another part acts as an input of A8 which is biased with the frequency, v_2 . Therefore, A8 reflects the frequency, v_2 at the drop port, and this frequency, v_2 , works as pump signal of RSOA-12 which gives v_1 at the dibit output terminal, Y''_{even} . Finally, the dibit output v_2 and v_1 are obtained at the output terminals, Y'_{even} and Y''_{even} , respectively, which indicates the digital logic state "1". v_1 and v_2 are obtained at the dibit output terminals, Y'_{odd} and Y''_{odd} that altogether indicates the digital logic state "1". Therefore, when inputs, A=1, B=0, C=1, and D=1, then the outputs show even parity, $Y_{\text{even}}=1$ and odd parity, $Y_{\text{odd}}=0$. In this way, other outputs can be obtained and these are given in Table 1.

5 The Simulation of 4-Bit Generator

In the previous section, the operational scheme of the proposed design is explained theoretically. Now, we discuss the simulation model of the proposed design. Using MATLAB (R2018a) software, the proposed design of the parity generator is verified. RSOAs and ADMs are programmed on the basis of their characteristics using MAT-

Table 1 Simulation results of dibit-based frequency encoded parity generator (all the frequencies are given in THz range)

Input									Output			
Time (ps)	A'	A''	B'	B''	C'	C''	D'	D''	Y' _{even}	Y'' _{even}	Y' _{odd}	Y'' _{odd}
0–10	193.5	194.1	193.5	194.1	193.5	194.1	193.5	194.1	193.5	194.1	194.1	193.5
10–20	193.5	194.1	193.5	194.1	193.5	194.1	194.1	193.5	194.1	193.5	193.5	194.1
20–30	193.5	194.1	193.5	194.1	194.1	193.5	193.5	194.1	194.1	193.5	193.5	194.1
30–40	193.5	194.1	193.5	194.1	194.1	193.5	194.1	193.5	193.5	194.1	194.1	193.5
40–50	193.5	194.1	194.1	193.5	193.5	194.1	193.5	194.1	194.1	193.5	193.5	194.1
50–60	193.5	194.1	194.1	193.5	193.5	194.1	194.1	193.5	193.5	194.1	194.1	193.5
60–70	193.5	194.1	194.1	193.5	194.1	193.5	193.5	194.1	193.5	194.1	194.1	193.5
70–80	193.5	194.1	194.1	193.5	194.1	193.5	194.1	193.5	194.1	193.5	193.5	194.1
80–90	194.1	193.5	193.5	194.1	193.5	194.1	193.5	194.1	194.1	193.5	193.5	194.1
90–100	194.1	193.5	193.5	194.1	193.5	194.1	194.1	193.5	193.5	194.1	194.1	193.5
100–110	194.1	193.5	193.5	194.1	194.1	193.5	193.5	194.1	193.5	194.1	194.1	193.5
110–120	194.1	193.5	193.5	194.1	194.1	193.5	194.1	193.5	194.1	193.5	193.5	194.1
120–130	194.1	193.5	194.1	193.5	193.5	194.1	193.5	194.1	193.5	194.1	194.1	193.5
130–140	194.1	193.5	194.1	193.5	193.5	194.1	194.1	193.5	194.1	193.5	193.5	194.1
140–150	194.1	193.5	194.1	193.5	194.1	193.5	193.5	194.1	194.1	193.5	193.5	194.1
150–160	194.1	193.5	194.1	193.5	194.1	193.5	194.1	193.5	193.5	194.1	194.1	193.5

LAB language. If frequencies, $\nu_1=193.5$ THz (wavelength=1550 nm) and $\nu_2=194.1$ THz (wavelength=1545 nm) are considered as the probe signal and pump signal, then 193.5 THz is obtained at the output port, whereas 194.1 THz is obtained at the output port when $\nu_1=193.5$ THz (wavelength=1550 nm) and $\nu_2= 194.1$ THz (wavelength=1545 nm) are considered as the pump and probe signal. If frequencies, $\nu_1=193.5$ THz and $\nu_2=194.1$ THz, are considered as the input and biasing signal, then ADM selects the input frequency, 193.5 THz at the output port, whereas the

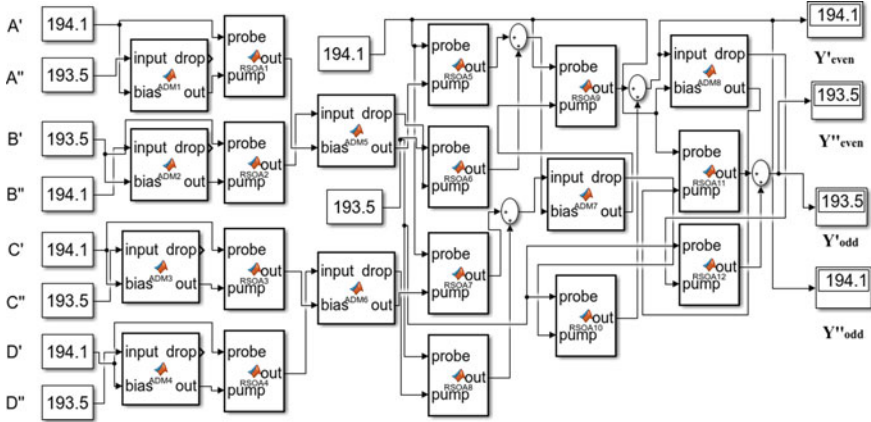


Fig. 3 Simulink model diagram of parity generator when dibit input signal is “1011”

drop port gives nothing. But if the input and biasing frequencies are the same then, ADM reflects the input frequency at the drop port through the circulator “C” whereas output gives nothing. On the basis of these considerations, the proposed design is simulated. In Figs. 3 and 4, dibits $\langle 193.5 \rangle \langle 194.1 \rangle$ and $\langle 194.1 \rangle \langle 193.5 \rangle$ represent the digital logic state “0”, and “1”, respectively. From Fig. 3, the dibits $\langle 194.1 \rangle \langle 193.5 \rangle$, $\langle 193.5 \rangle \langle 194.1 \rangle$, $\langle 194.1 \rangle \langle 193.5 \rangle$, and $\langle 194.1 \rangle \langle 193.5 \rangle$ are injected into the dibit inputs “A”, “B”, “C”, and “D” terminals of the device, respectively. After the simulation, $\langle 194.1 \rangle \langle 193.5 \rangle$ and $\langle 193.5 \rangle \langle 194.1 \rangle$ are obtained as dibit even parity, Y_{even} and dibit odd parity, Y_{odd} at the output terminals respectively. From Fig. 4, $\langle 194.1 \rangle \langle 193.5 \rangle$, $\langle 193.5 \rangle \langle 194.1 \rangle$, $\langle 193.5 \rangle \langle 194.1 \rangle$, and, $\langle 194.1 \rangle \langle 193.5 \rangle$ are applied to the dibit inputs “A”, “B” and “C”, and “D” terminals, respectively, then at the output terminals $\langle 193.5 \rangle \langle 194.1 \rangle$ and $\langle 194.1 \rangle \langle 193.5 \rangle$ are obtained as dibit even parity, Y_{even} and dibit odd parity, Y_{odd} at the output terminals, respectively. Similar way, other outputs are obtained corresponding to the applied inputs. Simulation results are given in Table 1.

6 Results and Discussion

In this results section, the verification of the simulation outputs and the theoretical predictions are discussed. Here, different frequency signals are applied to the proposed design in 10 ps time intervals. The input and output signal waveforms are given in Figs. 5 and 6. After the verification of simulation results (Table 1) and graphical results, it is concluded that the outputs of the 4-bit parity generator is accurately verified. A comparative study with the previous work is also given in Table 2.

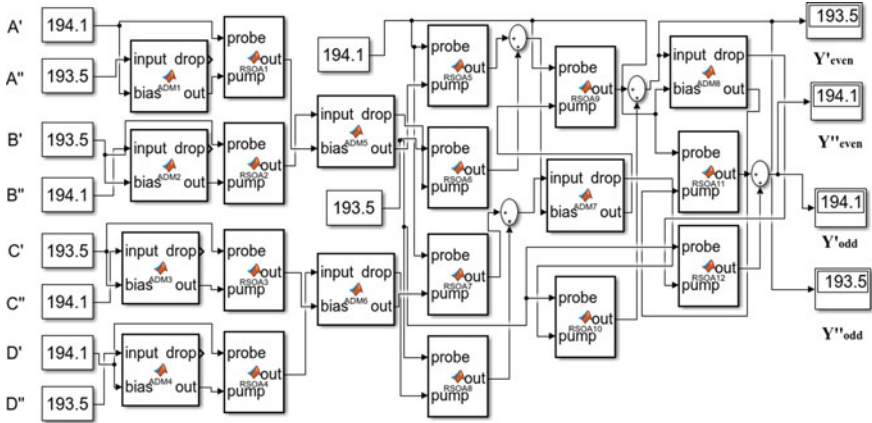


Fig. 4 Simulink model diagram of parity generator when dibit input signal is “1001”

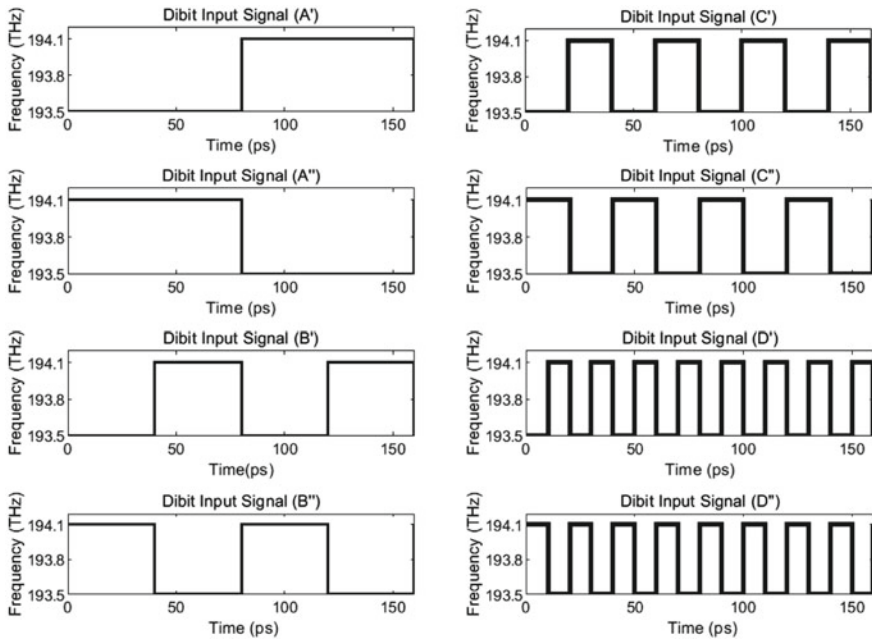


Fig. 5 Dibit input signal waveforms of parity generator

Table 2 A comparison with the previous work

Work	Bit pattern given	Dibit logic used or not	RSOA used or not	Even and odd parity generator in a single device	No. of bit operation	Without extra control signal
Rakshit et al. [6]	Yes	No	No	No	3-bit	No
Nair et al. [8]	Yes	No	No	No	3-bit	No
Bhattacharyya et al. [9]	Yes	No	No	No	3-bit	No
Dimitriadou et al. [11]	Yes	No	No	No	4-bit	No
Kumar and Raghuvanshi [13]	Yes	No	No	No	4-bit	No
Majik et al. [17]	Yes	No	Yes	Yes	4-bit	No
This Work	Yes	Yes	Yes	Yes	4-bit	Yes

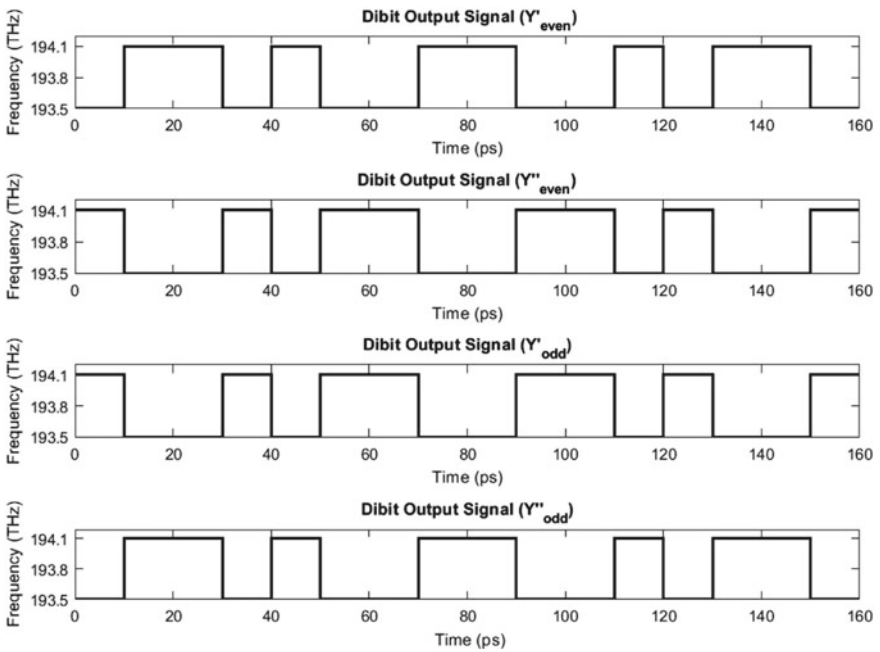


Fig. 6 Dibit output signal waveforms of parity generator

7 Conclusion

A frequency encoded dibit-based parity generator using RSOA and ADM is designed. Here, for error-free dibit logic, we introduced a dibit control units at the input section that decreases the bit error problems at the output. The proposed design promotes reliable and faithful operation. Here, MATLAB Simulink (R2018a) software is used to verify the proposed design. Frequency encoding and dibit representation technique give more advantages to the all-optical domain. In future, we intend to develop a higher bit parity generator. We also intend to use the proposed design in cryptographic systems as well as in the development of binary to gray code converter device in future.

Acknowledgements We are thankful to the Department of Physics, Bankura University, Bankura, Pin-722155, West Bengal, India for support to conduct this research work.

References

1. Samanta D (2019) Implementation of polarization-encoded quantum Fredkin gate using Kerr effect. *J Opt Commun*
2. Mandal M, Mukhopadhyay S (2021) Photonic scheme for implementing quantum square root controlled z gate using phase and intensity encoding of light. *IET Optoelectron* 15(1):52–60
3. Bosu Surajit, Bhattacharjee Baibaswata (2021) A novel design of frequency encoded multiplexer and demultiplexer systems using reflected semiconductor optical amplifier with simulative verification. *J Opt* 50(3):361–370
4. Bosu S, Bhattacharjee B (2021) All-optical frequency encoded dibit-based half adder using reflective semiconductor optical amplifier with simulative verification. In: 2021 devices for integrated circuit (DevIC. IEEE,), pp 388–392
5. Bosu S, Bhattacharjee B (2021) All-optical frequency encoded dibit-based half subtractor using reflective semiconductor optical amplifier with simulative verification. In: International conference on paradigms of communication, computing and data sciences (PCCDS 2021). Springer, Berlin
6. Rakshit JK, Roy JN, Chattopadhyay T (2013) Design of micro-ring resonator based all-optical parity generator and checker circuit. *Opt Commun* 303:30–37
7. Wang F, Gong Z, Hu X, Yang X, Yang H, Gong Q (2016) Nanoscale on-chip all-optical logic parity checker in integrated plasmonic circuits in optical communication range. *Sci Rep* 6(1):1–8
8. Nair N, Kaur S, Goyal R (2018) All-optical integrated parity generator and checker using an SOA-based optical tree architecture. *Curr Opt Photonics* 2(5):400–406
9. Bhattacharyya A, Gayen DK, Chattopadhyay T (2014) All-optical parallel parity generator circuit with the help of semiconductor optical amplifier (SOA)-assisted Sagnac switches. *Opt Commun* 313:99–105
10. Chowdhury KR, De D, Mukhopadhyay S (2005) Parity checking and generating circuit with nonlinear material in all-optical domain. *Chin Phys Lett* 22(6):1433
11. Dimitriadou E, Zoiros KE, Chattopadhyay T, Roy JN (2013) Design of ultrafast all-optical 4-bit parity generator and checker using quantum-dot semiconductor optical amplifier-based Mach-Zehnder interferometer. *J Comput Electron* 12(3):481–489
12. Ghosh AK (2010) Parity generator and parity checker in the modified trinary number system using savart plate and spatial light modulator. *Optoelectron Lett* 6(5):325–327

13. Kumar A, Raghuwanshi SK (2015) Implementation of optical gray code converter and even parity checker using the electro-optic effect in the Mach-Zehnder interferometer. *Opt Quant Electron* 47(7):2117–2140
14. Mehra R, Jaiswal S, Dixit HK (2013) Parity checking and generating circuit with semiconductor optical amplifier in all-optical domain. *Optik* 124(21):4744–4745
15. Kaur S, Shukla MK (2017) All-optical parity generator and checker circuit employing semiconductor optical amplifier-based mach–zehnder interferometers. *Opt Appl* 47(2)
16. Singh L, Bedi A, Kumar S (2017) Modeling of all-optical even and odd parity generator circuits using metal-insulator-metal plasmonic waveguides. *Photon Sens* 7(2):182–192
17. Maji K, Mukherjee K, Raja A, Roy JN (2020) Numerical simulations of an all-optical parity generator and checker utilizing a reflective semiconductor optical amplifier at 200 gbps. *J Comput Electron* 1–15
18. Kotb A, Zoiros KE, Li W, Guo C (2021) Theoretical investigation of 120 gb/s all-optical and and or logic gates using reflective semiconductor optical amplifiers. *Opt Eng* 60(6):066107
19. Saha S, Mukhopadhyay Sourangshu et al (2020) All optical frequency encoded quaternary memory unit using symmetric configuration of MZI-SOA. *Opt Laser Technol* 131:106386
20. Ghosh B, Hazra S, Haldar N, Roy D, Patra SN, Swarnakar J, Sarkar PP, Mukhopadhyay S (2018) A novel approach to realize of all optical frequency encoded dibit based XOR and XNOR logic gates using optical switches with simulated verification. *Opt Spectrosc* 124(3)

Design and Implementation of an Efficient QCA-Based Multilayer Multi-Bit Parallel Shift Register Using Reversible Level-Sensitive ‘D’ Flip-Flop



Rupsa Roy, Swarup Sarkar, and Sourav Dhar

Abstract Quantum dot cellular automata (QCA) is a current low-power nano-technology that is an effective replacement of popular CMOS technology in this recent nano-technical digital world. The QCA offers high-speed yet less complex digital circuitry. Further, QCA supports multilayer design with reversibility. In this paper, a novel nano-sized, high-speed temperature tolerance low-power multi-bit (4 bit and 8 bit) shift register with parallel-in parallel-out (PIPO) operation is designed in multilayer QCA platform using less complex reversible ‘D’ flip-flop. In this proposed design, the pipelined structure is not used up to 4 bit. Two 4-bit registers are placed in a pipelined manner to form an 8-bit structure. QCA designer software is mainly used in this research work to get the QCA-based designs and then check and calculate the required parameters. To establish the novelty of the proposed design, a parametric comparison among this proposed design and most optimized existing designs is shown in this paper based on the parameters like occupied area, cell complexity, cost, and delay.

Keywords ‘D’ flip-flop · Multilayer · QCA · Register · Reversibility

1 Introduction

Low-power high-speed high-dense less complex integrated circuit design has become a challenge in this nano-technical era. Advance complex digital circuit designs are formed by using transistor-level technologies; mainly, CMOS technology is highly available in the recent electronic market due because it maintains Moore’s law [1] to maintain the high scalability of devices. But, now, more advanced technology beyond CMOS technology is required for device area, complexity, delay, power, and cost

R. Roy (✉) · S. Sarkar · S. Dhar
Department of ECE, Sikkim Manipal Institute of Technology, Sikkim Manipal University,
Sikkim, India
e-mail: rupsa_202010004@smit.smu.edu.in

S. Sarkar
e-mail: swarup.s@smit.smu.edu.in

optimization. Thus, in this paper, an advanced technology QCA with electron-spin operational criterion is selected to design a novel advanced shift register.

QCA is established by Lent et al. in 1993 [2, 3]. This technical platform is considered as a low-power high-speed technology due to the presence of quantum wire, which can flow information from input to output by electro-repulsion criterion between two consecutive quantum cells, and it can be operated in tetra-Hz frequency range [7]. A basic 4-dot quantum cell is used in this work, and when these quantum cells are specified one after another, it presents a quantum wire [4]. 2-dots among the 4-dots in a quantum cell are occupied by moving electrons, and they are always placed diagonally to neighbor electrons because of the electrostatic repulsive force between the same charged carriers. The movement takes place by tunneling, which is happened from one dot to another in a cell. Thus, in this technology, leakage current flow is very low and energy dissipation is in the pico-joule range. These quantum cells can be placed in different layers easily in this technology to convert it into a 3D manner to reduce the unit area-occupation of proposed circuitry.

In this paper, the selected advanced technology QCA with the advancement of multilayer circuitry, is utilized for getting a novel advanced shifter. There are four different types of shift registers basically, and among them, serial-in-serial-out (SISO) and parallel-in-parallel-out (PIPO) are fewer complexes compare to others. But, QCA-based design of PIPO is more optimized than QCA-based design of SISO, which is proved in paper [6]. Thus, the PIPO shift register is selected in this paper, where more optimization is done by using a multilayer structure and reversible ‘D’ flip-flops. The main contributions of this paper are:

- Design a reversible QCA-based ‘D’ flip-flop with less complexity, area-occupation, power-dissipation, delay, and cost compared to a most optimized existing design.
- Design a QCA-based multilayer 3D PIPO shift register up to 4 bit using the proposed D flip-flop with less complexity, area-occupation, power-dissipation, delay, and cost compared to a most optimized existing design.
- Increase the bit number of proposed circuitry up to 8-bit by adding two 4-bit proposed registers in the pipelined manner and check the complexity, area-occupation, power-dissipation, delay, and cost.
- Check the output function of the proposed register design at the time of room temperature increment and layer separation gap decrement.

The whole contribution is thoroughly presented in five different sections: Sect. 2 presents the theory-based background of proposed technical field and logical field, Sect. 3 presents the review of related work, Sects. 4 and 5 present the design and outcomes of the proposed ‘D’ flip-flop and multilayer multi-bit proposed register, respectively, and Sect. 6 presents the conclusion of this proposed work.

2 Theoretical Background

The “3-input majority gate or MG,” “5-input majority gate or MG,” and “inverter gate” (give inverted form of input) are the most effective and highly utilized gates in our proposed low-power 4-dotted QCA-design technology, based on the previously discussed binary ‘0’ and binary ‘1’ selection in QCA platform. The polarity of inputs and output in a “3-input MG” are matched. The output of “3-input MG” is shown in Eq. 1, where A, B, and C are the three inputs. “AND Gate” and “OR Gate” are also designed by changing the polarity -1 and + 1 of one of the three inputs in “3-input MG” (given in Eqs. 2 and 3, respectively) (Fig. 1a presents a clear image of “3-input MG”) [7–10]. Another important multi-input MG is the “5-input MG.” Equation 4 represents a “5-input MG” with input A, B, C, D, and E. If the three inputs among five are merged and change the clock zone (discussed below in QCA-based clock scheme part) from clock 0 to clock 1 near output section, it gives a “3-input XOR” output, and without changing the clock zone, it presents the output of normal “3-input MG” with 4% output strength increment, but the cell complexity is increased from 5 to 11. The “5-input MG” is shown in this section in Fig. 1b, which is representative of the “3-input XOR” operation. As we know, “inverter gate” is rapidly required to design any digital-based circuitry, which gives a “NOT-Gate” outcome. In this theory about QCA-based widely used conventional logic gates, this clear reflection of basic single later “inverter gate” is also included, which is given in Fig. 1c.

$$3 \text{ input MG}(A, B, C) = AB + BC + AC \tag{1}$$

$$3 \text{ input MG}(A, B, 0) = A.B \tag{2}$$

$$3 \text{ input MG}(A, B, 1) = A + B \tag{3}$$

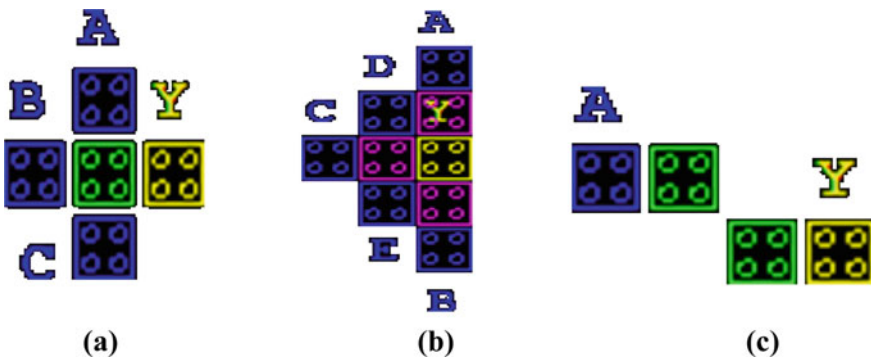


Fig. 1 QCA-based basic single-layer structure of **a** “3-input MG,” **b** “5-input MG,” and **c** “Inverter Gate”

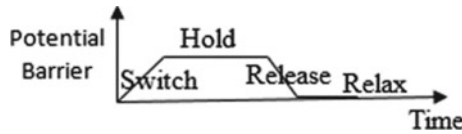


Fig. 2 Clock phases used in QCA technology

$$\begin{aligned}
 5 \text{ input MG}(A, B, C, D, E) = & ABC + ABD + ABE + ADE \\
 & + ACE + ACD + BCE + BCD \\
 & + BDE + CDE
 \end{aligned} \tag{4}$$

In QCA, a different clocking scheme helps to control the information flow from one part to another in a circuit, maintain the power gain by restoring the lost signal energies in the environment, and identify the delay of the design. This is a pipelined structure, where four clock zones with four clock phases are presented. Four clock zones are: clock zone-1, clock zone-2, clock zone-3, and clock zone-4, and 4 clock phases with 90° phase difference are: switch, hold, release, and relax, which are given in Fig. 2 [11]. In this clocking scheme, when the given clock is high, the potential barrier between two dots becomes low and the total polarization of the circuit is 0, and when the clock becomes low, the potential barrier between two dots is high, and electrons are placed in the dots through a tunnel according to the polarization of cells, which depends on the specified neighbor-cells.

In the above-discussed conventional logic gates, the “information erase with the copy” cannot be possible. So, energy is dissipated per bit, which can be maintained through design-adiabaticity, and this adiabatic logic can be followed by using reversible gates, where “information erase with the copy” can be maintained by the “Bennet clock scheme” [12–14]. So, energy dissipation per bit can be maintained by adding this reversible gate. In a conventional gate, only the outcomes depend on the inputs, but in this type of gate (reversible gate), the inputs are also outcome dependent. That means the arrangement of outputs is also able to represent the arrangement of inputs and vice versa in this proposed reversible gate. To form this arrangement in a reversible gate, it is required to maintain the same number of inputs and outputs. Thus, we can say that a proper exploration of the advantages of QCA-based circuitry can be possible by using a reversible gate, and in a multilayer platform, it becomes more effective because of its energy-controlled nature. Due to this reason, this proposed design is formed in a hybrid manner by adding a reversible gate with widely used “3-input MG.” Figure 3 represents a proper block diagram of a basic reversible gate.

In a circuit design, crossing-criterion of two wires is a very common and important thing, which becomes more complex at the time of operations of a circuit increment. Delay, area, output strength, and power-dissipation also depend on this criterion. Thus, crossover design selection in circuit formation is a challenging part. In our proposed technology, QCA, coplanar, multilayer, and crossover by changing clock zones of two different crossing wires is presented [15]. In our work, multilayer

Fig. 3 Block diagram of basic reversible gate

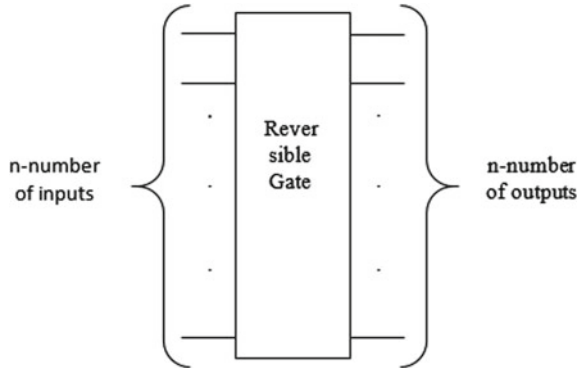
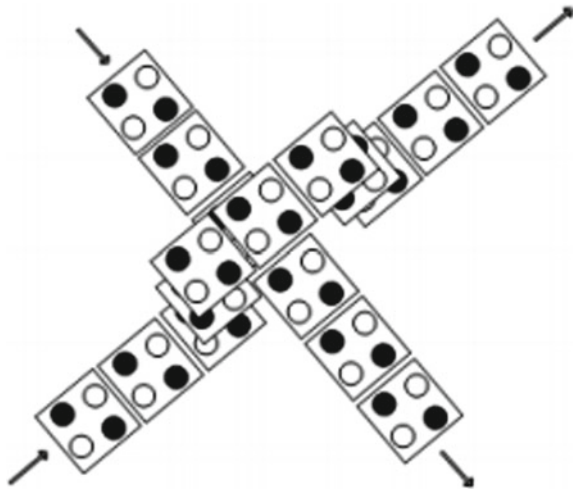


Fig. 4 Multilayer QCA-based structure [5]



crossover is used, where different cells are specified in different layers, which acts as an inverter in two different consecutive layers. But, this type of structure can increase the output strength compare to coplanar form and also compare to a single later “inverter gate” with a 25% delay reduction. In this multilayer structure, the vertically separated quantum cells are tuned to match their kink energy in the horizontal plane unlike the transistor-based structure [16]. In Fig. 4, this bridge-looking multilayer QCA-based structure is presented.

3 Review of Related Work

As we know, D flip-flop is an important widely used sequential circuit in the digital world due to its simple operation, and in this paper; ‘D’ flip-flop is utilized to form

a PIPO shift register in the QCA platform. Because of the simplicity and huge application of this proposed circuitry using ‘D’ flip-flop, different novel designs are published in different years related to this proposed work. In this work, some optimized-related designs from 2019 to 2021 are discussed. In 2019’s March, a QCA-based level-sensitive ‘D’ flip-flop is introduced by Ting Li et al [22], where only 28 quantum cells are required, and in that paper, the optimization-based advantages of 3-bit PIPO register based on the proposed D flip-flop compare to 3-bit SISO register based on the proposed ‘D’ flip-flop is also proved in single-layer platform.

After 4 months, a novel Universal Shift Register is presented by Jun-Cheol Jeon (published in 2020), where 4-bit PIPO is a part, and in this design, a 24 quantum cell-based ‘D’ flip-flop is used with 0.25 higher delays compared to the previous one in a single-layer platform [23]. Next, in 2020, 3-bit PIPO and SISO are presented by Shuyan Fan et al to prove the advancement of PIPO compared with SISO again based on area-occupation, cell complexity, and delay, where dual edge-triggered ‘D’ flip-flop is used [6].

In 2021, another novel ‘D’ flip-flop is presented applying only 21 quantum cells and the same delay by Salma Yaqoob et al [24] to form a single-layer SISO register. But this ‘D’ flip-flop formation can be optimized more, and optimization of shift registers can also be possible, which is shown in this paper.

4 Proposed ‘D’ Flip-Flop

‘D’ flip-flop or Delay flip-flop or Data flip-flop can store data, and it can work by maintaining reversible logic [4]. This data-storing process follows Eq. 5, and in this work, a level-sensitive ‘D’ flip-flop is proposed, where the reversible logic is applied. The block diagram of ‘D’ flip-flop is presented in this part in Fig. 5, and based on this diagram, a novel level-sensitive ‘D’ flip-flop is formed in this work using 20 quantum cells (18*18 nm² per cell area).

$$Q = D.CLK + Q.CLK' \quad (5)$$

The proposed ‘D’ flip-flop is presented in this portion in Fig. 6, which follows reversibility. In this proposed operation, output C follows the input CLK to make the design reversible and when CLK is 0, the output Q shows the previous state,

Fig. 5 Block diagram of proposed ‘D’ flip-flop [22]

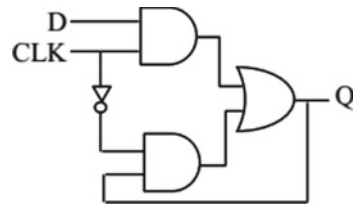




Fig. 6 Proposed level-sensitive ‘D’ flip-flop with reversibility

and otherwise, the ‘D’ input is followed by output Q, which is presented here by the outcome of the proposed level-sensitive ‘D’ flip-flop (in Fig. 7), achieved through QCA Designer 4.0 [25, 26] software.

The outcome of the proposed level-sensitive ‘D’ flip-flop is presented in Fig. 7 with THz frequency range and 2 ps delay in output C. This proposed structure is more optimized, where delay reduction of 33.3% is possible for only 2% output strength reduction compare to previously discuss most optimized structures, which is shown in this section in Table 1.

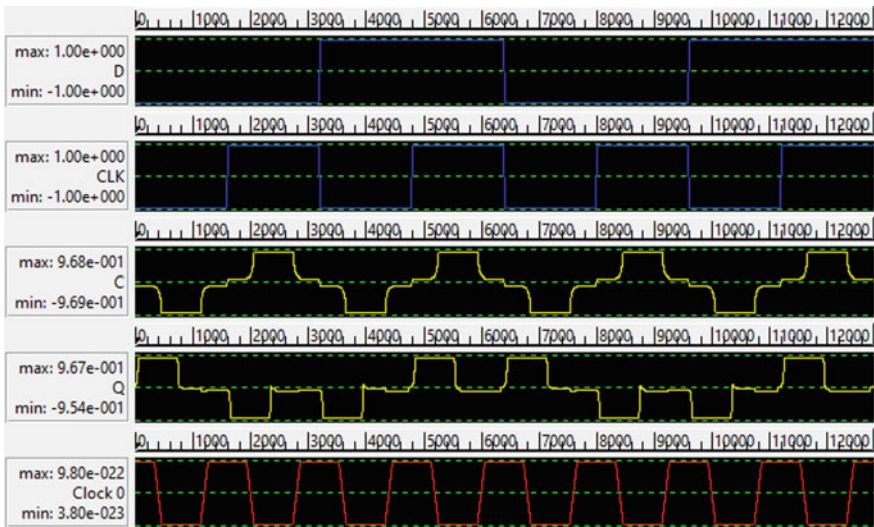


Fig. 7 Simulated outcomes of proposed level-sensitive ‘D’ flip-flop with reversibility

Table 1 Comparison table: proposed level-sensitive ‘D’ flip-flops of different paper

Referred papers of different years	Cell complexity	Area-occupation (μm ²)	Delay (ps)	Areal-power-dissipation (nW)	Cost (area*delay) [27]
Li and Kornovich [22]	28	0.03	3	30	0.09
Jeon [23]	24	0.02	4	20	0.08
Yaqoob et al. [24]	21	0.02	4	20	0.08
Proposed	20	0.02	2	20	0.04

5 Proposed PIPO Shift Register

The advancement to get an optimized register based on area-occupation, delay, areal-power-dissipation [17–21], complexity, and cost PIPO is better than SISO, which is already discussed previously. In this work, the optimization of the PIPO register is increased by using the proposed ‘D’ flip-flop in the multilayer QCA platform. The novel advanced proposed structure of multilayer 3D PIPO with the simulated result and proper parametric comparison is revealed in this portion. At first, the 4-bit structure of the proposed PIPO is presented here, and then, this is converted to 8 bit by adding another 4-bit structure.

The proposed three layers of novel 4-bit PIPO register is presented separately and in the combined manner in Fig. 8 and 9 respectively, where D0-D3 data are given and the outputs are got from Q0-Q3 with another output C which is the direct output of CLK signal (used to make the design reversible without increase the cell count and area-occupation). The outcome of this proposed 4-bit PIPO is presented in Fig. 10, where the output is achieved after 0.5 clock cycles for Q0-Q3. Next, this presented 4-bit structure is used to design an 8-bit PIPO with the same number of layers to control the high-temperature and complexity issues. These three layers of the proposed 8-bit

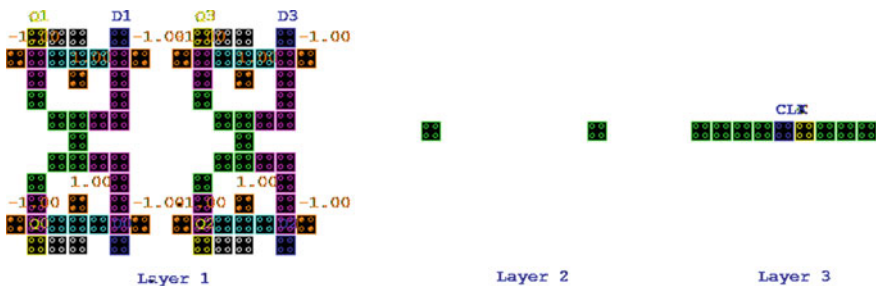


Fig. 8 Three layers of proposed 4-bit PIPO register using level-sensitive ‘D’ flip-flop with reversibility

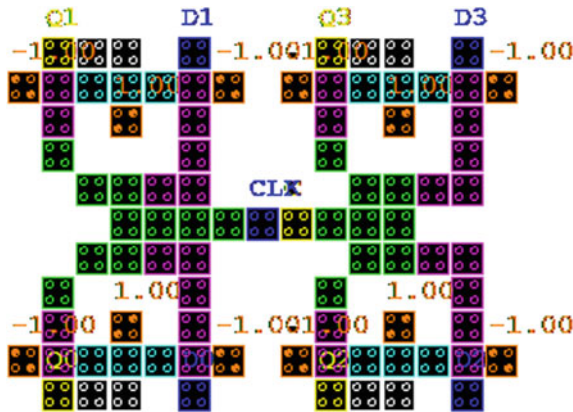


Fig. 9 Proposed combined multilayer 3D 4-bit PIPO register using level-sensitive ‘D’ flip-flop with reversibility

structure, where two presented 4-bit structures are added are shown here separately and also in a combined manner in Fig. 11 and 12, respectively, with the simulated outcomes (in Fig. 13).

The parametric and simulated outcomes of proposed designs show the advancement of the 4-bit PIPO compared to the 3-bit design of paper [22] and the 4-bit design of paper [23]. After checking the advancement of the proposed novel 4-bit PIPO structure of this paper, this optimized structure is doubled without increasing the layer number. In Table 2, a parametric investigation of different structures of multi-bit shift registers is presented based on the cell complexity, area-occupation, delay, areal-power-dissipation, and cost. The temperature increment effects on device performance are also checked in this work. This design can work properly up to 5 K temperature more than room temperature with the same output strength, but after 5 K, the output strength is reduced due to electron scattering at high temperature. In 24 K temperature, the output strength is reduced 37% compared to real output strength, and above this, the device starts to do malfunction. Multilayer structures face another problem, which is the volume increment of the proposed structure. But, this proposed design can work properly with the same output strength at the time of layer separation gap reduction from 11.5 nm to 2.5 nm (78% reduction in layer separation gap can be possible) with the same temperature tolerance, power-dissipation, and delay. A graphical representation is also shown in this section in Fig. 14 for a better understanding of these high-temperature and layer separation gap reduction effects.

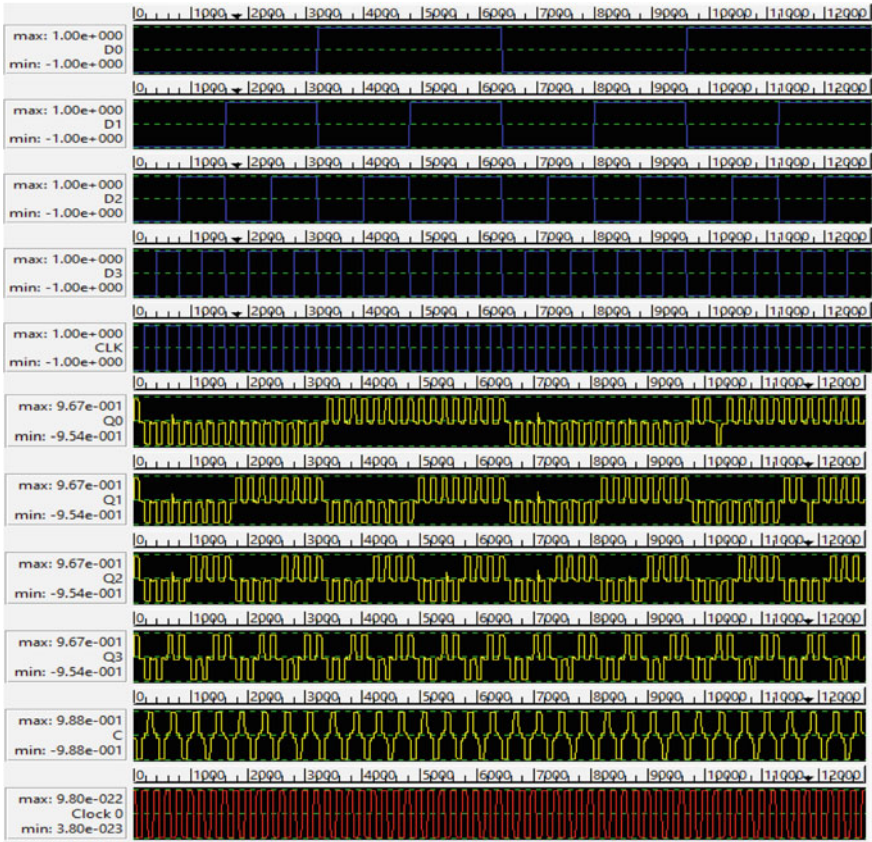


Fig. 10 Simulated result of proposed multilayer 3D 4-bit PIPO register using level-sensitive ‘D’ flip-flop with reversibility

6 Conclusion

A novel area, delay, complexity, cost, and dissipated power efficient QCA-based multilayer 3D 4-bit and 8-bit PIPO shift register are presented in this paper using optimized novel level-sensitive ‘D’ flip-flop with reversibility. In the proposed ‘D’ flip-flop, 4.8% cell complexity, 33.3% delay, and 50% cost decrement with same area-occupation and areal-power-dissipation is possible compared to previously published most optimized ‘D’ flip-flop parameters [22, 22]. Further, the proposed 4-bit multilayer PIPO structure is capable of reducing 19% cell complexity, 20% unit area-occupation and areal-power-dissipation, 33.3% delay, and 46.7% cost compared to previously published optimized parameters of single-layer 3-bit PIPO register [22]. The novel efficient multilayer 8-bit structure is formed in this work by using the proposed optimized previously discussed components with the same number of layers compared to the 4-bit structure. This proposed 8-bit 3-layered PIPO register

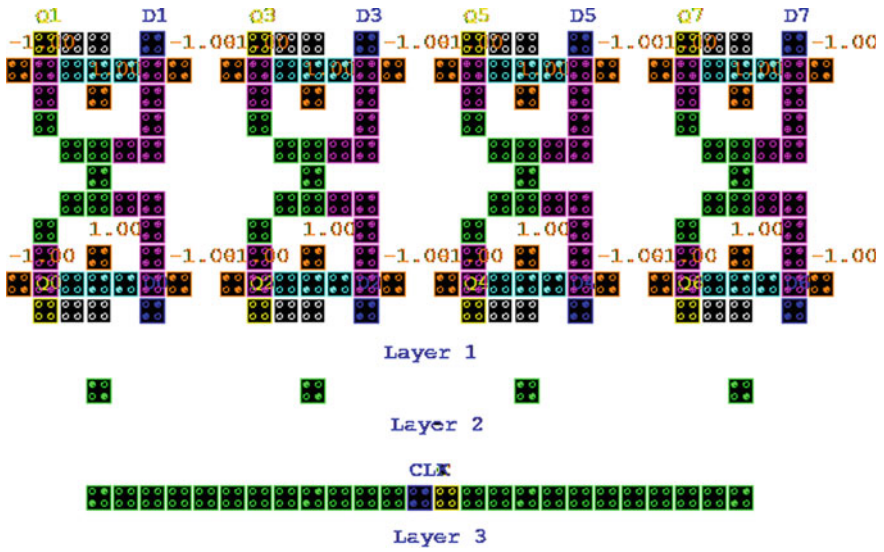


Fig. 11 Three layers of proposed 8-bit PIPO register using level-sensitive ‘D’ flip-flop with reversibility

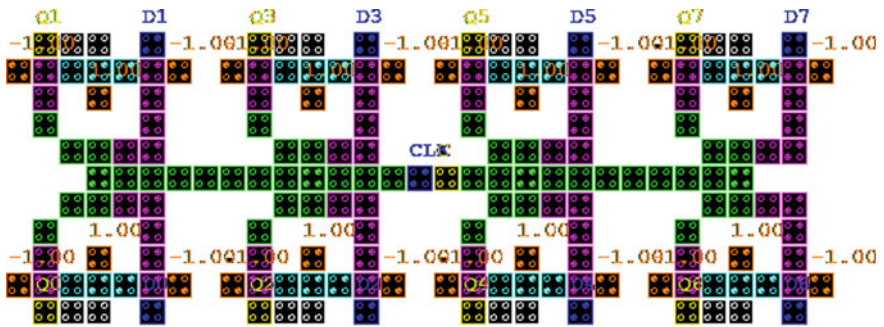


Fig. 12 Proposed combined multilayer 3D 8-bit PIPO register using level-sensitive ‘D’ flip-flop with reversibility

can perform with efficient output strength up to 5 K temperature more than room temperature, and this performance is continued at the time of layer separation gap reduction from 11.5 nm to 2.5 nm. Proper fabrication with hardware verification of multi-bit advanced shift registers can be possible in the future.

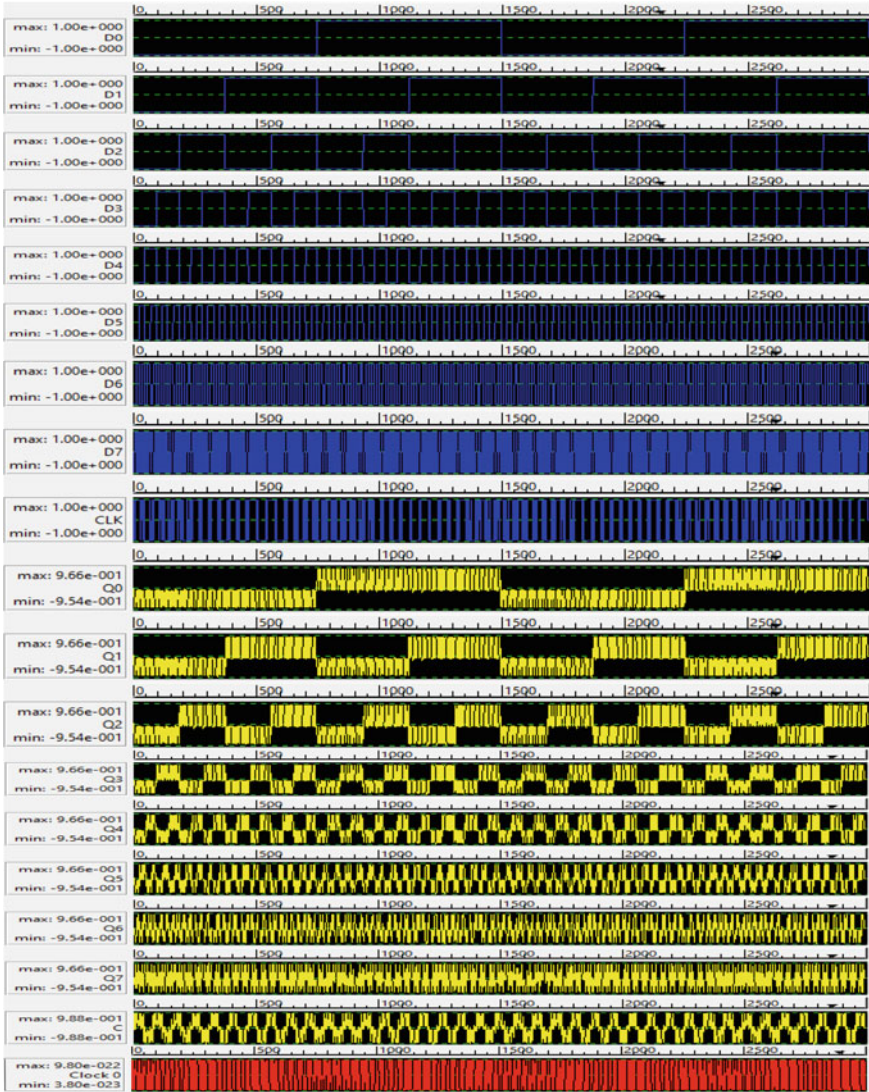


Fig. 13 Simulated result of proposed multilayer 3D 8-bit PIPO register using level-sensitive ‘D’ flip-flop with reversibility

Table 2 Table of parametric investigation: proposed multi-bit PIPO shift registers of different papers

PIPO designs of different papers	Cell Complexity	Area-occupation (μm^2)	Latency (ps)	Cost	Layer number	Areal-power-dissipation (nW)
3-bit PIPO [22]	115	0.10	3	0.30	1	100
4-bit PIPO [23]	136	0.099	4	0.396	1	99
Proposed 4-bit PIPO	93	0.08	2	0.16	3	80
Proposed 8-bit PIPO	193	0.16	3	0.84	3	160

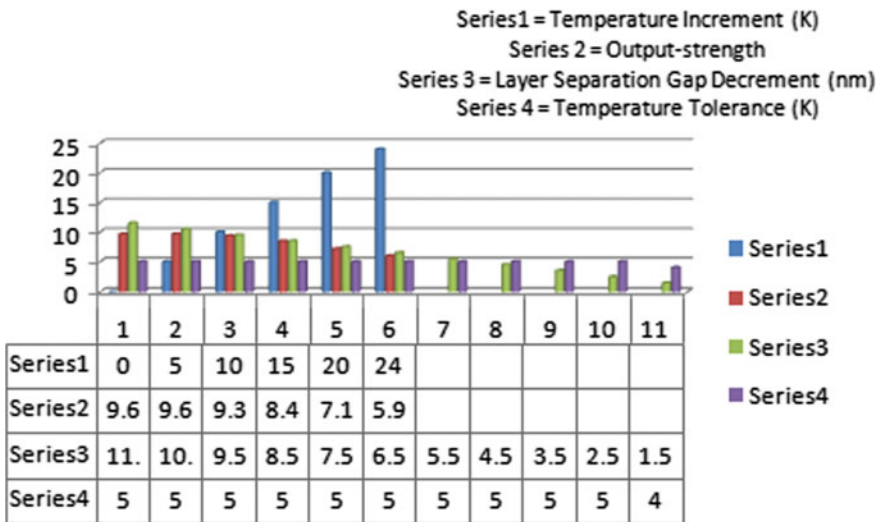


Fig. 14 Graphical representation of temperature increment affects device output strength and temperature tolerance for layer separation gap reduction

References

1. Moore GE (1965) Cramping more components onto integrated circuits. Electronics 38(2):114–117

2. Lent CS, Tougaw PD, Porod W, Bernstein GH (1993) Quantum cellular automata. *Nanotechnology* 4(1):49–57
3. Tougaw PD, Lent CS (1994) Logical devices implemented using quantum cellular automata. *J Appl Phys* 75:1818–1825
4. Rad SK, Heikalabad SR (2017) Reversible Flip-Flops in Quantum-Dot Cellular Automata. *Int J Theor Phys* 56(9):1–15
5. Jeon J-C (2019) Time-efficient parity generator based on quantum-dot cellular automata. *Int J Civ Eng Technol (IJCIET)* 10:715–723
6. Fan S, Khamesinia MS (2021) An efficient design of parallel and serial shift registers based on quantum-dot cellular automata. *Int J Theor Phys* 60:2400–2411
7. Oskouei SM, Ghaffari A (2019) Designing a new reversible ALU by QCA for reducing occupation area. *J Supercomput* 75(8):5118–5144
8. Babaie S et al (2019) Design of an efficient multilayer arithmetic logic unit in quantum-dot cellular automata (QCA). *IEEE Trans Circuits Syst* 66(6):963–967
9. Walus K, Dysart TJ et al (2004) QCA designer: a rapid design and simulation tool for quantum-dot cellular automata. *IEEE Trans Nanotechnol* 3(1):26–31
10. Roy SS (2016) Simplification of master power expression and effective power detection of QCA device. In *IEEE students' technology symposium*, pp 272–277
11. Askari M, Taghizadeh M (2011) Logic circuit design in nano-scale using quantum-dot cellular automata. *Eur J Sci Res* 48(3):516–526
12. Narimani R, Safaei B, Ejlali A (2020) A comprehensive analysis on the resilience of adiabatic logic families against transient faults Integration. *VLSI J* 72:183–193
13. Pidaparathi SS, Lent CS (2018) Exponentially adiabatic switching in quantum-dot cellular automata. *J Low Power Electron Appl* 8:1–15
14. D'Souza N, Atulasimha J, Bandyopadhyay S (2012) An energy-efficient bennett clocking scheme for 4-state multiferroic logic. *IEEE Trans Nano Technol* 11(2):418–425
15. Abedi D, Jaberipur G, Sangsefidi M (2015) Coplanar full adder in quantum-dot cellular autmatavia clock-zone based crossover. In: *IEEE transactions on nanotechnology, 18th CSI international symposium on computer architecture and digital systems (CADS)*
16. Waje MG, Dakhole P (2013) Design implementation of the 4-bit arithmetic logic unit using quantum-dot cellular automata. *IEEE, IACC*, pp 1022–1029
17. Timler J, Lent CS (2002) Power gain and dissipation in quantum-dot cellular automata. *J Appl Phys* 91:823–831
18. Barughi YZ et al (2017) A three-layer full adder/subtractor structure in quantum-dot cellular automata. *Int J Theor Phys* 56:2848–2858
19. Ganesh EN (2015) Power analysis of quantum cellular automata circuit. *Procedia Mater Sci* 10:381–394
20. Roy SS (2017) Generalized quantum tunneling effect and ultimate equations for switching time and cell to cell power dissipation approximation in qca devices. *Phys Tomorrow*, pp 1–12
21. Zahmatkesh M, Tabrizchi S, Mohammadyan S, Navi K, Bagherzadeh N (2019) Robust coplanar full adder based on novel inverter in quantum cellular automata. *Int J Theor Phys* 58:639–655
22. Li T, Kornovich R (2019) An Optimized design of serial-input-serial-output (SISO) and parallel-input-parallel-output (PIPO) shift registers based on quantum dot cellular automata nanotechnology. *Int J Theor Phys* 58:3684–3693
23. Jeon J-C (2020) Low-complexity QCA universal shift register design using a multiplexer and D flip-flop based on electronic correlations. *J Supercomput* 76:6438–6452
24. Yaqoob S, Ahmed S, Naz SF, Bashir S, Sharma S (2021) Design of efficient N-bit shift register using optimized D flip flop in quantum-dot cellular automata technology. *IET quantum communication*, pp 1–10
25. Verhoeven M (2016) A brief introduction to QCA, pp 2–18
26. India Documents (2020) Chapter 3 QCA Introduction, pp 24–48
27. Maharaj J, Muthurathinam S (2020) Effective RCA design using quantum-dot cellular automata. *Microprocess Microsyst* 73:1–8

A 12-bit Low-power 50MS/s SAR ADC Optimized for Speed and Power in 45 nm CMOS Technology



M. S. Akshatha and M. Nagabushanam

Abstract The paper reports a 12-bit successive approximation register (SAR)-based analog to digital converter (ADC) with minimal power consumption, better accuracy, and high speed suitable for biomedical applications. The major role in ADC conversion operation is mainly played by the feedback. It is the feedback which drives the approximation of input signal samples in an analog form to the required digital form. Also, design and performance analyses are carried out for charge scaling DAC technique such as charge redistribution and two-stage open loop comparator. SAR control logic is designed using TG-based D flip-flop (DFF) for minimum power consumption. This proposed design operates at a supply voltage of 1 V. This SAR ADC architecture is designed and optimized for power area and speed using Cadence 45 nm gpdk and simulated to extract various parameters with sampling rate of 50MS/s, and power consumed by the design is 205.9 μ W at 1 V operating supply with a delay of 249 μ s.

Keywords Successive approximation register ADC · Digital to analog converter · Charge redistribution DAC · Sample and hold circuit · Comparator

1 Introduction

In the last few decades, integrated circuit designs are intended for applications operating at low voltage, high gain, and low-power applications. Design of the analog circuits with a few innovative technologies will always be challenging to achieve low power consumption. One of the great challenges is to obtain design using minimum supply voltage. Designers are forced to develop new designs mostly applicable in integrated circuits which consume as less power as possible and at lower voltages.

The world today has been inundated with many Internet and wireless applications. There are many challenges for the design engineers to convert analog signals into digital signals that are used for applications such as biomedical applications with

M. S. Akshatha · M. Nagabushanam (✉)

Department of Electronics and Communication Engineering, Ramaiah Institute of Technology, Affiliated to Visvesvaraya Technological University, Belagavi, Bangalore, Karnataka 560054, India

e-mail: nagabushanam1971@gmail.com; nagabushanam1971@msrit.edu

the sensor platform using ADC as their main component. It becomes imperative that ADC conversion is deployed in view of the fact that most of the information available today is in analog form. The analog signals are processed with the help of an ADC circuit to obtain resultant digital signal. Binary form of outputs can be obtained from the ADC when applied input is an electric signal like voltage or current in analog form. Normally, most of the data can be seen in the analog form for real-world applications. It is utmost necessary to manipulate such data for microprocessor blocks. With the assistance of signal conversion to digital form, microprocessor will be able to understand, read, and if required, it is possible to fine-tune data.

2 Literature Survey

This section provides an understanding of different propositions proposed by many authors on the converter design in the last few years to achieve a novel design. Carandang et al. [1] developed low-power application-based 8-bit SAR ADC suitable for biomedical application. The SAR ADC design used a binary weighted capacitive array technique and dynamic latch comparator. ADC design is implemented using the synopsis tool in 90 nm CMOS technology with 1.2 V supply voltage. ADC is designed by targeting an ideal sampling rate for biomedical application with low power consumption. DelgadoRestituto et al. [2] presented SAR ADC architecture designed for biomedical applications. For implementation purpose, 180 nm CMOS split technology is used. It consists of offset cancelation technique for its time domain comparator design architecture having advantage of low switching energy consumption. Krishna et al. [3] came up with an energy-efficient SAR ADC for biomedical applications. To minimize consumed energy during ADC operation, charge redistribution technique is used. Also, comparator offset cancelation is achieved through auto-zero technique. SAR ADC is beneficial compared to conventional design because of the low-power ability in its individual blocks such as latched comparator, DAC, and effective digital circuitry. ADC is designed to satisfy minimum total power for 8-bit resolution. Singh et al. [4] came up with a work related to high speed, power efficient SAR ADC design using 90 nm CMOS technology. Work is based on voltage mode R-2R DAC. Inverters with resistors are used in this case. Since DAC used is designed without Op-Amp, minimum kickback noise can be observed. Bobadeet al. [5] proposed SAR ADC with certain modification in its DAC design. Generally, power dissipation is high in DAC module of ADC. Hence, the purpose of modification in DAC architecture is to obtain power efficiency for a design of moderate resolution. This technique is designed on the basis of split capacitive-based DAC array. Architecture consists of inverters and capacitors with series-based connection. The main purpose is to use a minimum number of capacitors, which reduces overall power consumption to achieve excellent DAC design for ADC. Authors, Diana et al. [6] came up with the design of an 8-bit SAR ADC architecture which is more suitable for medical applications. DAC design is obtained from charge redistribution concept to make possible ADC operation in low power. Dummy switch-based S/H

design is used. Such type of DAC technique is better applicable to meet biomedical requirements. Faster operation is achieved by using bootstrap switch. Sharuddin et al. [7] proposed an SAR ADC with ultra-low power capability by focusing on area efficiency. The authors have designed a circuit using a 4-bit coded thermometer in an array of 6-bit C-2C DAC. For ultra-low power performance parameter, DAC architecture is implemented with a 4-bit thermometer coded and a 6-bit C-2C array design. Technique is responsible for reduced switching activity with a minimum number of capacitive arrays. Also, technique results in better linearity. Nazari et al. [8] came up with new DAC array using split capacitive technique with an advantage of reduced capacitive area. DAC is designed for resolution of 8-bits. By using this DAC architecture, reduction in chip area and improved speed are possible. Kulkarni et al. [9] came up with a design and simulation of charge scaling DAC using array applicable for 6-bit resolutions. Capacitors array in binary weighted fashion, more suitable for high-resolution applications, are used. Switch design for DAC is implemented with 2:1 multiplexer to select between reference voltage and ground signals. Such DAC designs are helpful for obtaining implementation of ADC with minimum capacitors. Mehta et al. [10] designed 6-bit SAR ADC in 45 nm technology. Design uses tanner tool for its implementation. In this design, SAR control logic is designed using combination of shift register and code register. DFFs are used to design SAR control logic. For low power purpose, transmission gate (TG)-based flip-flops circuit design is being employed. Here, transistors with a minimum size are chosen for the design which in turn reduces the overall size. Their lengths are doubled and used for obtaining improved power performance.

Invariably today, when more than one channel are in play necessitating multiplexing of input, the most preferred application for data acquisition is the SAR ADC. This is more so when the power consumption needs to be minimal; the speed, rapid; a high precision in accuracy and a good degree of reliability. SAR ADC's in embedded microcontrollers are widely used. Typically, SAR ADC's converts 8–16 bit range and have the sampling speed from kilo-samples to giga-samples per second.

Low-power SAR ADC can be designed by using different DAC techniques such as R-2RDAC, charge redistribution principle-based binary weighted capacitive array, and split capacitive DAC. Using charge redistribution DAC, it is possible to achieve better accuracy. Hence, a majority of SAR ADC architectures contains DAC design using charge redistribution technique. Basic S/H circuit using single switch and capacitor suffers from charge injection problem, thereby resulting in a small output voltage error. To overcome charge injection problem, different S/H circuits can be designed depending on applications. Comparative analysis indicates that a dummy switch-based design is more suitable for biomedical applications. An efficient switching scheme is necessary to obtain low power consumption. Some of the energy-efficient SAR ADC can be designed using energy efficient comparators such as dynamic latched comparator, rail-to-rail comparator, and two-stage open loop comparator which are suitable for biomedical applications. The presence of comparator offsets the cancellation mechanism and results in improved energy efficiency as compared to conventional designs. SAR control logic can be designed using DFF's where TG-based implementation can be used to reduce power consumption.

Table 1 Summary of the literature survey

Reference journal	[1]	[2]	[4]	[5]	[6]	[7]	[8]	[11]
Process technology	90 nm	0.18 μm	90 nm	0.18 μm	0.18 μm	90 nm	0.18 μm	65 nm
Supply voltage	1.2 V	1 V	1.2 V	0.6 V	1.8 V	1 V	1.8 V	1 V
Sampling rate	100kS/s	4kS/s	250MS/s	500kS/s	80kS/s	40MS/s	40MS/s	20MS/s
Resolution	8 bit	10 bit	8 bit	10 bit	8 bit	8 bit	8 bit	12 bit
Power consumption	$\leq 10\mu$	76.2nW	0.98mW	691nW	164.97 μW	144nW	760 μW	472.2 μW

A charge redistribution capacitor DAC is proposed literally to attenuate the output voltage by distributing the charge present at each capacitor, resulting in a reduction in total power due to zero static power consumption under ideal conditions and also resulting in a reduction in switching energy (Table 1).

A two-stage comparator is proposed utilizing asynchronous SAR logic for high speed with a supply of 1 V, eliminating the need for an external clock and resulting in minimal power. The main goal of the proposed paper is to achieve higher sampling rates than 50MS/s in comparison to the literature, as well as further improvements in speed, delay, and power, to present a charge redistribution-based DAC methodology to save switching energy and increase DAC linearity in comparison to conventional architecture.

3 SAR ADC Key Building Blocks

The conceptual SAR ADC architecture is illustrated in the Fig. 1. It consists of building blocks such as S/H circuit, comparator, DAC, and SAR control logic. The output of DAC block present in SAR ADC is in the analog form. Role of comparator is mainly to quantize the signal. Digital logic circuits will be responsible for memorizing the comparator's output and that signal is given to DAC array in the form of feedback to obtain the signals in analog form for next step. Another important block in almost all ADC's is the comparator design. For the effective SAR ADC operation, the type of comparator design used also plays a vital role. The acquired output from the effective comparator can be either differential or it can also be single ended, where type of design hierarchy will decide the output signal. While quite a few permutations and combinations are possible from the available variety of designs, techniques, and methods, vis-à-vis the desired optimal system efficiency from performance metrics used, it boils down to picking the best one for the applications in mind. Also, the

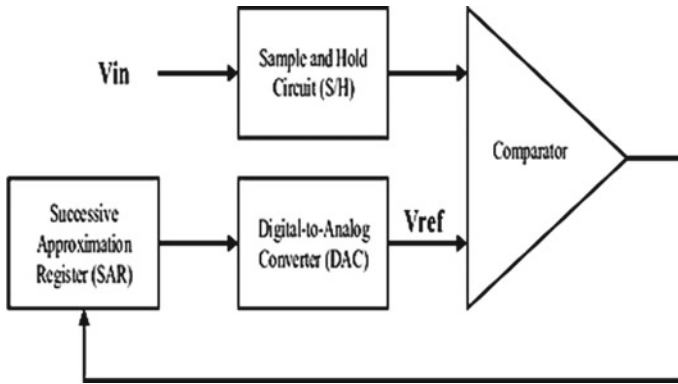


Fig. 1 Complete block diagram of SAR ADC

overall performance can be improved by taking into consideration different challenges for several parameters like managing the frequency of sampling to achieve design which gives better resolution.

A higher power consumption is also seen in DAC architecture of ADC. Individual building blocks of SAR ADC are direct functions of amount of power consumed in overall ADC design. Since capacitors are associated with switching action, each unit capacitor is responsible for calculated amount of power consumption affecting overall architecture performance. Better accuracy is required in comparator design to obtain energy-efficient SAR ADC design.

3.1 *Sample and Hold Circuit*

The simulation is carried out for two different S/H circuit design such as basic S/H design and transmission gate-based S/H design.

The basic S/H circuit is the switch and the buffer in series. The switch can be designed using MOSFET's such as NMOS, PMOS, or the combination of both. The capacitor can be used as buffer. The NMOS/PMOS-based switch is one of the simplest S/H circuits. The working of the S/H circuit involves sampling and hold duration. When the switch is in 'ON' position, the input is followed by the output. When the switch is in 'OFF' position, the electric charge is stored in capacitor. Output of S/H unit can be degraded due to certain kinds of errors. Mainly during the hold period, the errors cause difficulties in the operation than in sample mode. There are two main types of errors. They are—charge injection and the clock feed through which the requirements need to be considered in the S/H circuit.

Figure 2 depicts the schematic of a basic S/H circuit. Simulation carried out using gpdK 45 nm technology for basic S/H circuit is shown in Fig. 3. It is observed that basic S/H suffers charge injection, resulting output voltage error.

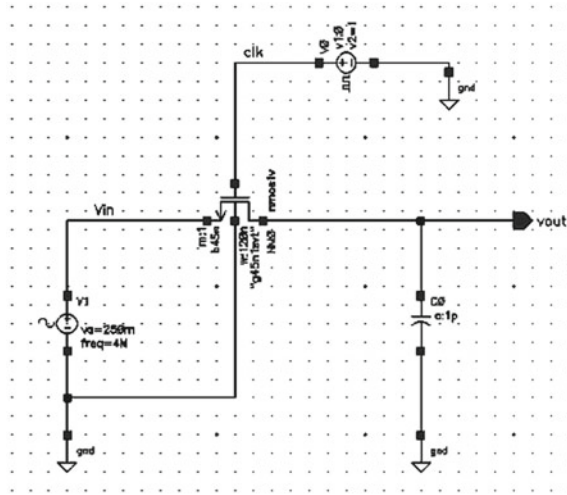


Fig. 2 Schematic of basic sample and hold circuit

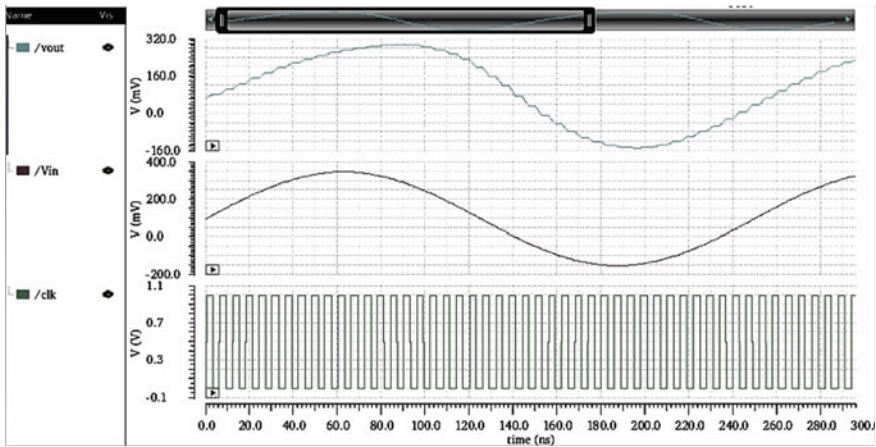


Fig. 3 Transient analysis of a simple sample and hold design

The charge injection problem can be avoided by using TG in S/H circuit. Figure 4 shows TG-based S/H circuit. NMOS and PMOS both are used to get to form TG with complementary clocks driving their gates which provide the relatively constant overall resistance. Hence, the linearity is increased (Fig. 5).

Linearity and speed of the S/H circuit rely on the ON resistance (R_{on}) and capacitance. For NMOS sampling switch, the R_{on} is given in Eq. (1).

$$R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})} \tag{1}$$

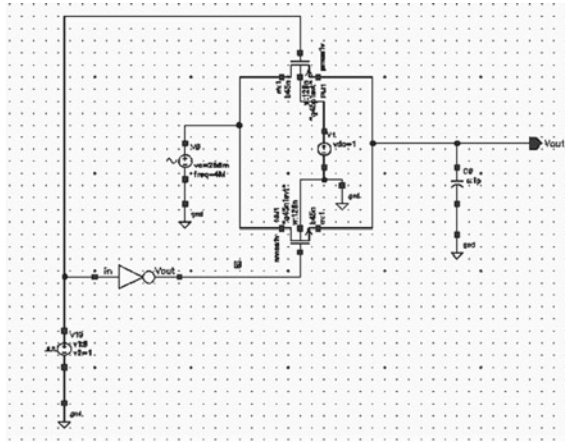


Fig. 4 Schematic of TG-based sample and hold circuit

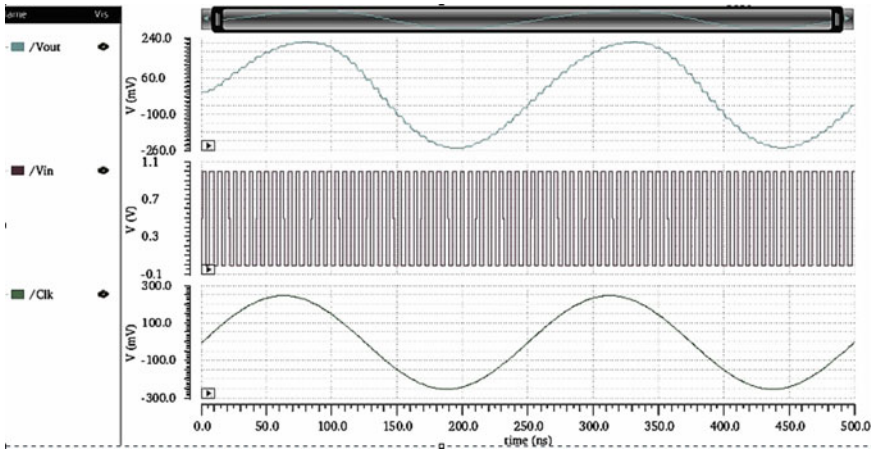


Fig. 5 Transient analysis of TG-based sample and hold circuit

where

$$V_{GS} = V_{DD} - V_{in} \tag{2}$$

Sampling capacitor C_H is chosen to cancel the amount of output noise present at the sampling node. Output noise considered is thermal noise given by the Eq. (3).

$$\text{Output noise} = \sqrt{\frac{KT}{C_H}} \tag{3}$$

where

- K is the Boltzmann constant
- T is the temperature
- C_H is the sampling capacitor.

3.2 Comparator Circuit

Two-stage open loop-based comparator design is done by focusing on various design parameters which is shown in Fig. 6. The two stages of comparator are differential stage followed by common source amplifier. Aspect ratios of each MOSFET in the comparator are calculated and are selected to obtain the required specifications such as gain requirement and low power consumption. This circuit is used as comparator in ADC where inputs for V_{in+} and V_{in-} are given from the sampled analog input signal V_{in} and DAC output signal V_{ref} .

The designed comparator circuit results in gain margin of -30.267 dB (-180.4650°) and phase margin of -126.37 m dB (-138.870°). The overall gain result from this comparator design is 52.06 dB. The simulation plots showing these results are shown in Figs. 7, 8, and 9.

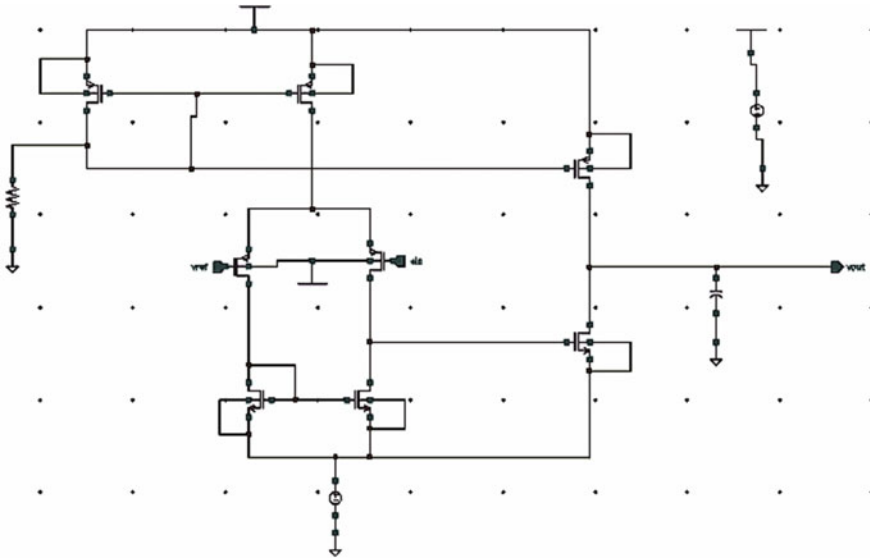


Fig. 6 Schematic of a two-stage open loop comparator

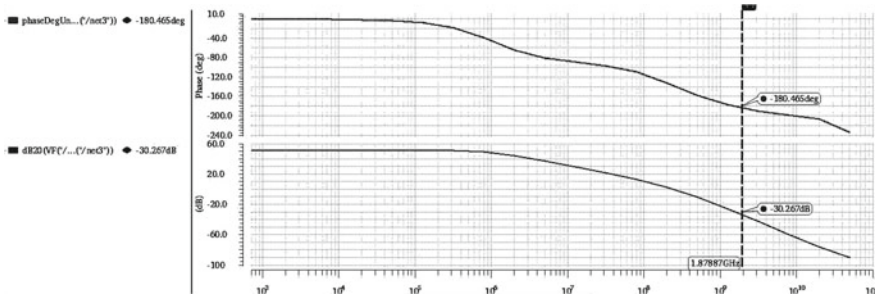


Fig. 7 Gain margin plot of comparator

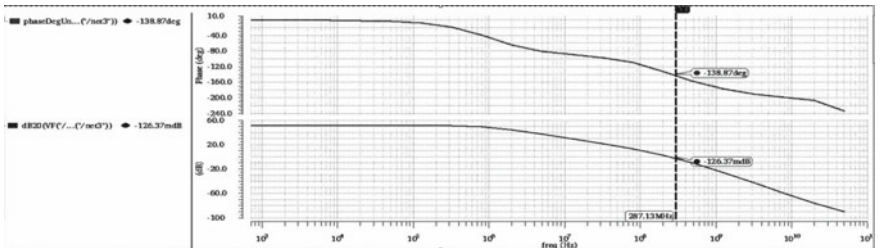


Fig. 8 Phase margin of comparator circuit

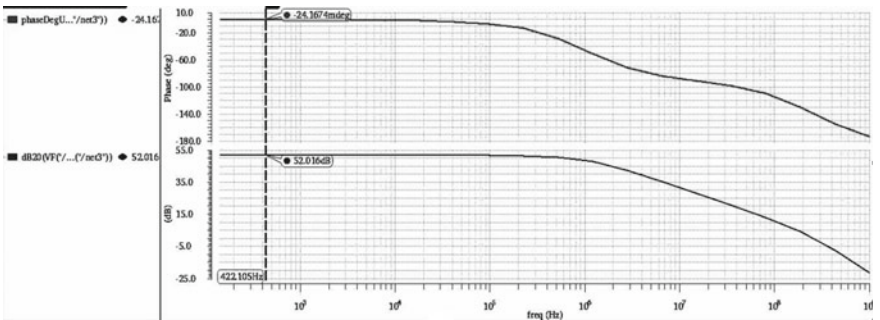


Fig. 9 Gain and phase plot of comparator circuit

3.3 SAR Control Logic Circuit

SAR logic architecture has two building blocks such as shift register and code register. The top bank of flip-flops acts as the shift registers, and the bottom bank of flip-flops acts as code registers as shown in Fig. 10. It is constructed using DFF. DFFs are constructed using TG to achieve low power using set and reset input signals. The construction of the N-bit SAR logic requires 2 N flip-flops.

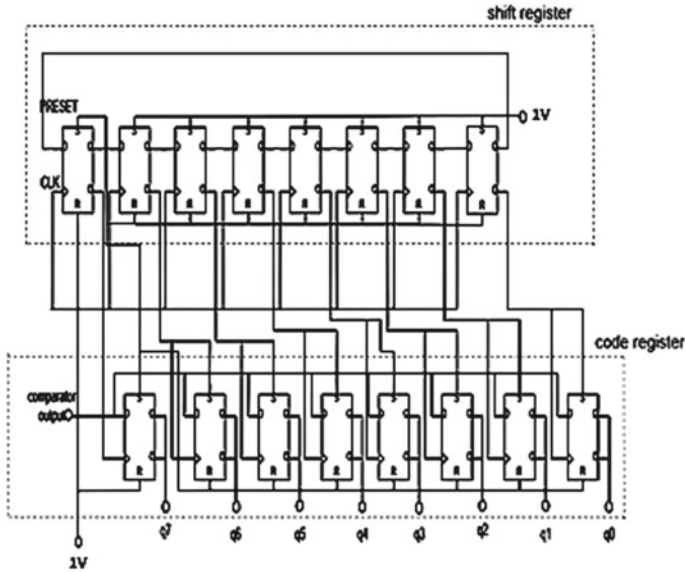


Fig. 10 General SAR logic architecture

SAR control logic working is based on binary search algorithm. A word in digital format is converted to analog using DAC by first setting the value to ‘1’ and ‘0’ for the most significant bit (MSB) and the remaining bits, respectively. The sampled input in analog form is compared with the output in analog form. Depending on comparator output, the SAR logic decides the code register’s value. If sampled analog input is greater than DAC output, the MSB bit is retained ‘1’ and next MSB bit is set to ‘1’, else MSB bit is reset, and next MSB bit is set. The succeeding bits are decided in a similar manner. This working principle is shown in Fig. 11.

3.4 Charge Redistribution DAC

The signals in digital form are converted to analog form using the DAC. The schematic diagram of the 12-bit charge redistribution DAC architecture is shown in Fig. 13. It consists of capacitors and switches. The circuit has two modes of operation. One is the reset mode, and other is the conversion mode. In reset mode, the bottom plate of the capacitors gets discharged, since all the switches are connected to ground. In conversion mode, the switch connection changes from ground to V_{ref} depending on the digital code with the reference voltage of 1 V. Equivalent analog-converted output voltages for charge distributed DAC are shown in Fig. 14 with respect to reference voltage 1 V. The analog equivalent voltage for different binary

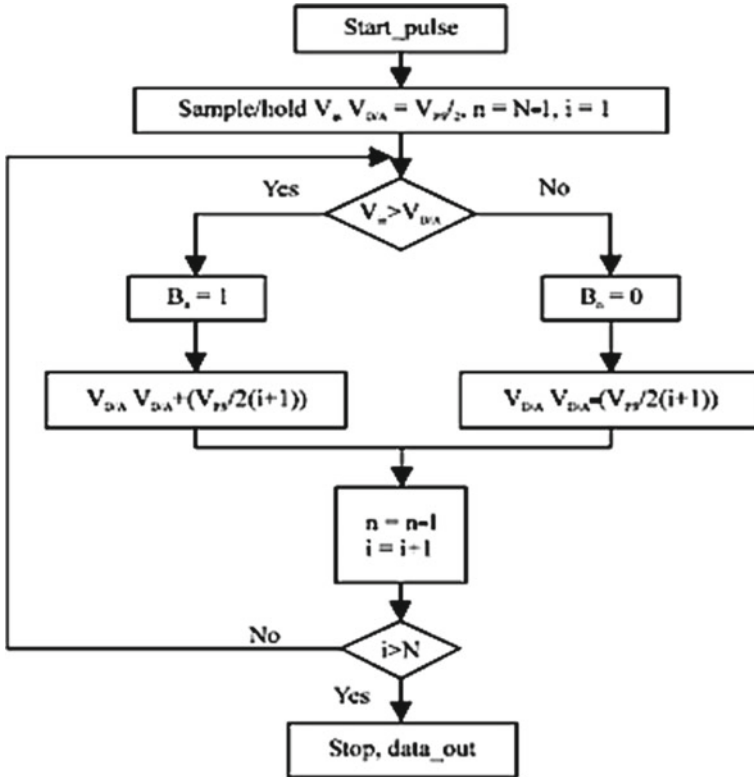


Fig. 11 Flowchart showing the working principle of SAR control logic circuit

bit combination is obtained and is observed to be equal in comparison with the theoretical voltage values (Fig. 12).

The analog output voltage for a given digital bits is calculated using equation,

$$V_{out} = V_{ref}D \tag{4}$$

V_{ref} is the reference voltage. In this case, the reference voltage is kept as 1v. D is the equivalent output voltage of the binary bits.

The equivalent output voltage of the binary bits is calculated using equation,

$$D = \frac{b_0}{2^0} + \frac{b_1}{2^1} + \frac{b_2}{2^2} + \dots + \frac{b_N}{2^N} \tag{5}$$

where

$b_0, b_1, b_2, \dots, b_N$ are the binary bits.

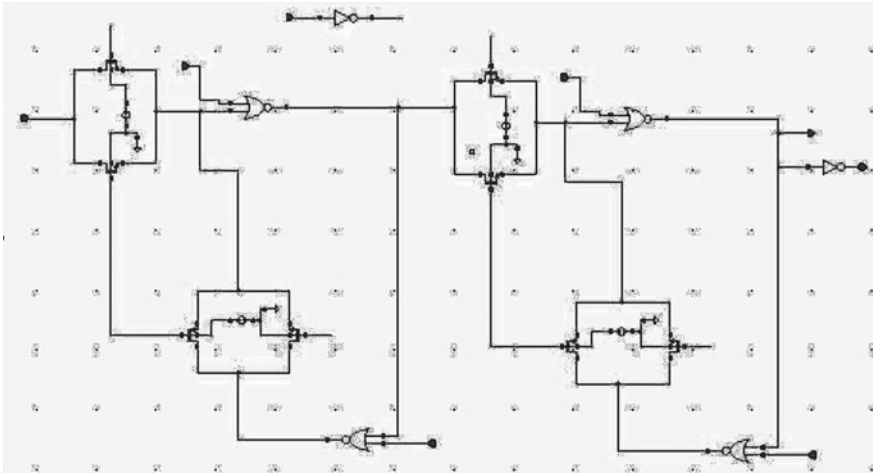


Fig. 12 Schematic of DFF employed in SAR control logic circuit

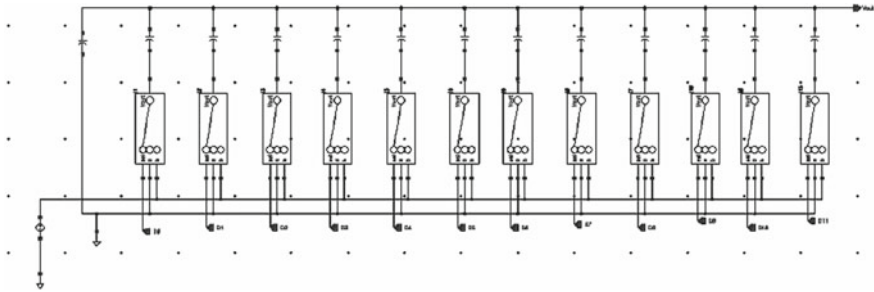


Fig. 13 Schematic of a 12-bit charge redistribution DAC

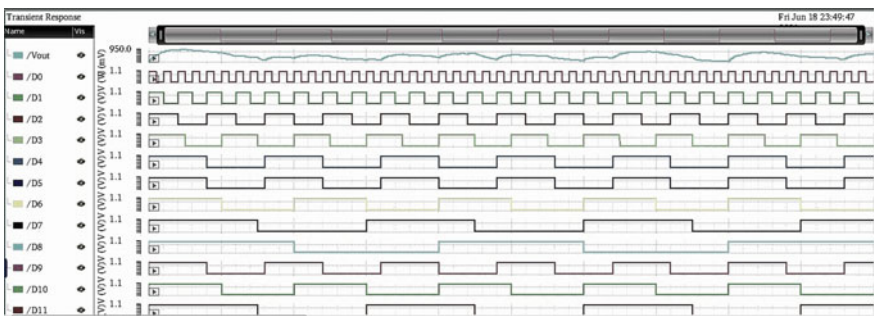


Fig. 14 Transient analysis of a 12-bit capacitive DAC

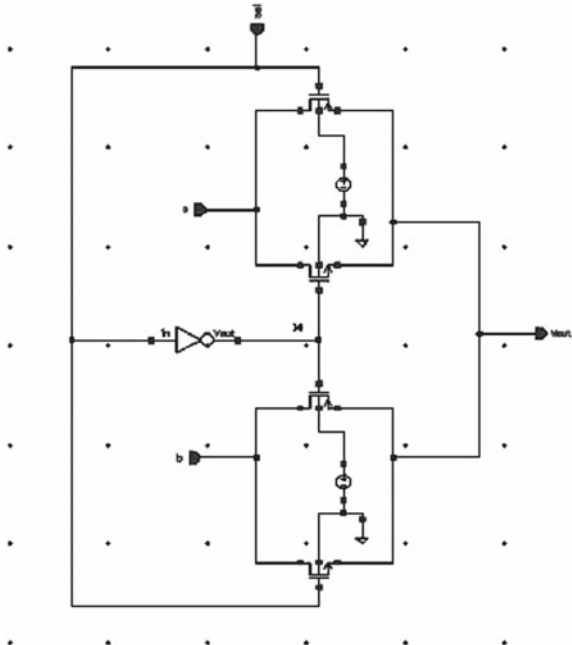
For a 12-bit DAC,

$$D = \frac{b_0}{2} + \frac{b_1}{4} + \frac{b_2}{8} + \frac{b_3}{16} + \frac{b_4}{32} + \frac{b_5}{64} + \frac{b_6}{128} + \frac{b_7}{256} + \frac{b_8}{512} + \frac{b_9}{1024} + \frac{b_{10}}{2048} + \frac{b_{11}}{4096} \tag{6}$$

3.5 Switch Design

The switch circuit is made up of an NMOS and a PMOS, operating as a 2:1 multiplexer, which is simple to implement in CMOS technology. It is because switch requires only two TG's and one inverter. One among two TG-based switches is activated at a time using selection signal (SEL). When signal SEL is high, input 'a' is connected to the output Vout. When SEL is low, 'b' is connected to the output. Figure 15 shows the schematic of 2:1 multiplexer circuit which is operating as a switch. The switch circuit is being designed with all the MOSFETs having aspect ratio of 2.6.

Fig. 15 Schematic of 2:1 multiplexer as a switch circuit



4 Integration of 12bit SAR ADC

The complete 12-bit SAR ADC design is implemented with all the submodules as discussed in the previous sections, and by aggregation of all the submodules a complete 12-bit SAR ADC is constructed. The complete VLSI design is implemented in 45 nm CMOS technology as shown in Fig. 16. EDA tools are used for the schematic entry, simulation environment setup, and functional verification. The transient analysis of the final integrated 12-bit SAR ADC circuit is shown in Fig. 17.

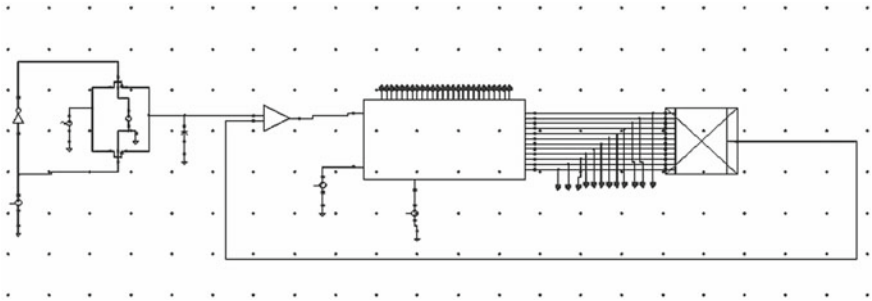


Fig. 16 Complete schematic of 12-bit SAR ADC

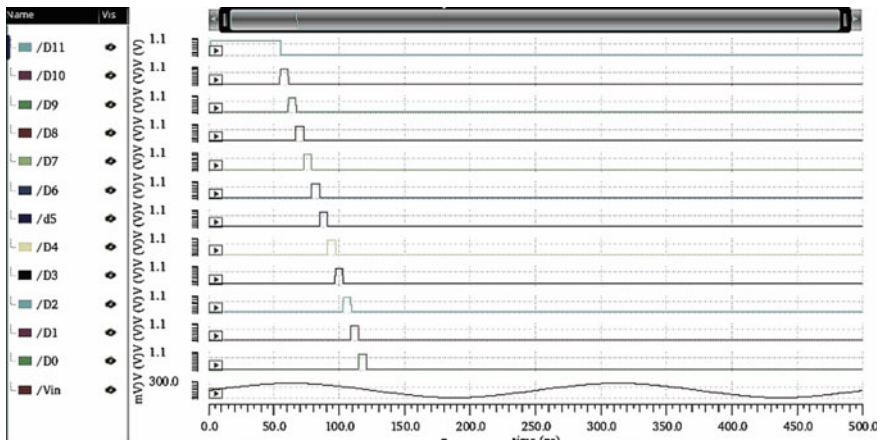


Fig. 17 Transient response of 12-bit SAR ADC

Table 2 Comparison with other works

	This work	[11]	[12]	[13]	[14]
Process technology	45 nm	65 nm	180 nm	180 nm	0.13 μm
Resolution (bits)	12	12	12	12	12
Sampling rate (S/s)	50 M	20 M	20 M	10 M	0.83 M
Supply voltage (V)	1	1	1.5	1.8	3.3
Power consumption (W)	205.9 μ	472.2 μ	1220 μ	820 μ	3200 μ

5 Conclusion

A 12-bit asynchronous SAR ADC with improved linearity and reduced noise is proposed. To achieve essential parameters such as slew rate and low power consumption, a two-stage comparator and TG-based sampling are employed with PMOS input drivers. The proposed ADC is simulated on a 45 nm CMOS chip with a 1 V supply, and the results show that it outperforms previous techniques. Finally, with a low power of 3200 μW , the ADC achieves a sampling rate of 50MS/s. The capacitor-efficient DAC's switching implies that the proposed ADC is monotonic with linear characteristics for a high resolution of 12 bits; the performance analysis is carried out by focusing on low power and accuracy. Based on these findings, the proposed ADC is suitable for high-speed real-time biomedical applications (Table 2).

References

1. Carandang J, Fortu J, Mojares J, Santos A (2017) Development of a lowpower8-bit successive approximation register adc in 90nm process technology for biomedical application. In TENCON IEEE Region 10 Conference, pp 1315–1320
2. Delgado-Restituto M, Carrasco-Robles M, Fiorelli R, Gines-Arteaga AJ, Rodriguez-Vzquez A (2016) A 76nW, 4kS/s 10-bit SAR ADC with offset cancellation for biomedical applications. In IEEE Asia Pacific conference on circuits and systems (APCCAS), pp421–424
3. Suneel Krishna RVNR, Shankar A (2013) Design of low power SAR-ADC in 0.18 μm mixed-mode CMOS process. *Int J Res Comput Commun Technol* 2(8): 654–658.
4. Singh VP, Sharma GK, Shukla A (2017) 8-Bit high speed, power efficient SAR ADC designed in 90 nm CMOS technology. IN: International conference on computing, communication and networking technologies (ICCCNT), pp 1–5
5. Bobade N, Mahendra, Gaikwad A, Dhande JD (2014) Design of successive approximation analog to digital converter with modified DAC. *Int J Eng Res Technol (IJERT)* 3(7)
6. Babita Diana K, Lackulinemoni D (2014) An Ultra low power 8-bit SAR ADC suitable for wireless medical applications. In: International conference on communication and signal processing, pp 1969–1973
7. Sharuddin IB, Lee L (2014) An ultra-low power and area efficient 10-bit digital to analog converter architecture. In: IEEE international conference on semiconductor electronics (ICSE2014), pp 305–308
8. Nazari M, Aghajani A, Hashemipour O (2015) Design of a new split-capacitive-array DAC based on distribution of attenuation capacitor. In: Iranian conference on electrical engineering, pp 1370–1373

9. Kulkarni M, Shingadi M, Kulkarni GH (2014) 6-bit charge scaling DAC and SAR ADC. *Int J Adv Res Electric Electron Instrument Eng* 3(12)
10. Mehta, NA, Nagpara BH (2014) Design and implementation of 6-bit successive approximation register (SAR) analog to digital convertor (ADC) using 45nm technology. *Int J Sci Res Develop (IJSRD)*
11. Shehzad K, Verma D, Khan D, Ain QU, Basim M, Kim SJ, Rikan BS, Pu YG, Hwang KC, Yang Y, Lee KY (2021) A low-power 12-Bit 20 MS/s asynchronously controlled SARADC for WAVE ITS sensor based applications. *Sensors*
12. Chung YH, Zeng QF, Lin YS (2019) A 12-bit SAR ADC With a DAC-configurable window switching scheme. In: *IEEE transactions on circuits and systems—I: regular papers*
13. Liu S, Shen Y, Zhu Z, A 12-Bit 10 MS/s SAR ADC with high linearity and energy-efficient switching. In *IEEE transactions on circuits and systems—I: regular papers*
14. Zhou H, Gui X, Gao P Design of a 12-bit 0.83 MS/s SAR ADC for an IPMI SoC, 978–1–4673–9094–1/15/\$31.00 ©2015 IEEE

Decision Tree-Based Classification of sEMG and Accelerometer Data of Sign Language



Akhtar Ismail Nadaf and Sanjay A. Pardeshi

Abstract Sign language is a worldwide gesture-based technique for communication in mute and hearing-impaired communities. It is rich in its own vocabulary and grammar. The ideas, emotions and information are conveyed only through physical and non-physical gestures of hands, facial muscular movements, and eye's movement. Sign language recognition is broadly classified as vision-based and sensor-based techniques. However, vision-based technique is having issues like low illumination, bad camera setting, etc. We prefer here sensor-based technique for sign word prediction by using decision tree algorithm. In this study, we are using surface electromyography and accelerometer data of both fore arms of subject for classification and prediction of isolated sign words; this database is freely available. We have used decision tree classifier and found accuracy of 95% with k-fold validation. Sign language recognition will have grate impact on hearing-impaired community for expressing their views, emotions, thoughts with hearing individuals.

Keywords Decision tree · Machine learning (ML) · Supervised learning · Support vector machine (SVM)

1 Introduction

All over globe, sign language is an ocular technique for communication in hearing-impaired and mute community. These communities use hand gestures and signs, facial expressions, leap movements, and eye movement for communication of information. The sign language dose not rely on grammar like other spoken language, rather it has its own technique for passing information. World Health Organization (WHO) claims more than 466 million people suffering from hearing loss. Birth complication, noise exposure, medication and infection are major reasons behind this disability [1]. Universally, sign language is not unique; it varies according to culture. As well as it varies according to single-handed and two-handed sign languages.

A. I. Nadaf (✉) · S. A. Pardeshi
Shivaji University Kolhapur, Kolhapur, India
e-mail: akhtarnadaf@gmail.com

Worldwide, more than 300 types of sign languages are used. Generally, series of single-hand and two-hand gestures constitute for formation of statement. It includes numbers, alphabets, and isolated words. Predominantly, numbers and alphabets are represented by finger gestures, whereas isolated words are encoded in sign language by specific hand gestures.

Hearing individuals face tremendous issues while communicating with disabled person. This requires a sign language translator for smooth communication. On other hand, information and communication technology has made remarkable advancement. So here, we have a room for development of technologies for this barrier. Recently, many researchers have worked on vision-based sign language recognition technology and sensor-based sign language recognition technology. However, vision-based technology may have limitations such as camera position, brightness, and complexity of system. Sensor-based technology includes glove based systems, surface electromyography (sEMG) systems, and inertial sensor-based systems and overcomes limitations of vision-based system [1].

In current scenario, artificial intelligence (AI) and machine learning (ML) have influenced every field. In addition, AI and ML had revolutionized many technologies. It has opened doors for new challenges. In this paper, we propose decision tree-based classifier for freely available two-handed sEMG and accelerometer signal dataset of 15 isolated words [2]. The applied algorithm shows 93% accuracy. Objective of this study is to create a machine learning model, which is capable for prediction and classification of sEMG data for American sign language.

2 Related Work

Sign language recognition is broadly classified as vision-based and sensor-based technologies. Here we summarize contribution by various researchers for sensor-based sign language recognition. Sensor-based sign language recognition comprises use of surface electromyography sensors mounted on fore arm with combination of accelerometer.

Yang et al. applied sliding window on eight-channel surface electromyography signals obtained by Myo Arm band on fore arm. Author has applied feed-forward artificial neural network (ANN) classifier [2] and had achieved accuracy of 96%. Author has limited its research up to five hand gestures of single fore arm.

Shigeyuki Tateno et al. proposed a long short-term memory (LSTM) neural network [3] as classifier. Author experimented in two steps by capturing sEMG data through Myo Arm band placed on fore arm of subject. In first step, experiment was carried out with 20 sign hand motions on single subject. Further, 20 healthy subjects were assigned with same task, and system was examined. In this study, continuous sign statements are considered for recognition. Author claims to achieve 99.7% accuracy.

In another study, Zhang et al. developed a framework for three-axis accelerometer and three-channel sEMG data of fore arm. This framework was developed for

72 Chinese sign language words and 40 sentences [4]. Results provide average accuracies of 95.3 and 96.3% for two subjects for 72 CSL words.

Wu et al. compared four popular classifiers—Naive Bayes, decision tree, nearest neighbor, and LibSVM—with two-handed 80 American signs commonly used in daily conversation [5]. Average accuracies of 93.68%, 83.85%, 96.11%, and 97.89% are achieved, respectively, through experimentation. Author has performed experiment on ten American isolated sign words.

Divya et al. [6] proposed a single-hand Indian sign language recognition (ISLR) for five public speaking words like “Please”, “Sorry”, “Good Night”, “Good Morning”, and “Hello”. Author captured sEMG signals for these words by placing two sEMG electrodes on flexor carpi radialis and carpi radialis longus and reference electrode in the palm of the right hand. These signals were acquired by using BIOPAC MP-45. Further features like mean absolute value (MAV), standard deviation (STD), and root mean square (RMS) are extracted, and a support vector machine classifier is used for classification. In this methodology, authors have achieved accuracy of 90%.

In another study, Suri et al. [7] proposed sequential master–slave architecture. The performance of master–slave architecture is improved by additional features of long short-term memory network (LSTM). This study elaborates 14% improvement in convolutional neural network and 9% improvement in master–slave network. An accuracy of 93.5% was achieved through this experimentation.

Sharma et al. in [8] employed multi-modality data acquisition for better classification accuracy. The study compares the findings of sensors such as an accelerometer–gyroscope, a sEMG–accelerometer, a sEMG–gyroscope, and a sEMG–accelerometer–gyroscope. The use of sEMG sensors in conjunction with an accelerometer and gyroscope improves signal categorization significantly, according to the author.

Gupta et al. [9] comparatively analyzed continuous Indian sign language classification with two CNN classifier models, viz. MC-DCNN and t-LeNet. The signals were captured with two IMU units placed on the fore arm of the subject and recorded with Delsys Trigno wireless digital system. Author performed an experiment with eleven sentences constructed from 15 commonly used words in ISL. Ten different people signed these phrases ten times. They also put MC-DCNN and t-LeNet to the test. Finally, it was discovered that MC-DCNN has an accuracy of 83.94% and t-LeNet has an accuracy of 79.70%.

This literature does not provide evidence specifically, for decision tree classifier for eight-channel surface electromyography and accelerometer signals of both hands.

3 Methodology

3.1 Database Description

The database used for this study is freely available at <https://data.mendeley.com/datasets/wgswcr8z24/2>. Researcher has collected sEMG and accelerometer data from Myo Arm band. This dataset comprises. CVS files of ten American sign language

signs SL and 15 health signs. Ten American sign language signs consist of 20 signs collected from 10 healthy subjects. On other hand, 15 health signs consist of 15 health signs collected from five users. All signals are collected by wearing Myo Arm band on both hands [10]. This data consists of eight-channel sEMG data and nine-axis accelerometer data.

3.2 Experimentation

3.2.1 Selection of Machine Learning Platform

We chose Google Colab as the machine learning platform for this study because it is free and allows you to develop and execute Python code, import other databases, and interface with PyTorch, TensorFlow, Keras, and OpenCV. It also offers a free cloud service with a free GPU.

3.2.2 Preprocessing

For this study, we have used 10 American sign language data. All CSV files of sEMG signals are combined to form a master file. This file contains samples of eight sEMG signals of each hand and nine accelerometer samples of each hand along with associated label of American sign. This data is checked for cleaning using heat map. It is found that it does not require any cleaning as it is not containing any null value (Fig. 1).

3.2.3 Exploratory Data Analysis

To understand pair-wise relationship between dataset, pair plot is plotted. Pair plot is useful for finding correlation in dataset. Following figure shows relationship of four attributes of dataset. Remaining are explored in result sections. It shows that sEMG signals from channels are related to each other, but when we study for higher channels, that are distinct (Fig. 2).

3.2.4 Selection Machine Learning Algorithm

Surface electromyography is a time series data [10]. The database for this experimentation is time-series, labeled data. Hence for this study, we are focusing decision tree algorithm, as it is supervised machine learning algorithm, and it can be used for classification and regression. It uses tree representation for solving problem. Goal

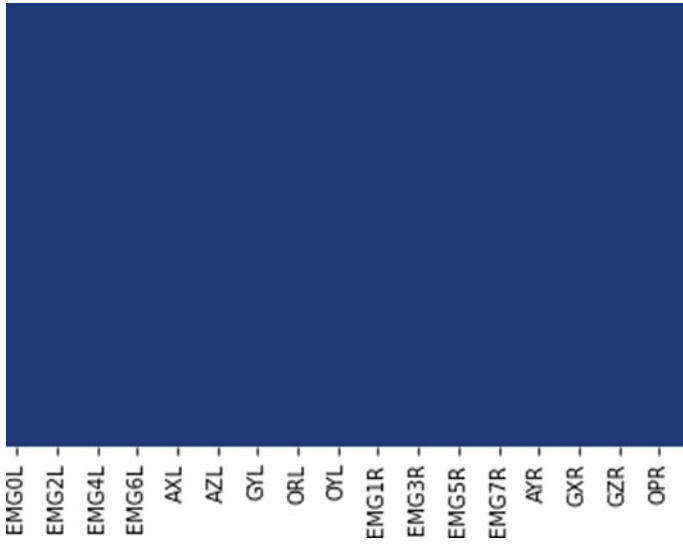


Fig. 1 Heat map of dataset

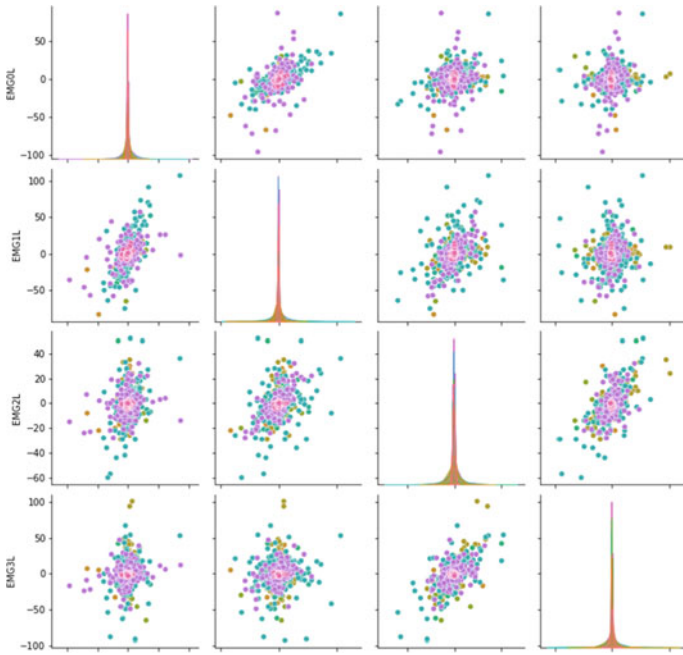


Fig. 2 Pair plot of four attributes of left-hand sEMG signal of four channels

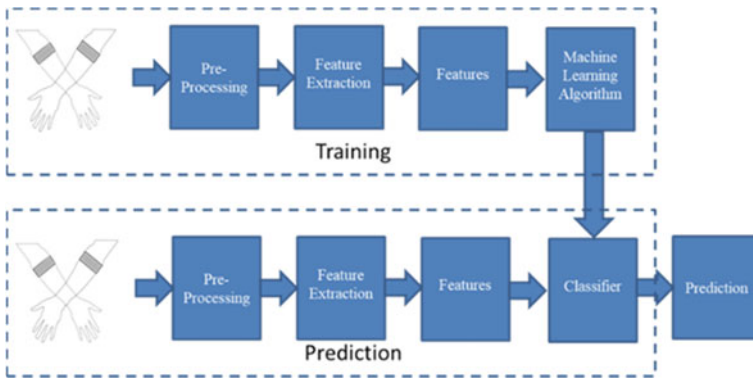


Fig. 3 Research method design

of this algorithm is to create a model that predicts sign based on its learning from training dataset, which will have features and associated labels.

The sEMG and accelerometer data visualized in pair plot (Fig. 3) shows that data points are complex and overlapped, as well as there is not any distinct and separate data group. For this purpose, decision boundary cannot be found by linear regression. Hence, data needs to split again and again with decision boundaries, and this can be achieved by decision tree.

3.2.5 Gini Impurities

This is another important term while dealing with decision tree. It indicates impurity in dataset. Its purpose is to calculate purity of data split in decision tree.

$$G.I. = 1 - \sum_{i=1}^n p^2$$

$$G.I. = 1 - \sum_{i=1}^n [p_+^2 + p_-^2]$$

Gini impurity is computationally efficient and requires shorter time for execution.

3.2.6 Setting Test Data

In this section, there are two steps:

(A) Cross-validation

The attainment of machine learning models is determined by cross-validation technique. It is used to produce model predictions and measure accuracy of predictive model once it has been deployed. In prediction model, dataset is divided into two sets

- (1) Training dataset
- (2) Testing dataset.

Training dataset is used to train the model. On other hand unknown dataset (test dataset) is used for testing of model.

(B) Data split

The dataset consists of 2700 instances; hence, for a study we have used 80% training data and 20% testing data. Therefore, that performance statistics will not have greater variance (Figs. 4 and 5).

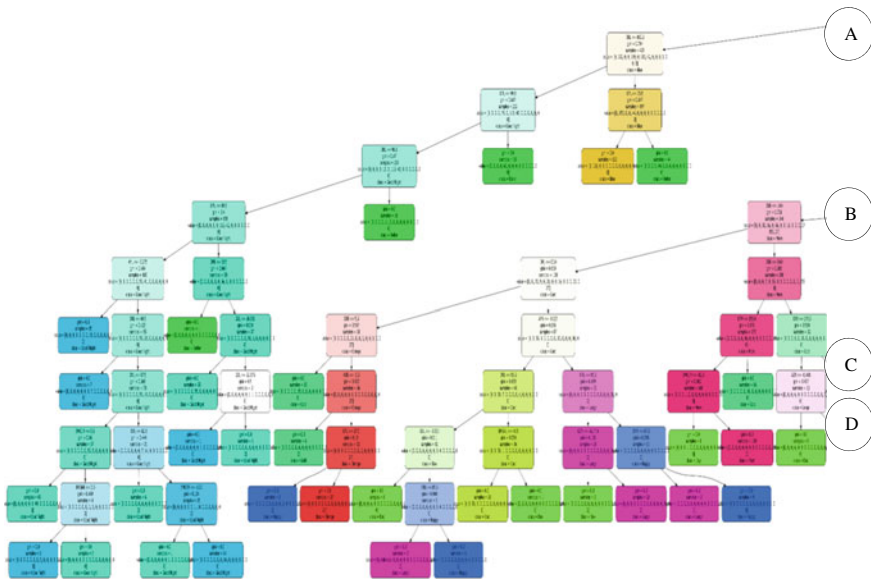


Fig. 4 Decision tree part A

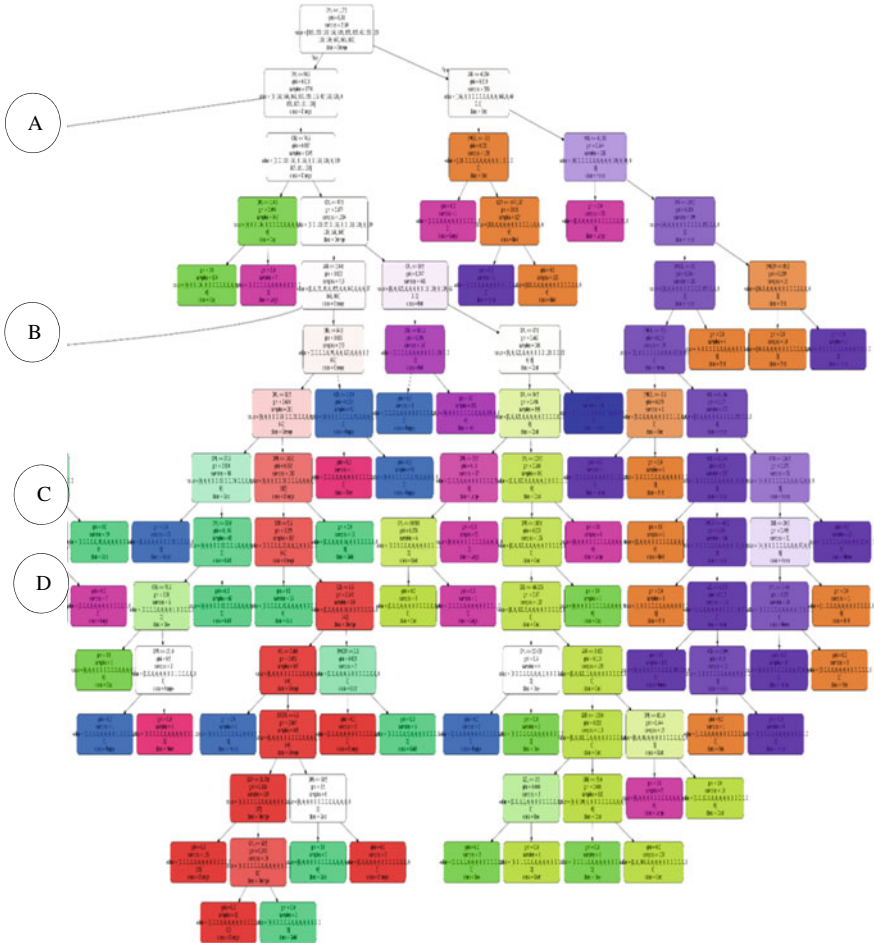


Fig. 5 Decision tree part B

4 Result and Discussion

4.1 Classification Report

See Table 1.

From the classification report, we infer following points.

- (1) The precision shows percentage of sign word prediction with higher values ranging from 85 to 100%.
- (2) Recall indicates correctly identified positive fractions many of with average of 96%.

Table 1 Classification report

Classification report				
Sign words	Precision	Recall	F1-score	Support
Bird	0.85	0.91	0.88	32
Blue	1.00	1.00	1.00	34
Cost	0.96	1.00	0.98	47
Day	0.97	0.94	0.95	33
Doller	1.00	1.00	1.00	41
Gold	0.98	0.91	0.95	47
Good Night	1.00	0.86	0.92	28
Happy	1.00	0.96	0.98	51
Home	1.00	1.00	1.00	29
Horse	0.92	0.90	0.91	40
Hot	1.00	1.00	1.00	41
Large	0.98	0.98	0.98	43
Mom	0.97	0.97	0.97	34
Orange	0.91	1.00	0.95	30
Accuracy			0.96	540
Macro avg	0.95	0.96	0.95	540
Weighted avg	0.96	0.96	0.96	540

- (3) F1-score shows what percentage of positive prediction is correct and its average accuracy is 96%.

4.2 Confusion Matrix

The confusion matrix is summary of prediction result of classification problem. The rows indicate predicted class, and columns indicate actual class (Fig. 6).

5 Conclusion and Future Scope

From above experiment, we conclude that decision tree classifier is robust technique to accurately classify isolated sign words encoded by eight-channel sEMG signals and accelerometer. Through this, we have achieved accuracy of 96% and in future, it can be used for classification of continuous statements and finger spelled words accurately. In future, we decide to use other machine learning model and test performance. In addition, it will be utilized for Indian sign language.

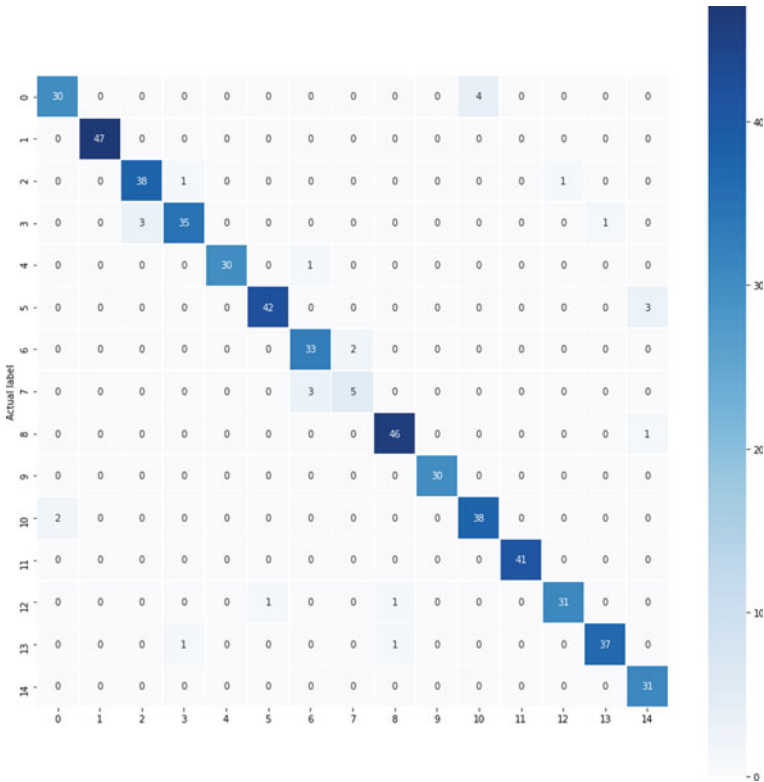


Fig. 6 Confusion matrix

References

1. Kudrinko K, Flavin E, Zhu X, Li Q (2020) Wearable sensor-based sign language recognition : a comprehensive review 3333. <https://doi.org/10.1109/RBME.2020.3019769>
2. Yang K, Zhang Z (2019) Real-time pattern recognition for hand gesture based on ann and surface emg, 799–802
3. Signal E, Tateno S, Liu H, Ou J (2020) Development of sign language motion recognition system for hearing-impaired people using
4. Zhang X, Chen X, Member A, Li Y, Lantz V, Wang K, Yang J (2011) A framework for hand gesture recognition based on accelerometer and emg sensors 41:1064–1076
5. Wu J, Sun L, Jafari R (2016) A wearable system for recognizing american sign language in real-time using imu and surface emg sensors 20:1281–1290
6. Divya B, Delpha J, Badrinath S (2018) Public speaking words (Indian sign language) recognition using EMG. In Proceedings 2017 international conference smart technology smart nation, SmartTechCon 2017, pp 798–800. <https://doi.org/10.1109/SmartTechCon.2017.8358482>
7. Suri K, Gupta R (2018) Transfer learning for sEMG-based hand gesture classification using deep learning in a master- slave architecture. In Proceedings 3rd international conference contemporary computing informatics, IC3I 2018, pp 178–183. <https://doi.org/10.1109/IC3I44.769.2018.9007304>

8. Sharma S, Gupta R, Kumar A (2019) On the use of multi-modal sensing in sign language classification. In 2019 6th international conference signal processing integrated networks, SPIN 2019, pp 495–500. <https://doi.org/10.1109/SPIN.2019.8711702>.
9. Gupta R, Rajan S (2020) Comparative analysis of convolution neural network models for continuous indian sign language classification. *Procedia Comput Sci* 171:1542–1550. <https://doi.org/10.1016/j.procs.2020.04.165>
10. Ma F, Song F, Liu Y, Niu J (2020) SEMG-based neural network prediction model selection of gesture fatigue and dataset optimization. *Comput Intell Neurosci* 2020. <https://doi.org/10.1155/2020/8853314>

Bibliometric Analysis of Published Literature on Mobile Healthcare in the Past One Decade



Saibal Kumar Saha, Sangita Saha, and Ajeya Jha

Abstract The study aims to highlight the bibliometric analysis of published literature on mobile healthcare in the past one decade. The data consist of literature indexed in Scopus database from 2011 to 2020. The keywords used for the search were m-healthcare, m-healthcare, mobile healthcare, mobile healthcare. Microsoft Excel version 2016 and VOS Viewer have been used for data analysis. Bibliometric analysis was done for co-authorship based on countries, co-occurrence of all keywords using network visualization and overlay visualization. The co-occurrence analysis of all keywords resulted in formation of four clusters. The most impactful keywords were human, health care, m-healthcare, telemedicine and m-health. The keywords cryptography, network security, cloud computing, Internet of Things, big data, wearable sensors and mobile application are being used more often in the recent years.

Keywords Bibliometric analysis · Mobile healthcare · Electronic healthcare · VOS viewer · Citation analysis · Co-occurrence analysis · Telemedicine

1 Introduction

The promising approach of mobile healthcare in the field of medicine is realized with the help of biomedical devices connected to the cloud [1]. It is an emerging technique for clinical applications. Usage of mobile and Internet technologies directly and indirectly with healthcare systems has increased the efficiency and reliability of healthcare systems. Real-time communication and convenient healthcare services are possible due to intellectualization of mobile devices and automation of mobile healthcare networks. Mobile devices help to collect data related to patient health and upload it in the healthcare server. Doctor, patient, healthcare centers and researchers

S. K. Saha (✉) · A. Jha

Department of Management Studies, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Gangtok, India

e-mail: saibal115@gmail.com

S. Saha

Department of Management Studies, MIT University, Shillong, India

may use these data [2]. Doctors have a chance to see the medical records, history, similar cases and its treatment in the past, electronically. It helps them to take an informed decision about their patient and decide the right form of treatment. Patients on the other hand may get timely reminders for taking their medication [3], doctor appointments and other health-related information. Healthcare centers get benefited with the data as all the records are in digital format and provide good resource pool for treatment analysis and future course of action. It also helps them to interact with the patients, doctors and healthcare staff in real-time and share vital information [4]. It also facilitated monitoring in real time [5]. Researchers benefit from the data as they get historical data for analysis and designing artificial intelligence (AI)-based solutions for doctors and patients to get better and predictive treatment analysis. Storage of data, processing in real time, ease of availability of records and information in the fingertips have helped the stake holders to take informed decisions.

The paper is divided into different sections. In the first section, the top five authors with highest number of contribution, top five subject area, top five journals, top five universities in which the works related to the topic of this research were carried out and the top funders have been discussed. In the second part, analysis with respect to number of publications per year and the trend has been shown. It is followed by an analysis of citations for the years under study. The analysis has been done with the help of a citation matrix. It is followed by an analysis of yearly average citation ranking of the articles. In this analysis, the top ten results have been discussed.

2 Literature Review

A mobile healthcare system is a well administered system which helps to connect patient, physician and the healthcare system. The development of wearable gadgets, evolution of Internet of Things—IOT along with cloud-based versatile web services—has changed the dynamics of therapeutic environment [6]. The advancements made in m-healthcare have helped patients, physicians and other stakeholders to gather and assess data related to patient health, providing support for remote monitoring and efficient decision making [7].

Using electrocardiogram (ECG) signals, Qaisar and Hussain [8] presented a novel solution to detect arrhythmia. A system was developed to use samples of level-crossing-based ECG signals to compress them in real time and process it for transmission. Researchers were able to achieve a sevenfold reduction in the data amount for cloud transmission which helped to reduce usage of bandwidth and consumption of transmitter power along with reduced cloud application processing load. The study of Alam [9] presented a blockchain-based healthcare system. The objective was to reduce pressure on patients and costs associated with clinical services. The researcher proposed a blockchain-based architecture to develop a system for identification and authentication of COVID-19 suspects. The m-healthcare monitoring system of Arora et al. [10] comprised of wireless sensor network and gadget based on analog-to-digital converter (ADC). It helped to track the humidity, temperature and

heartbeat of the patients. Information is stored in cloud server using Wi-Fi module. This information is easily accessible to the users in the platform of android and iPhone operating system (IOS). The Hong Kong-based work of Lam et al. [11] focused on decline of fertility rate, increase in life expectancy and aging problem of the country. The researchers developed an Intelligent mHealthcare System (ImHS) which aimed to reduce the load on manpower by integrating technologies like face recognition and FaceAPI-based search algorithms of elderly medical records. It helped to increase the operation efficiency and effectiveness by better allocation of healthcare resources.

Trend of publications and impact of an article is best understood with the help of bibliometric analysis [12]. Publication pattern and its impact is well explained with the help of article citations, funding agencies, collaborating institutions and organizations, use of keywords, etc. [13, 14]. The VOS viewer software [15] provides a convenient platform to do the bibliometric analysis of published literature by creating analysis like text data visualization, term co-occurrence, bibliographic coupling, co-occurrence, etc. [16, 17].

3 Objective

The research aims to highlight the bibliometric analysis of published literature on mobile healthcare in the past one decade—2011 to 2020.

4 Methodology

To achieve the objective of this study, metadata of published literature which were indexed in the Scopus database was extracted for the period 2011 to 2020. The keywords used for the search were m-healthcare, mobile healthcare, mobile healthcare. Only articles which were published in journals and in English were used for the research. The search resulted in 344 documents. All the analyses were done in MS Excel version 2016. In the third section, bibliometric analysis of the metadata has been done using VOS Viewer. Analysis has been done for co-authorship based on countries, co-occurrence of all keywords using network visualization and overlay visualization. Graphs have been prepared and explained.

5 Results and Discussion

The top five authors with highest number of contribution in the field of mobile healthcare are: (i) Ahmad, M. (6), (ii) Omar, M. (6), (iii) Lin, X. (5), (iv) Sood, S.K. (5) and (v) Cao, Z. (4). The top five subject areas in which articles related to mobile healthcare have been published are: (i) Computer Science (173), (ii) Engineering

(145), (iii) Medicine (80), (iv) Biochemistry, Genetics and Molecular Biology (49) and (v) Chemistry (26). The top five journals which have published articles related to the topic of this research are: (i) IEEE Access (14), (ii) Sensors Switzerland (12), (iii) Journal of Medical Systems (8), (iv) IEEE Journal of Biomedical And Health Informatics (7) and (v) Journal of Ambient Intelligence And Humanized Computing (7). The top five universities in which the research have been carried out are: (i) King Saud University (14), (ii) Xidian University (7), (iii) University Utara Malaysia (6), (iv) Korea University (6) and (v) Huazhong University of Science and Technology (6), and the top funders are: National Natural Science Foundation of China and Ministry of Education of the People's Republic of China.

Publications Per year

Figure 1 describes the publication pattern of the topic under study in the past 10 years. The number of publications increased from 13 in the year 2011 to 52 in the year 2020 which indicates the popularity of research in the field of mobile healthcare.

10 years' citation of the data extracted from Scopus database is analyzed in Table 1. The table consists of total number of citations in a particular year (TC), total number of publications in a year (TP). TC/TP is the average number of citations in a year, and different thresholds like citations greater than 99, 49, 24, 9, 4 and 0. 89.17% of the publications were cited at least once. Nine publications received citations more than 99. Twenty-four publications received more than 49 citations. The papers published in 2015 received the highest number of citations (1016).

The yearly average citation ranking of articles is shown in Table 2. In the table, TC = total number of citations, AG = age of the article as of 2021, TC/AG = average citation per year. In the TC/AG rank analysis, the paper titled "Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring" [18] ranked 1st. The paper was published in 2016 and has received 201 citations. In the citation ranking, the paper ranks second. The paper titled "A generalized adoption model for services: A Cross-Country Comparison of Mobile Health (m-health)" [19]

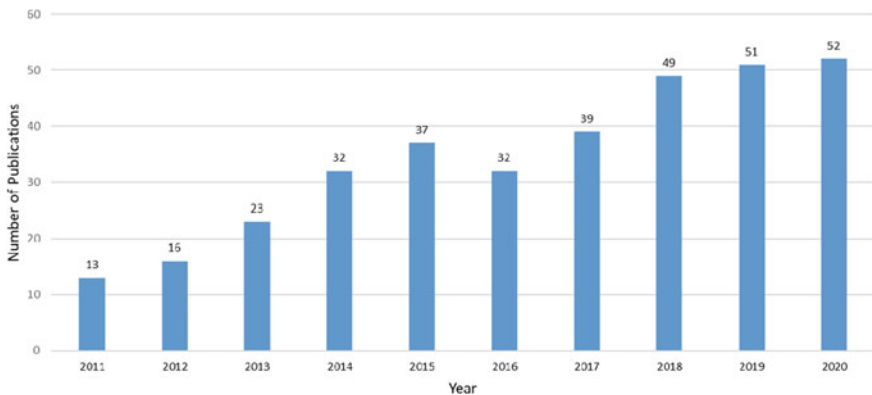


Fig. 1 Number of publications per year

Table 1 10 years' citation matrix

Year	TC	TP	TC/TP	>99	>49	>24	>9	>4	>0
2011	622	10	62.2	2	3	1	1		1
2012	189	15	12.6			3	3	7	
2013	667	17	39.23	2	2	3	3	4	3
2014	747	23	32.47		7	4	3	3	5
2015	1016	34	29.88	3	3	7	8	7	5
2016	818	27	30.29	2	3	4	5	3	7
2017	598	30	19.93		3	5	15	2	4
2018	757	45	16.82		3	6	13	10	10
2019	333	42	7.92			1	10	13	11
2020	127	34	3.73				1	12	11
Total	5874	277		9	24	34	62	61	57

ranked second in the TC/AG rank analysis. This paper was also published in 2016 and has 176 citations. It ranked 4th in the citation ranking. The most cited paper was “The adoption of mobile healthcare by hospital’s professionals: An integrative perspective” [20] published in 2011. It has received 214 citations. The paper ranked 7th in the TC/AG ranking analysis.

Bibliometric Analysis

Co-authorship based on countries

The VOS Viewer-based bibliometric analysis on the data extracted from Scopus database reveals that for co-authorship analysis based on countries out of 57 countries 18 met the threshold when the minimum documents of a country were set to 5. Table 3 gives the details of number of documents, citations and total link strength of the 18 countries that satisfied the processing criteria. Figure 2 gives a graphical representation of the analysis on co-authorship based on countries. It is found that USA and China have the highest link strength (15). Publications from China, Australia, Saudi Arabia and Malaysia are mostly of recent origin (around 2018). Malaysia have co-authorship mainly with Saudi Arabia. Saudi Arabia’s co-authorship is primarily with Malaysia, Australia, USA and China. Australia’s co-authorship is with USA, China, Canada, India, Hong Kong and Saudi Arabia. India’s co-authorship is with Taiwan, UK, China, USA and Australia. Canada’s co-authorship is with Japan, USA, UK, Australia and China.

Co-occurrence of All Keywords

By setting 10 as the criteria for minimum number of occurrence of a keyword, co-occurrence analysis of all keywords was done in VOS Viewer using full counting method. Out of 3410 keywords, 71 met the threshold. Figure 3 represents the co-occurrence of all keywords network visualization. In the figure, four clusters can be found. Cluster 1 consists of 27 items. The most prominent keywords of cluster 1 are: health care, m-health, diagnosis, cryptography and cloud computing. Cluster

Table 2 Yearly average citation ranking of articles

TC/AG rank	Citation rank	Title	Citation	Year	TC	AG	TC/AG
1	2	“Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring”	[18]	2016	201	5	40.20
2	4	“A Generalised Adoption Model for Services: A cross-Country Comparison of Mobile Health (m-health)”	[19]	2016	176	5	35.20
3	12	“Cloud and IoT-Based Disease Prediction and Diagnosis System for Health care Using Fuzzy Neural Classifier”	[21]	2018	94	3	31.33
4	5	“MHealth Adoption in Low-resource Environments: A Review of the Use of mobile healthcare in Developing Countries”	[22]	2015	158	6	26.33
5	3	“SPOC: A Secure and Privacy-preserving Opportunistic Computing Framework for Mobile Healthcare Emergency”	[23]	2013	200	8	25.00
6	6	“Fast Fabrication of Flexible Functional Circuits Based on Liquid Metal Dual-Trans Printing”	[24]	2015	150	6	25.00
7	1	“The Adoption of mobile healthcare by Hospital’s Professionals: An Integrative Perspective”	[20]	2011	214	10	21.40

(continued)

Table 2 (continued)

TC/AG rank	Citation rank	Title	Citation	Year	TC	AG	TC/AG
8	23	“Policy-Based Secure and Trustworthy Sensing for Internet of Things in Smart Cities”	[25]	2018	62	3	20.67
9	24	“A Provably Secure Cross-Domain Handshake Scheme with Symptoms-Matching for mobile healthcare Social Network”	[26]	2018	61	3	20.33
10	8	“4S: A Secure and Privacy-preserving Key Management Scheme for Cloud-assisted Wireless Body Area Network in m-healthcare Social Networks”	[27]	2015	120	6	20.00

Table 3 Co-authorship based on countries

S. No.	Country	Documents	Citations	Total link strength
1	USA	65	1340	35
2	China	75	1671	34
3	Australia	11	176	13
4	India	42	713	11
5	Saudi Arabia	22	193	10
6	UK	15	718	10
7	Canada	15	659	8
8	Hong Kong	6	186	8
9	Pakistan	6	31	5
10	Singapore	5	297	5
11	Italy	9	82	4
12	Japan	6	68	3
13	Malaysia	13	27	2
14	South Africa	7	102	2
15	South Korea	52	649	2
16	Taiwan	15	628	2
17	France	10	90	0
18	Spain	5	91	0

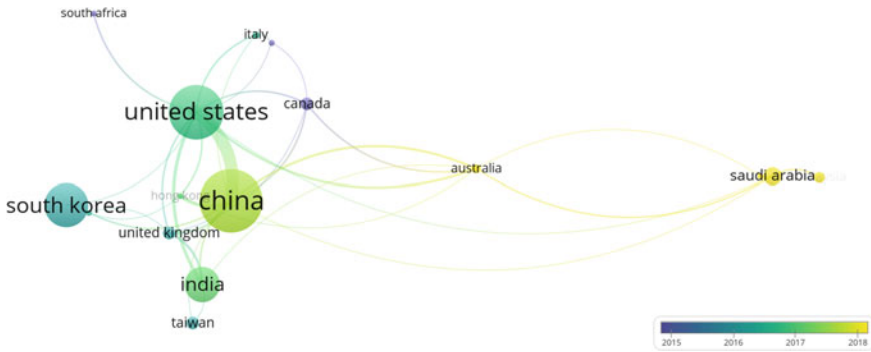


Fig. 2 Co-authorship based on countries

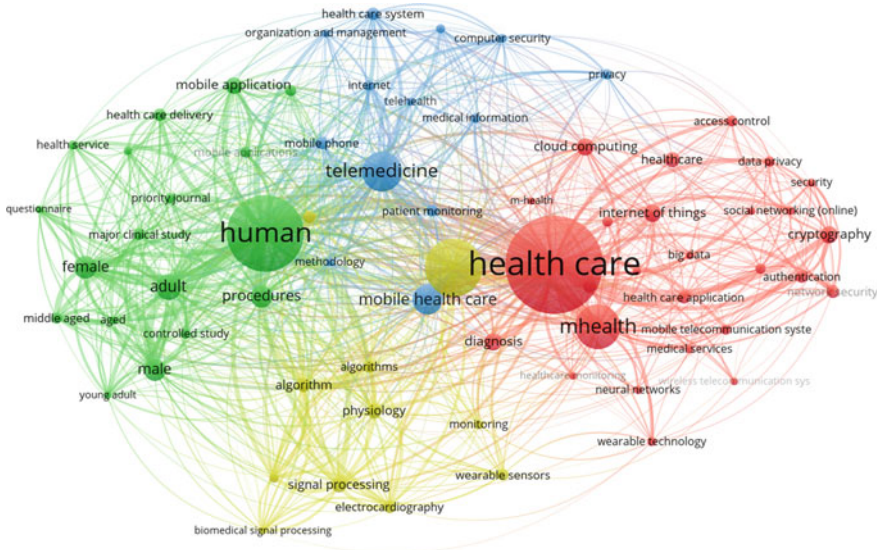


Fig. 3 Co-occurrence of all keywords network visualization

2 consists of 18 items. The prominent keywords of cluster 2 are: human, female, male, adult, procedures, aged, controlled study and questionnaire. In cluster 3, the prominent keywords are telemedicine, mobile healthcare, Internet, mobile phone, privacy, etc. A total of 13 items form this cluster. Cluster 4 consists of 11 items which include algorithm, signal processing, wearable sensors, devices, biomedical signal processing, electrocardiography etc.

The overlay visualization of Fig. 3 is displayed in Fig. 4. In this, it is found that the keywords m-health, cryptography, network security, cloud computing, Internet of Things, big data, wearable sensors, mobile application, healthcare application, authentication, etc., have been used in the recent years (2018).

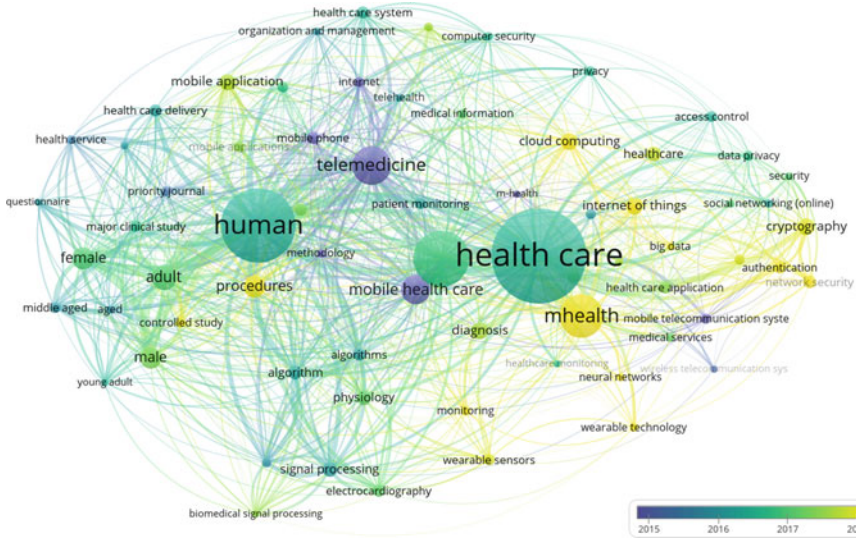


Fig. 4 Co-occurrence of all keywords overlay visualization

An analysis of health-related issues was conducted using text data analysis of title and abstract fields using full counting method. On setting 5 as the minimum number of term occurrence, 556 texts were obtained out of 9196 terms. On manual examination of each term, following 21 terms were extracted related to health issues. Table 4 shows the text analysis of major health issues, their occurrence and relevant scores.

6 Conclusion

The study aims to highlight the bibliometric analysis of published literature on mobile healthcare in the past one decade. The data consist of the literature indexed in Scopus database from 2011 to 2020. Microsoft Excel version 2016 and VOS Viewer have been used for data analysis.

The study shows that the number of publications increased from 13 in the year 2011 to 52 in the year 2020. 89.17% of the publications were cited at least once. The co-occurrence analysis of all keywords resulted in formation of four clusters. The most impactful keywords were human, health care, m-healthcare, telemicine and m-health. The keywords cryptography, network security, cloud computing, Internet of Things, big data, wearable sensors and mobile application are being used more often in the recent years. The yearly average citation ranking of articles revealed that the paper titled “Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring” [18] had the highest yearly average citation rank (40.2),

Table 4 Health issues

S. No.	Term	Occurrences	Relevance score
1	Arrhythmia	7	0.9088
2	Asthma	9	1.5985
3	Blood	9	0.5511
4	Cancer	10	2.6513
5	Cardiovascular disease	8	0.9526
6	Chronic disease	13	0.6535
7	Diabetes	19	1.7175
8	Disability	14	2.1735
9	ECC	6	1.3428
10	ECG	24	0.7554
11	EEG	5	1.3206
12	Electrocardiogram	15	0.6991
13	Emergency	20	1.0515
14	Exercise	11	1.7295
15	Exercise Promotion	6	2.5586
16	Food	6	1.3375
17	HIV	6	2.538
18	Hypertension	8	2.2066
19	iheart	9	1.0447
20	Malaria	5	0.9744

With every good, there is an evil associated with it. Mobile healthcare is nothing different. Although there are numerous benefits and advantages of using advanced technologies in the healthcare system, but there is a chance of data loss, theft and misuse. Hence, the recent studies are based on reduction of such threats and making the platform of mobile healthcare more robust. The works of Wang et al. [2] try to provide a solution for “context-aware cyber-physical system” by mitigating the storage and computational cost of a mobile terminal and strengthening the security and cyber-physical privacy of healthcare data. Hence, recent works focus more on resolving these challenges.

The research is limited to the metadata of one database i.e., Scopus. There are numerous other databases which hold tons of information about the topic of this research. Future research could be directed toward analysis of metadata obtained from different databases. Similar research may also be carried out with other keywords like electronic healthcare or other burning topics of this era.

References

1. Mian Qaisar S, Subasi A (2020) Effective epileptic seizure detection based on the event-driven processing and machine learning for mobile healthcare. *J Ambient Intell Humaniz Comput* 1:1–13. <https://doi.org/10.1007/S12652-020-02024-9>
2. Wang S, Wang H, Li J, Wang H, Chaudhry J, Alazab M, Song H (2020) A fast CP-ABE system for cyber-physical security and privacy in mobile healthcare network. *IEEE Trans Ind Appl* 56(4):4467–4477. <https://doi.org/10.1109/TIA.2020.2969868>
3. Saha SK, Adhikary A, Jha A, Mehta VK (2021) Use of Interventions to overcome medication non-adherence. *Int J Asian Bus Inf Manage (IJABIM)* 12(3):289–318
4. Mani N, Singh A, Nimmagadda SL (2020) An IoT guided healthcare monitoring system for managing real-time notifications by fog computing services. *Procedia Computer Science* 167:850–859. <https://doi.org/10.1016/J.PROCS.2020.03.424>
5. Shuwandy ML, Zaidan BB, Zaidan AA, Albahri AS, Alamoodi AH, Albahri OS, Alazab M (2020) mHealth authentication approach based 3D touchscreen and microphone sensors for real-time remote healthcare monitoring system: comprehensive review, open issues and methodological aspects. *Comput Sci Rev* 38:100300. <https://doi.org/10.1016/J.COSREV.2020.100300>
6. Mohammed MN, Desyansah SF, Al-Zubaidi S, Yusuf E (2020) An internet of things-based smart homes and healthcare monitoring and management system: review. *J Phys: Conf Ser* 1450(1):012079. <https://doi.org/10.1088/1742-6596/1450/1/012079>
7. Shahzad SK, Ahmed D, Naqvi MR, Mushtaq MT, Iqbal MW, Munir F (2021) Ontology driven smart health service integration. *Comput Methods Programs Biomed* 207:106146. <https://doi.org/10.1016/J.CMPB.2021.106146>
8. Qaisar SM, Hussain SF (2020) Arrhythmia diagnosis by using level-crossing ECG sampling and sub-bands features extraction for mobile healthcare. *Sensors* 20(8):2252. <https://doi.org/10.3390/S20082252>
9. Alam T (2021) Blockchain-enabled mobile healthcare system architecture for the real-time monitoring of the COVID-19 patients. *SSRN Electron J*. <https://doi.org/10.2139/SSRN.3772643>
10. Arora M, Adholeya R, Sharan S (2021) An analytical hierarchical process evaluation on parameters apps-based Data Analytics for healthcare services. In: *Applications of big data in healthcare*, pp 215–239. <https://doi.org/10.1016/B978-0-12-820203-6.00011-4>
11. Lam HY, Tang YM, Tang V, Wu CH (2020) An Intelligent m-Healthcare system for improving the service quality in domestic care industry. *IFAC-PapersOnLine* 53(2):17439–17444. <https://doi.org/10.1016/J.IFACOL.2020.12.2113>
12. Cooper T, Aharony N, Bar-Ilan J (2021) Gender differences in the Israeli academia: a bibliometric analysis of different disciplines. *Aslib J Inf Manag* 73(2):160–179. <https://doi.org/10.1108/AJIM-05-2020-0170>
13. Liang J-C, Hwang G-J, Chen M-RA, Darmawansah D (2021) Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach, pp 1–27. <https://doi.org/10.1080/10494820.2021.1958348>
14. Zhang D, Zhang M, Peng C, Jung JJ, Xia F (2021) Metaphor research in the 21st century: a bibliographic analysis. *Comput Sci Inf Syst* 18(1):303–322. <https://doi.org/10.2298/csis201109059z>
15. Van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84(2):523–538
16. Orduña-Malea E, Costas R (2021) Link-based approach to study scientific software usage: the case of VOSviewer. *Scientometrics* 1–34. <https://doi.org/10.1007/S11192-021-04082-Y>
17. Sood SK, Kumar N, Saini M (2021) Scientometric analysis of literature on distributed vehicular networks: VOSViewer visualization techniques. *Artif Intell Rev* 2021:1–33. <https://doi.org/10.1007/S10462-021-09980-4>

18. Chen M, Ma Y, Song J, Lai C-F, Hu B (2016) Smart clothing: connecting human with clouds and big data for sustainable health monitoring. *Mobile Netw Appl* 21(5):825–845. <https://doi.org/10.1007/S11036-016-0745-1>
19. Dwivedi YK, Shareef MA, Simintiras AC, Lal B, Weerakkody V (2016) A generalised adoption model for services: a cross-country comparison of mobile health (m-health). *Gov Inf Q* 33(1):174–187. <https://doi.org/10.1016/J.GIQ.2015.06.003>
20. Wu IL, Li JY, Fu CY (2011) The adoption of mobile healthcare by hospital's professionals: an integrative perspective. *Decis Support Syst* 51(3):587–596. <https://doi.org/10.1016/J.DSS.2011.03.003>
21. Kumar PM, Lokesh S, Varatharajan R, Chandra Babu G, Parthasarathy P (2018) Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. *Future Gener Comput Syst* 86:527–534. <https://doi.org/10.1016/J.FUTURE.2018.04.036>
22. Chib A, van Velthoven MH, Car J (2015) mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries, vol 20, no 1, pp 4–34. <https://doi.org/10.1080/10810730.2013.864735>
23. Lu R, Lin X, Shen X (2013) SPOC: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency. *IEEE Trans Parallel Distrib Syst* 24(3):614–624. <https://doi.org/10.1109/TPDS.2012.146>
24. Wang Q, Yu Y, Yang J, Liu J (2015) Fast fabrication of flexible functional circuits based on liquid metal dual-trans printing. *Adv Mater* 27(44):7109–7116. <https://doi.org/10.1002/ADMA.201502200>
25. Li W, Song H, Zeng F (2018) Policy-based secure and trustworthy sensing for Internet of Things in smart cities. *IEEE Internet Things J* 5(2):716–723. <https://doi.org/10.1109/JIOT.2017.2720635>
26. He D, Kumar N, Wang H, Wang L, Choo KKR, Vinel A (2018) A provably-secure cross-domain handshake scheme with symptoms-matching for mobile healthcare social network. *IEEE Trans Depend Secure Comput* 15(4):633–645. <https://doi.org/10.1109/TDSC.2016.2596286>
27. Zhou J, Cao Z, Dong X, Xiong N, Vasilakos AV (2015) 4S: A secure and privacy-preserving key management scheme for cloud-assisted wireless body area network in m-healthcare social networks. *Inf Sci* 314:255–276. <https://doi.org/10.1016/J.INS.2014.09.003>

Shetkari Mitra App—An Application to Maximize the Profit of Farmers



Niranjan Girhe, Divya Chaudhari, Prachi Channe, and Avinash Bhute

Abstract In this era of technology, we have seen many applications of data science around us, such as in medical science, business analytics, finance, and many more. But there is a huge room for the applications of data science in the agricultural sector also, but very few developments are seen in this sector. In view of the need and the potential of data science in agriculture, an application is developed with the objective to assist farmers toward smart farming practices. This application will be in the form of a mobile application and website. It serves as a platform to connect farmers with the market. It provides current market prices and the trend of prices for different crops in the local market. It also provides a better communication mode between farmers in form of forums. On the sidelines, relevant government schemes with the all-at-one-place idea. It will also be able to push notifications and alerts for market-rate peaks, weather alerts, and government schemes. This is achieved by taking data from various sources and government websites. ISRO satellites have collected humongous weather and soil data over the period. The app could serve as a crop and fertilizer guide by using the available data from the aforementioned sources and user input. Additionally, there are many available open data sources which could be used effectively to augment current functionalities as well as new ones. For example, Google Maps could be integrated which can locate buyers and agri shops in the vicinity. The farmers could make use of the web-based forum platforms which will link farmers with professional experts as well as other farmers all over India. This will open an incredible avenue for farmers to get the solution to their problems as well as get a sense of community with farmers across the states. Some many farmers and communities are still deprived of new techniques. This application will help them in changing their crop patterns and also to act smarter and more precisely.

Keywords Agriculture · Farming · Mobile application · Smart farming · Crop suggestion

N. Girhe (✉) · D. Chaudhari · P. Channe · A. Bhute
School of Computer Engineering, MIT Academy of Engineering, Alandi Pune, Maharashtra, India
e-mail: niranjangirheindia@gmail.com

1 Introduction

In the current age of information technology and digital reach, farmers are far-off from smart farming practices. With the deprivation of these advances in the farming sector, the logistical supply chain remains untapped of its true potential. The reach of the product is limited to local markets. Due to a lack of information about the global markets, they are forced to sell the crops to monopoly buyers at unimaginably lower prices. This remains the main cause of farmer distress in rural India. Their production is limited to the local level, yet they don't get proper information and news about their regional market, schemes, and weather. The climate change crisis is a challenge to the farming community. As is seen in humans, crops are attracted to different diseases. Even if the scientists could come up with guidelines, the information does not reach farmers in good time. Advanced technology could lead to higher productivity as compared to traditional and local practices, but lack of awareness has put local farmers in tough situations.

India is the second-largest smartphone market. Most Indians have at least one smartphone in their family. The fact that the Internet is the cheapest in India and opens a lot of opportunities for applications to help out farmers. Currently, there are some mobile applications that assist farmers in smart farming. However, most of them remain out of reach of the farmers due to their complex user interfaces.

Today, the information is readily made available via different sources. It is too abundant and complex that hardly farmers keep track of them. We aim to bridge this gap by creating a mobile application that will assimilate information from various sources and present it in a lucid comprehensible manner that anyone can understand. The need of the hour is to make use of all the research in agriculture, much like technology transfer. The data provided by farmers could help understand the supply and thus better plan food security. The platform could be used by all farmers to form large communities where they could solve each other's problems. It will guide them in basic things like weather forecast, quality and quantity of pesticides and fertilizer to be used as per crop, and provide helpful government and non-government schemes.

This paper includes a literature survey that describes the use of technology to improve farming. It also describes the survey conducted to know the current approach taken by farmers and the methodology they use. The paper proposes a mobile app 'Shetkari Mitra'—an ideal mechanism that could assist farmers toward increasing their profits.

2 Literature Survey

- R The aim of Lomotey, Chai et al. and Giri et al., [1, 2], is to give farmers mobile access to up-to-date pesticide information so that they can make decisions about which pesticides to use, how to use them, when to use them, and so on. Mobicrop is a mobile distributed system with a three-layered deployment strategy that includes

mobile nodes, cloud-hosted middleware, and a cloud-hosted database server. The app also contains a caching system that allows it to be used offline.

- Daniel Mutembesa, Ernest Mwebaze, Solomon Nsumba, Christopher Omongo, and Humphrey Mutaasa [3] proposed to complement low-resourced agricultural expert surveys, surveillance of crop diseases, and pests. It also consists of training the farmers. Authors suggest a very smart method to diagnose crop diseases, it uses ML to tell the disease by looking at photos. The paper also suggests data collection which will help agents to keep track in real-time.
- Monika Jadhav, Vishakha Jagtap, and Shankar M. Patil [4] proposed an app that can get real-time updates about the rates of vegetables, fruit, and sell their products at proper rates. Farming-related notices from the government will get proper information about different schemes. Feature of weather information which will help farmers to plan for the next 2–3 days. This system combines mobile communication systems with GPS for economical and sleek farming. This review paper presents a Database Management System and the use of smartphones in agriculture. They have developed a bot that automates the existing manual system with the help of computerized equipment and software.
- The app [5] allows farmers to explore the latest trends in agriculture; informs them about methodologies involved in selecting and understanding soil health and specific crops based on the environmental parameters; provides comprehensive information on crop protection; and advises on managing plants against diseases, weeds and other pests
- The app [6] aims to improve decision-making by farmers which could result in a reduction of cost, quality improvement, increase in income, and enhanced opportunities for livelihood.
- Visconti, Giannoccaro et al. and Ganjekar et al., [7, 8], describe the design of an innovative IoT-oriented farm management system that employs low-cost WSN for detecting soil, environmental, and crops parameters that are properly processed, in combination with weather forecasts, are used to determine future farming activities based on agronomic models implemented in the software platform.

2.1 Literature Summary

See Table 1.

Table 1 Literature summary

Ref No	Methodology/technique /approach used	Pros	Cons
1	A mobile distributed system that follows a three-layered deployment, consisting of mobile nodes, a cloud-hosted middleware, and a cloud-hosted database server	<ul style="list-style-type: none"> • Up-to-date • Dual caching methodology (Supports offline accessibility) 	<ul style="list-style-type: none"> • Intermittent loss of connectivity leads to stale data
2	Geo-tag image with GPS coordinates and complete micro-tasks like labeling the image and writing comments	<ul style="list-style-type: none"> • Use of ML to detect disease • Q & A • News channel feel • Easy to use. Point and capture 	<ul style="list-style-type: none"> • Processing power to run disease detection model • Strong and reliable GPS • Needs good connectivity
3	Uses Database Management System a bot that automates the An existing manual system with the help of computerized equipment and software	<ul style="list-style-type: none"> • One-Stop Solution to all agricultural information need • Location-specific information delivery • Highly authentic and reliable database on agriculture 	<ul style="list-style-type: none"> • Does not provide the soil pattern details • Does not give crop suggestions
4	Chat with experts and crop calendar which uses a database. Fetches news articles related to agriculture	<ul style="list-style-type: none"> • User can download packages • User friendly • Available in different languages 	<ul style="list-style-type: none"> • The chat forum doesn't respond efficiently • Delivery of products doesn't happen sometimes
5	While Agri VAS channels can pose limitations in the format and extent of the information delivered, such as character restrictions for SMS, an app's capability to provide richer content	<ul style="list-style-type: none"> • An extra module such as News and Jobs • Ask the expert • Multilingual • Kisan Call Center 	<ul style="list-style-type: none"> • Not in all states • Chats names don't show up • Not updated timely • Slow performance • The user gets spam calls
6	Utilization of low-cost wireless electronic modules for acquiring data from connected sensors. The fertigation system operation is supported by solar-powered and constantly monitoring soil and environmental parameters; thanks to sensor nodes properly positioned in cultivated land or greenhouse, this allows monitoring and remote control from enabled users (farmer or agronomist)	<ul style="list-style-type: none"> • Live to monitor • Works on solar energy • Cloud storage 	<ul style="list-style-type: none"> • Regular maintenance • Works in a controlled environment only • Upgradation is costly

3 System Architecture and Module Description

The main target audience of the app is farmers. Keeping this in mind, the app has been developed with a super easy interface and less-educated farmers should be able to use it comfortably as well. The app helps farmers to access the weather forecast at their current location and make the decision accordingly. Internet is full of information, but accessing it is a major challenge for a non-English speaker, the app intends to solve this problem to give direct access to such useful websites in their choice of language. There are a large number of crops that a user can potentially plant, so this app includes a crop suggestion feature along with a fertilizer calculator and provide a proper plan for the cultivation of a selected crop. The selection of crop automatically subscribes the user to the forums of that selected crop, where users can share their queries. Every user subscribed to the same forum or crop will able to help each other and share their experiences. A careful balance is exercised between the comfort of use and abundancy of important features (Fig. 1).

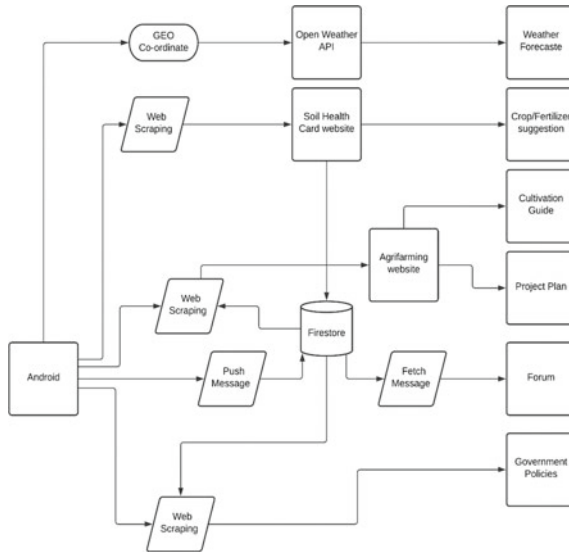


Fig. 1 System architecture

3.1 Module Description

There are several modules incorporated so that farmers can make a better judgment. For example:

- Weather Forecast module, in which users get to know the current as well as forecast weather. This module uses open weather API to fetch current as well as forecast weather at users' GPS location. The data is provided by openweathermap.org/.
- The crop/fertilizer suggestion module uses data from the soil health card government website maintained by the Ministry of Agriculture and Farmer Welfare India, to suggest crops based on the user's location and soil NPK values as well as suggest the appropriate fertilizer with quantity.
- The cultivation guide module provides guides from www.agrifarming.in [9], which are suitable for Indian users and have reliable in-depth information. This module gives very easy access to this website without any effort. This module is very effective to use for relatively less-educated users as there are no inputs necessary.
- The project plan modules provide information from www.agrifarming.in [9] regarding projects which have in-depth planning information for users.
- The forums module is one of the important modules that provides a unique platform to share ideas and problems of similar crops for all users. Users from all over the world can come together and form a community using this module. Forums are generated separately for each crop and only the subscriber of that crop can access them.
- Government policies related to agriculture are a very sensitive topic. Policies are different at different places and at different times. So to avoid misinformation, the Government policies module provides information from reliable third-party websites and provides the official government websites.
- Ask Developer which provides user support by giving a direct line of communication with the app developers. With a click, a user can mail the developer anonymously. And the last module Kisan Call center promotes the use of the Kisan call center and makes it easy to call. With a click, the user can place a direct call to the Kisan call center.

3.2 Algorithms and Smart Methods

Internet consumption, ram utilization, language barrier, and complicated UI are all important challenges in the usability of such apps. All of these concerns have been addressed, and we have developed smart methods and algorithms to handle them. One such method aids in the loading of forum messages, which are loaded in such a way that just the message now visible on the user screen is downloaded from the server, with new/old (depending on the scrolling direction) posts being fetched as the user scrolls. After then, all of the communications are cached for viewing offline.

Images should not be saved offline since they might take up a lot of space and slow down the user's device. The forum data is erased from the cache after removing the crop to clear the superfluous data.

To overcome the language barrier, the material is translated into several languages and presented to users based on their preferences. The app’s dynamic information was translated using Google Translate, while the UI text translation was hardcoded for better reaction time. The design of this app was influenced by popular messaging applications such as Telegram and WhatsApp. The forum’s user interface is extremely similar to that of a messaging app, which makes it easier for new users to use and doesn’t require any instructions.

3.3 Backend Configuration

All requests and authentication are handled by Firebase on the backend. Because it is the simplest and everyone has a smartphone, the phone number and OTP were utilized for verification. We also utilize the Firestore database, which is a No-SQL database, to handle the communications. Because the data stored from the user is not consistent, such as the number of crop options, language, and other characteristics, No-SQL was chosen.

For storing the photographs posted by users on the forums, we used Google cloud storage. Other users may then quickly view these photographs using the Shetkari Mitra App forums.

For improved reach, the Google cloud messaging platform was utilized to send notifications to a specific demographic. This also includes the ability to send peer-to-peer notifications after responding to someone’s question on the forum (Fig. 2).

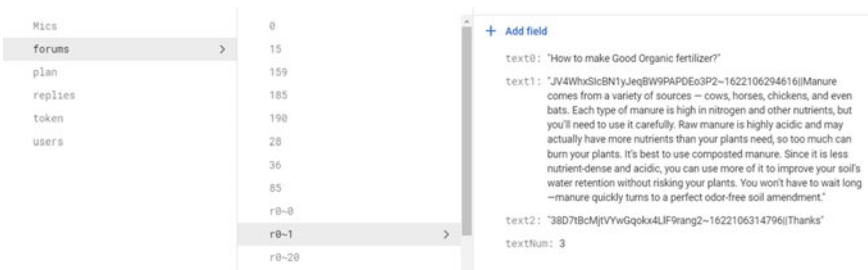


Fig. 2 Database structure

3.4 Innovative Content

This app provides the platform in the form of forums using firebase. Users can subscribe and unsubscribe to that module and share their ideas and problems. Users from all over the world can come together and help each other. The forum provides

anonymity to the users which can help them open up better. Technology such as web scraping is used to provide information that can be easily accessed even by illiterate users.

4 Methodology

- **Weather forecast**

Open weather maps provide a detailed report on current and forecast weather using geo coordinates. This app uses this fact and provides necessary data to the user using then geo coordinates.

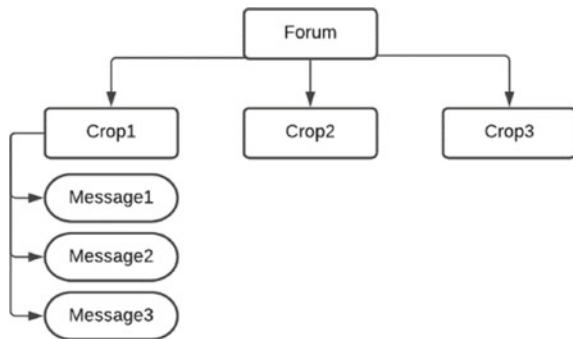
- **Web scraping**

Today in the age of information, information is readily available and in bulk but not easily accessible. To bridge this gap, we intend to provide easy access to very useful websites and portals to the user. This was archived using web views and simple web scraping using network monitoring.

- **Forums**

Forums are very useful to find a solution to particular problems quickly. But the problem with the forums is that they are not well organized and sometimes can be hard to access them particularly by illiterate users. To bridge this gap, this app intends to have a very easy and familiar approach to forums by separating them by the crops. Users having some crops will be able to communicate using that crop forum. This was achieved using firebase firestore, every forum has a unique id that is the same as the crop id to keep them separated (Fig. 3).

Fig. 3 Data storage structure



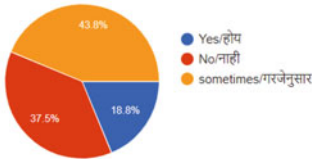
5 Pre-Survey

A survey was taken regarding the farming details of customers. The survey aimed to collect necessary information related to land and crop to form a problem statement and to note pre-state conditions. The data collected has to be analyzed and conclusions were made.

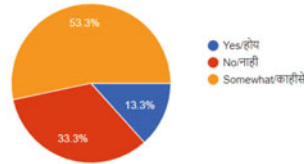
The surveys included important points such as:

- Land details (owner, size, type, etc.)
- Crop details
- Fertilizers and pesticides used
- Approach
- Satisfaction

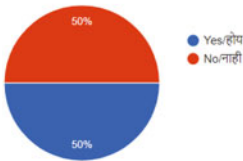
Q Do you take guidance from any professional before sowing and producing?



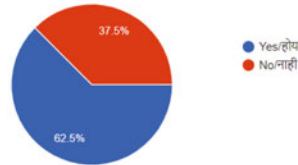
Q Are you Satisfied with the amount of production in agriculture?



Q Are you aware of Govt. schemes for farmers?



Q Do you take any advantage of Govt. schemes?



Complete Survey result in Appendix 1

6 Summary From the Survey

These survey results were quite enlightening about agricultural advancements and their practical outreach to rural farmers. We can see that almost 50% of farmers cultivate crops on less than 5 hectares of land. Farmers in India farm on small land and eventually result much lower profits. Farmers mostly opt for cereals and sugar crops. Apart from this, they also grow oilseeds plants. 50% of farmers are not at all satisfied with the market price for their produce. Also, there is some portion of farmers who are not completely satisfied. Some farmers do not have any guidance

from professionals before sowing and producing. Lack of advice and suggestion results in lesser produce. Farmers are unaware of the optimum usage of fertilizers and pesticides. Many are interested in business plans and assistance in farming.

From the survey, we also observed that most generally farmers are reluctant to ask for help even when needed. 78.6% of farmers have never consulted the Kisan Call Center and that's a matter of concern. Nearly 57% of farmers are unaware of the government schemes which are solemnly dedicated to their welfare. Refer Appendix for more Survey snaps.

7 Mobile Application

From the summary of the survey, it was clear that the problems that a farmer faces can be solved using better communication mode. An application is developed to assist farmers toward increasing communication within the community and with the experts.

7.1 User Flow

The user flow is designed by keeping in mind that the app will be majorly used by farmers. Further, the app has a super easy interface, and less-educated farmers would be able to use it comfortably as well.

The user's first screen will be the Homepage given that the user is logged in using a very simple login process. There on the user has the option to navigate to all the important modules from the Homepage itself. Other screens/modules are two clicks away from the Homepage making it at most simple design in terms of navigation (Fig. 4).

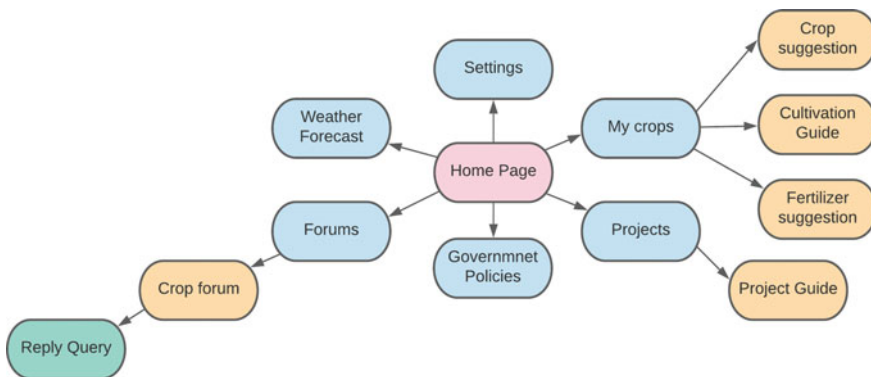


Fig. 4 Application user flow

Home Screen

Users will be greeted by the homepage of the app which consists of a dashboard with current weather details and links to all important modules.

Weather Forecast

A detailed weather forecast of 5 days is shown to the user, so the user can decide their future actions. This screen gives information on temperature, humidity, wind speed, and precipitation.

My Crops

This screen shows all the crops that users have added to his/her account for easy access. These crops can be removed by the user. Further on clicking, it will open a cultivation guide for that crop.

Crop Suggestion

Users can get crop suggestions based on their location. The screen also allows users to add the crop to their account so users can get the guides and project plans of that crop.

Cultivation Guide

A proper cultivation guide for the added crop can be viewed from this module. This module is automatically translated into the language of the user's choice.

Forum

In this screen/activity, the user can ask questions to the community in any language with images attached. The forum is divided into crops selected by the user to reach the correct audience or the users who added the same crop to their account. The app also provides a common forum so all users can ask common questions on that platform.

Reply to Query

Any user can answer the question asked in forums using this activity. The user who asked the question will get notified using notification. All users can access the reply section of every question.

Government Policies

Information on government policies and schemes is available on this screen. Users can access the websites and portals directly from this screen/module.

Forums Page

From this page, users can manage forums and ask queries directly to the developer. This helps the user to navigate much easier (Fig. 5).

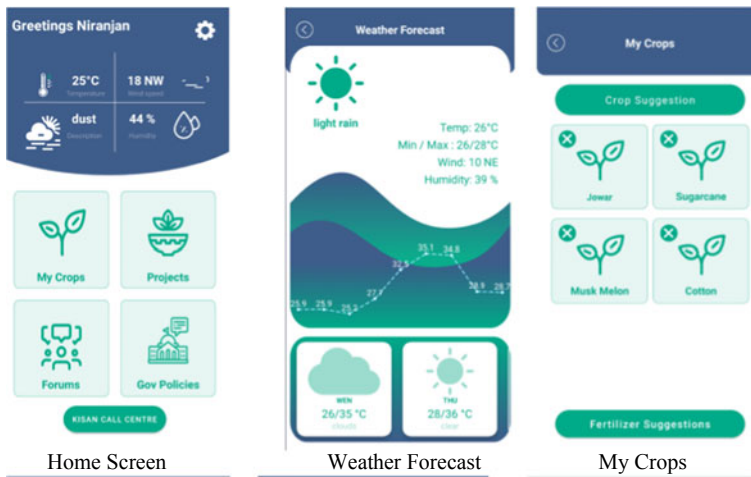


Fig. 5 Screenshots of Shetkari Mitra app (a) For all results screenshots are in Appendix 2

7.2 Result Analysis and Discussion

Users who installed the app asked more advice or questions than non-users. Users found that access to information and technology is much easier than ever thanks to multilingual support and properly organized modules. Also, the cultivation guides and project plans were helpful to make proper decisions. Many users contacted the developers reporting suggestions or asking for help using the Ask developer module which provides 24×7 support. We also found that most of the users were more likely to call the Kisan call center with the app.

8 Conclusion

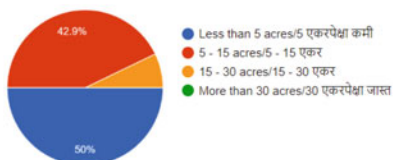
Today, the technology is available at everyone's fingertip but still, many farmers are unaware of the outside world and technical advancements in this sector. Our app provides a window for the user to look outside by providing them easy access to important and reliable websites. Most of the farmers have small farms and plant crops without diversity according to Q 1 and Q 2. The app overcomes these issues by providing project plans for both small farms and large farms with a wide range of crops. Most of the farmers aren't aware of the Kisan call center or not comfortable with the process (Q 9). This app promotes the Kisan call center and also makes the process easy. Kisan call center can help a farmer at a much deeper and personalized level. Most farmers don't take expert advice or professional guidance and they are not satisfied with the end product (Q 5 and Q 6). With the help of forums, this app

can resolve this issue, users can share their success stories and plans with each other to promote their best plans and have healthy competition. Some of the farmers aren't aware of any government schemes for their projects and don't take any advantage of them (Q 11 and Q 12). This app has a special module to spread awareness of such schemes.

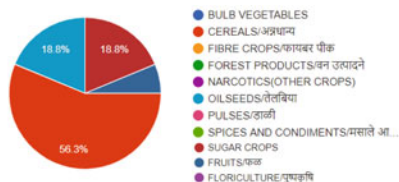
Appendix 1

Pre-survey report.

Q 1 How much land do you farm?



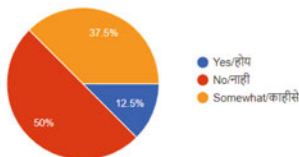
Q 2 What do you grow?



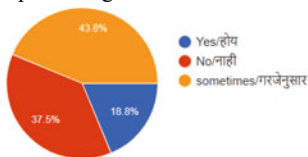
Q 3 Where do you sell your product?



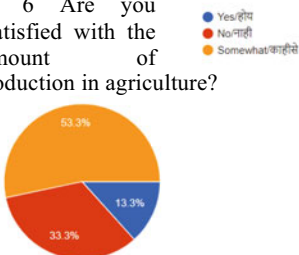
Q 4 Are you Satisfied with the market prices?



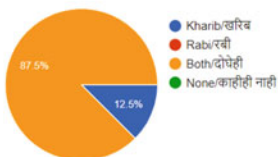
Q 5 Do you take guidance from any professional before sowing and producing?



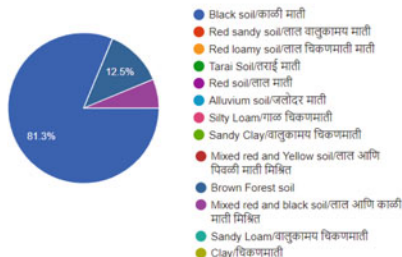
Q 6 Are you Satisfied with the amount of production in agriculture?



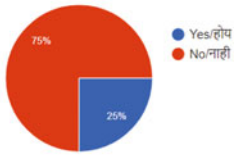
Q 7 Do you plant seasonal crops?



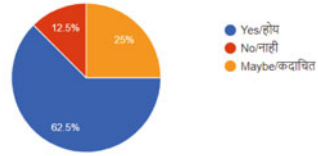
Q 8 Which type of soil is on your farm?



Q 9 Do you call the Kisan call center for help?



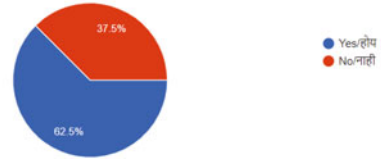
Q 10 Do you need help creating business plans or other assistance aimed at increasing production?



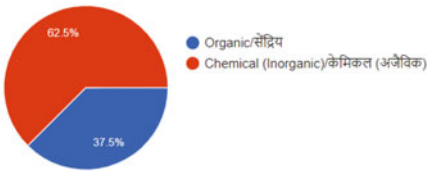
Q 11 Are you aware of Govt. schemes for farmers?



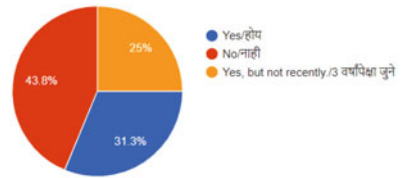
Q 12 Do you take any advantage of Govt. schemes?



Q 13 Which fertilizers do you use most?

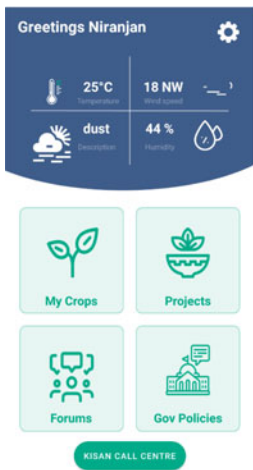


Q 14 Have you done a soil test?

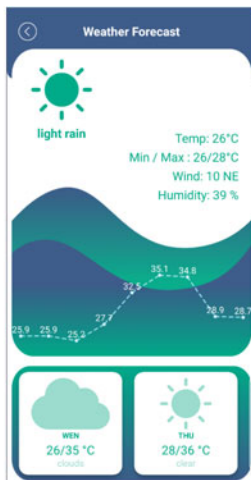


Appendix 2

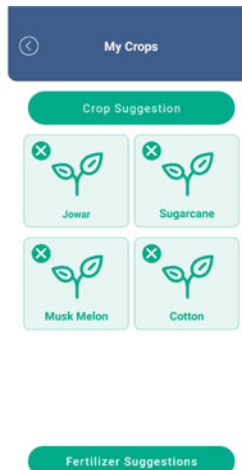
App Screenshots.



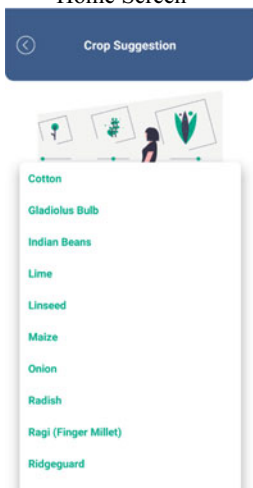
Home Screen



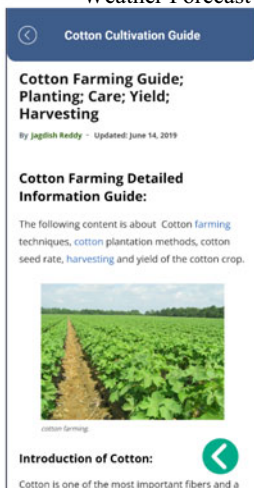
Weather Forecast



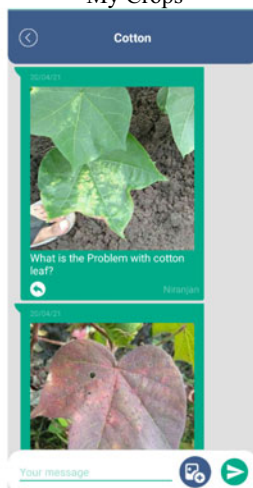
My Crops



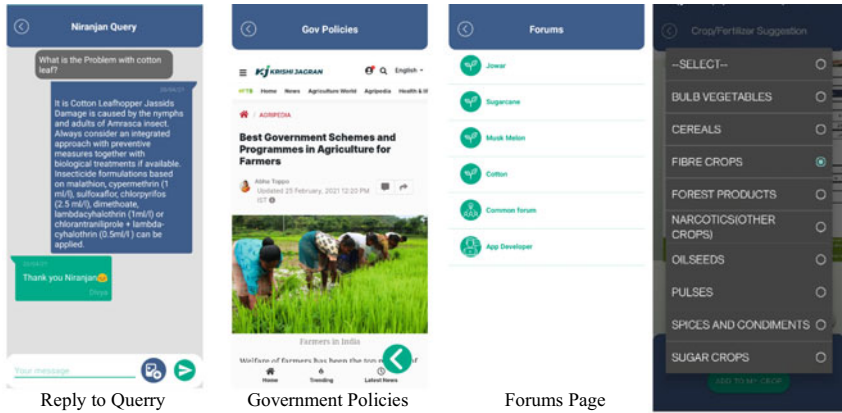
Crop suggestion



Cultivation Guide



Forum

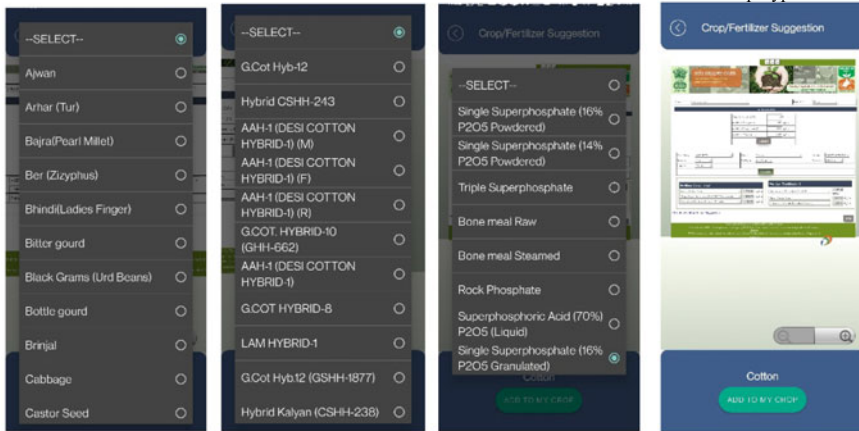


Reply to Query

Government Policies

Forums Page

Crop Type



Crops

Crop Variety

Fertilizer Combination

Fertilizer suggestions portal

References

1. Lomotey RK, Chai Y, Jamal S, Deters R (2013) MobiCrop: supporting crop farmers with a cloud-enabled mobile app. In: 2013 IEEE 6th international conference on service-oriented computing and applications, Koloa, HI, USA, pp 182–189. <https://doi.org/10.1109/SOCA.2013.19.IEEE>
2. Giri MB, Pippal RS (2017) A pervasive system for optimizing water utilization in agriculture using wireless sensor networks. *Int J Appl Environ Sci* 12(8):1605–1615
3. Mutembesa D, Mwebaze E, Nsumba S, Omongo C, Mutaasa H (2019) Mobile community sensing with smallholder farmers in a developing nation, a scaled pilot for crop health monitoring
4. Jadhav M, Jagtap V, Patil SM (2019) *Int Res J Eng Technol (IRJET)*, 06(04)
5. AgriApp Technologies Private Limited, AgriApp: Smart Farming App for Indian Agriculture, <https://play.google.com/store/apps/details?id=com.criyagen>
6. GSMA, Case Study IFFCO KISAN agriculture app evolution to data driven services in agriculture <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2016/10/IFFCO-Kisan-Agricultural-App.pdf>

7. Visconti P, Giannoccaro NI, de Fazio R, Strazzella S, Cafagna D (2020) IoT-oriented software platform applied to sensor based farming facility with smartphone farmer app 9(3)
8. Ganjewar PD, Barani S, Wagh SJ (2015) Data reduction using incremental Naive Bayes Prediction (INBP) in WSN. In 2015 international conference on information processing (ICIP), pp. 398–403. IEEE
9. AgriFarming, <https://www.agrifarming.in>

Performance Analysis of Spectrum Sensing Algorithms



Jaya Sharma and Kirti Vyas

Abstract The work involves designing and implementing conventional spectrum detection systems for the orthogonal frequency division multiplexing (OFDM) structure. The different detection algorithms are used to assess the performance of the system at the end of the receptor. The structure efficiency is explored by analysing Pd Vs SNR for various spectrum sensing techniques. The performance of the system was examined by simulating the energy detection (ED), matching filter (MF) and cyclostationary method in MATLAB 2014. The result of the projected work reveals that the cyclostationary exceeds the ED and MF technique and achieved a gain of 2 and 3.2 dB.

Keywords ED · Spectrum · Matched filter · Detection

1 Introduction

Over the past few years, a number of studies have been proposed for the use of unused spectrum works. The shortage of spectrum is increasing day by day due to the exponential rise of radio operations such as health care, education, industry and so on. Spectrum overlay is one of the techniques to solve the spectrum access. Cognitive radio (CR) is a spectrum overlay approach that detects the availability or unavailability of the main user (PU) and the spectrum assigned to the secondary user (SU) [1]. Several works are proposed on ED [2], MF [3], and cyclostationary [4], where spectrum can be detected only in absence of PU. However, several works projected the detection of spectrum in both presence and absence of PU. In spectrum sensing approaches, threshold is determined and compared with received signal [5]. ED is simple and easy detection approach. However, the requirement of highly received

J. Sharma (✉) · K. Vyas

Department of Electronics and Communication Engineering, ARYA College of Engineering and I.T, Jaipur, India

e-mail: jayasharma19994@gmail.com

K. Vyas

e-mail: kirtivyas.ec@aryacollege.in

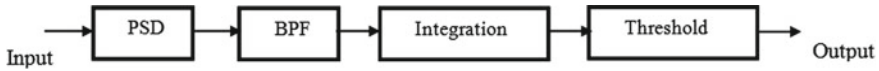


Fig. 1 Energy detection

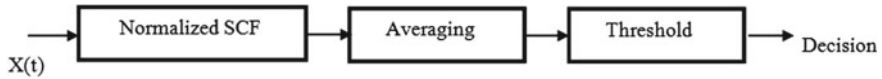


Fig. 2 Matched filter detection

power of one of the constraints needs to be fixed [6]. The schematic of ED is given in Fig. 1. MF is robust detection approached, given in Fig. 2. The requirement of received signal power is less, but channel condition should be estimated in advanced. Cyclostationary is one of the ideal approaches for spectrum detection, when channel information is known in advanced. It does not need channel prior channel estimation and highly received power signal [7]. In [8], the author designed an advanced energy detection system for Cr. The projected method estimates the availability of spectrum in presence and absent of primary user. However, high SNR for detection is seen as a major constraints in Ed. In [9], the author designs an unitary bit-based spectrum sensing algorithm for CR. The simulation result reveals that the performance of projected algorithm is better than the energy and quadratic detecting. However, high complexity is seen a major problem. In [10], a detailed study on spectrum detection algorithm for cognitive radio was carried out. The proposed work highlighted the constraints and benefit of different algorithms. The regularization of CR in advanced radio network was also discussed. Additionally, the upcoming challenges in design of Cr are also highlighted. In [11], spectrum sensing algorithms are designed for fourth-generation radio subsystem. The primary objective of the projected algorithm was to detect the idle spectrum without causing any interference and disturbance to PU. In [12], a cooperative algorithm was designed to detect the idle spectrum for dynamic network. The several parameters such as PFA, Pd were derived and estimated. The outcome of the work shows that the projected algorithm efficiently enhanced the performance of detection.

2 System Model

The schematic of OFDM with spectrum sensing algorithm model is given in Fig. 3. OFDM is the latest waveform method utilized in current radio system. It is a multi-carrier technique and offers no interference between the subcarriers. However, loss of spectrum due to insertion of cyclic prefix is seen as a significant disadvantage of OFDM structure. In this paper, we are applying spectrum sensing algorithm to the OFDM received signal and analysing the detection performance of the framework.

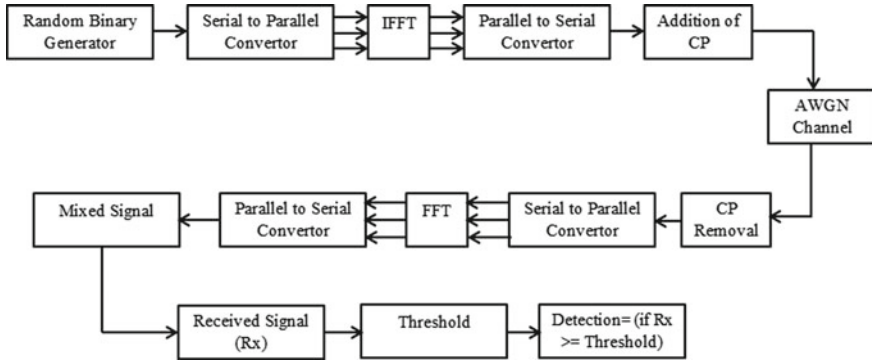


Fig. 3 OFDM system model schematic with spectrum sensing algorithm

Let us the consider an OFDM received signal of the spectrum sensing algorithm given as:

$$q(n) = q(0) q(1) q(2), \dots, q(L - 1) \quad 0 \leq ml \leq L - 1 \tag{1}$$

The time domain OFDM symbol is given as:

$$Q(n) = Q(0)Q(1)Q(2), \dots, Q(L - 1) \quad 0 \leq l \leq L - 1 \tag{2}$$

The modulated OFDM signal is represented as:

$$A(t) = \sum_{l=0}^{L-1} q(t) * l \exp^{i6.28l \Delta f t} \quad 0 \leq t \leq T_d \tag{3}$$

The received OFDM symbol is expressed as:

$$z_i(l) = x_i(l) + n_i(l) \tag{4}$$

where $z_i(l)$ denotes the received signal, $x_i(l)$ is original signal, and $n_i(l)$ is the noise.

The decision is based on following criterion:

$$E_i = \frac{1}{M} \sum_{i=1}^M |z_i(l)|^2 \tag{5}$$

The following hypothesis is suggested for the determination of

$$z_i(l) = n_i(l)H_o \quad (6)$$

$$z_i(l) = x_i(l) + n_i(l)H_1 \quad (7)$$

H_o denotes the absence of PU, and H_1 denotes the presence of PU.

3 Simulation Result

The projected work is rigorously analysed through the mathematical and computer simulation. The parameters of the simulation are given in Table 1.

The performance of probability of detection $P(d)$ for Ed is given in Fig. 4. The pd is obtained when probability of OFDM signal is greater than the threshold value. Spectrum analysis is performed by using Ed algorithm. The Pd is 1 at the SNR of 7 dB, which is considered to be very high and need to be reduced. However, the implementation of Ed is very simple and less complex. The MF detection performance is

Table 1 Simulation parameters

S. No	Parameters
1	OFDM system
2	16-QAM
3	ED, MF and cyclostationary algorithms
4	Subcarriers = 64
5	FFT = 64

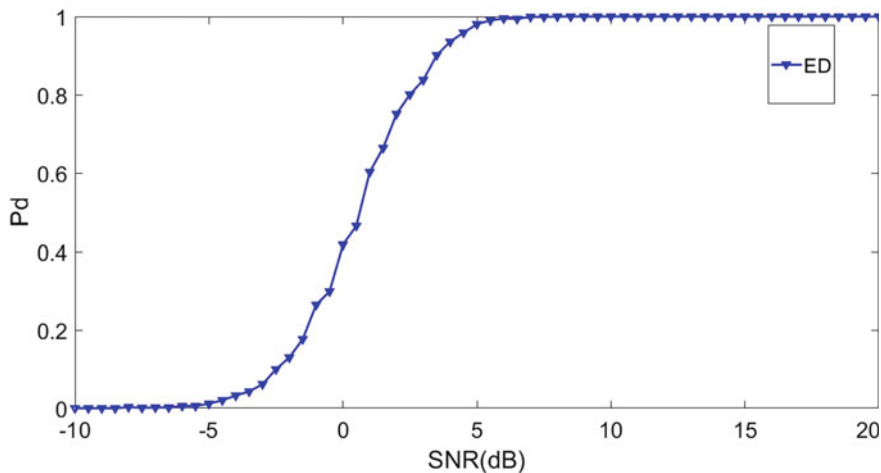


Fig. 4 ED detection

indicated by Fig. 5. It is seen that the Pd is 1 at the SNR of 4.2 dB and accomplished a gain of 2.8 dB better than Ed. The cyclostationary detection analysis is given by Fig. 6. The 100% detection is obtained at the SNR of -3 dB and accomplished a gain of 10 and 7.2 dB better than Ed and MF. The bit error rate (BER) performance of the projected methods is shown in Fig. 7. It is seen that MF outperforms the ED and cyclostationary algorithms. Table 2 indicates the comparison between the proposed algorithms.

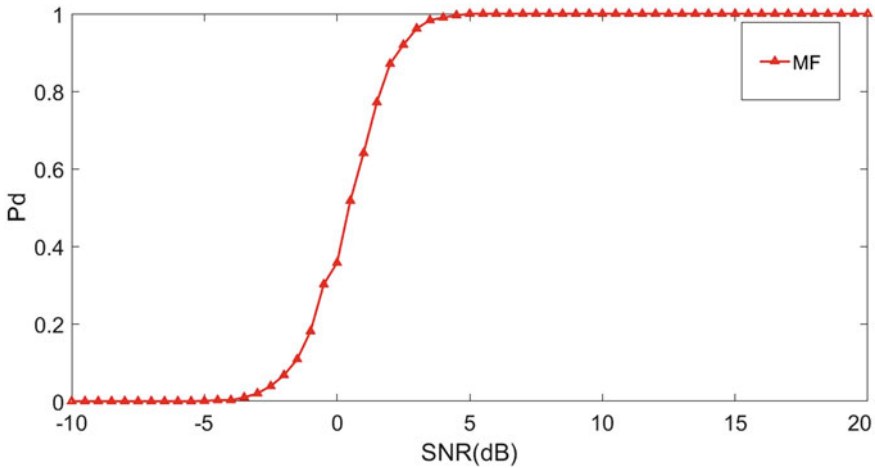


Fig. 5 MF detection

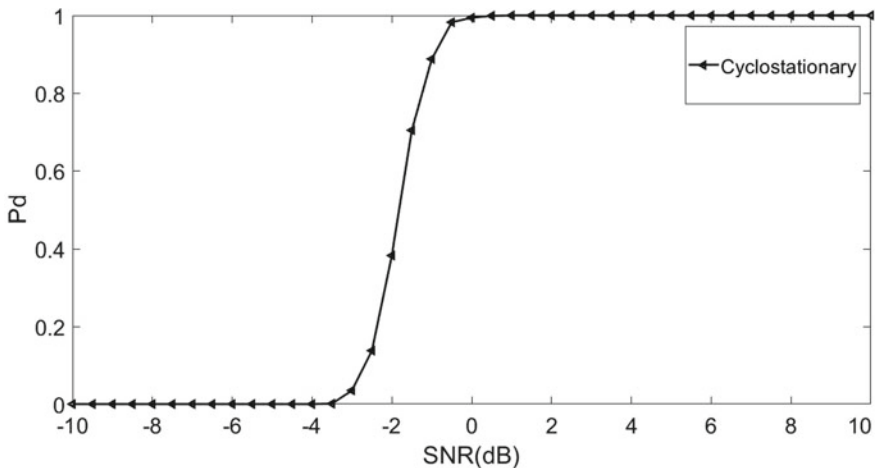


Fig. 6 Cyclostationary detection

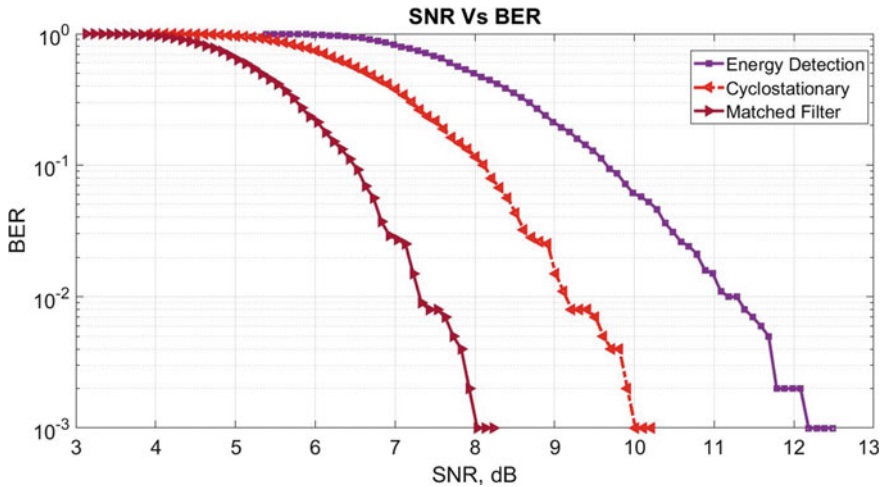


Fig. 7 BER curve

Table 2 Comparison of sensing algorithms

Techniques	Characteristics
Energy detection	<ol style="list-style-type: none"> 1. Easy to design 2. Time delay is high 3. Requirement of SNR is high 4. Performance is affected by noise
Cyclostationary detection	<ol style="list-style-type: none"> 1. Implementation is complicated 2. Time delay is medium 3. Requirement of SNR is less 4. Performance is affected by noise
Matched filter	<ol style="list-style-type: none"> 1. Implementation is complicated 2. Time delay is low 3. Requirement of SNR is less 4. Robust to noise 5. Performance is affected by interference among PU and Su

4 Conclusion

The implementation of CR in OFDM framework is one of the difficult tasks and propitious method for efficient spectrum access. We projected a CR-based spectrum detection algorithms and compared the detection performance. It is observed that the cyclostationary detection performance is better than the ED and MF. In the future, we will live in a digital world where our day-to-day needs will depend heavily on digital data of all kinds. In order to meet high-speed data requirements in real time, a very high-speed data transfer network will be required. All of this new next-generation communication system is going to be needed.

References

1. Ziafat S, Ejaz W, Jamal H (2011) Spectrum sensing techniques for cognitive radio networks: performance analysis. In 2011 IEEE MTT-S international microwave workshop series on intelligent radio for future personal terminals, pp 1–4. <https://doi.org/10.1109/IMWS2.2011.6027191>
2. Arun K, NandhaKumar P (2019) OFDM system with cyclostationary feature detection spectrum sensing. *ICT Express* 5(1):21–25
3. Choi NH, Hwang JH, Zheng G, Han N, Kim JM (2008) A cognitive UWB testbed employing adaptive pulse generation. In: 3rd International conference on cognitive radio oriented wireless networks and communications, CrownCom 2008, pp 1–6
4. Kockaya K, Develi I (2020) Spectrum sensing in cognitive radio networks: threshold optimization and analysis. *J Wireless Com Network* 255. <https://doi.org/10.1186/s13638-020-01870-7>
5. Claudino L, Abrão T (2017) Spectrum sensing methods for cognitive radio networks: a review. *Wireless Pers Commun* 95:5003–5037. <https://doi.org/10.1007/s11277-017-4143-1>
6. Arun K, Mohit KS, Kanchan S, Suraj K (2020) NOMA based CR for QAM-64 and QAM-256. *Egyptian Inf J* 21(2):67–71
7. Joshi GP, Kim SW, Kim BS (2009) An efficient MAC protocol for improving the network throughput for cognitive radio networks. In: Third international conference on next generation mobile applications, services and technologies, NGMAST 2009, pp 271–275
8. Nandhakumar P, Arun K (2016) Analysis of OFDM system with energy detection spectrum sensing. *Indian J Sci Technol* 9(16):1–6
9. Yadav PV, Alimohammad A, Harris F (2019) Efficient design and implementation of energy detection-based spectrum sensing. *Circuits Syst Signal Process* 38:5187–5211
10. Swetha N, Sastry PN, Rao YR (2014) Analysis of spectrum sensing based on energy detection method in cognitive radio networks. In: 2014 international conference on IT convergence and security (ICITCS), pp 1–4. <https://doi.org/10.1109/ICITCS.2014.7021738>
11. Mahendru G, Shukla A, Banerjee PA (2020) Novel mathematical model for energy detection based spectrum sensing in cognitive radio networks. *Wireless Pers Commun* 110:1237–1249

Impact of Benefits on Preference for Traditional Detailing



Saibal Kumar Saha, Bibeth Sharma, Sonia Munjal, and Ajeya Jha

Abstract Pharmaceutical corporations provide promotional inputs to physicians and advertise their medicines and other products with the help of detailing. The objective of the study is to determine if preference of the physicians for traditional detailing is an outcome of select benefits sought. Seven benefits that are frequently requested by physicians were identified and their preference for traditional detailing was measured using a seven-point Likert scale with the help of a questionnaire. 384 samples were taken from non-allopathic practitioners. A regression analysis was done with traditional detailing as the dependent variable and the benefits sought as the independent variables. It was found that preference for traditional detailing had strong linkage with certain benefits, important to the physicians like social interaction with MR, possibility of technology failures, and personalized information.

Keywords Traditional detailing · E-detailing · Physician · Regression · Medical representative · Information

1 Introduction

In 2005, the pharmaceutical sector in the USA spent more than \$18 billion on marketing pharmaceuticals, with drug sampling and detailing efforts accounting for the most of this investment. To remain in the competition, pharma companies need to strategize to target the physicians and optimize their return on the marketing investments [1]. According to studies, there is a strong positive correlation between

S. K. Saha (✉) · B. Sharma · A. Jha
Department of Management Studies, Sikkim Manipal Institute of Technology, Sikkim Manipal
University, Gangtok, India
e-mail: saibal115@gmail.com

B. Sharma
e-mail: bibeth.s@smit.smu.edu.in

S. Munjal
Department of Management Studies, DIT University, Mussoorie, India

a physician's exposure to a product and their prescribing habit, hence pharma corporations must promote their products aggressively [2]. There are additional charges for a higher promotion level as well; there are fewer blockbuster pharmaceuticals in the firm's new product offerings (brands with a sales potential of more than \$1 billion), and pharma firms are finding it harder to identify new drugs that can continue growth and compensate for revenue losses and short life cycles of new products [3]. E-detailing or electronic detailing is a marketing communication tool which is gaining traction in India in current years. Pharma companies need to be alert about the elements that impact Indian physicians' adoption of e-detailing to assure broad acceptance and long-term viability of this instrument [4].

2 Literature Review

Sales promotion in the pharmaceutical sector differs with respect to consumer product industries as most promotional activities are directed at physicians rather than end users. Pharmaceutical corporations focus on physicians, advertise their medicines, and provide promotional information to them since physicians serve as gatekeepers [5]. Due to the enormous number of pharmaceuticals available for prescription, building a brand is critical to a company's success. As a result, product differentiation can be aided by branding. Branding is vital since it promotes and enhances brand awareness while also helping to generate a competitive advantage [5]. The pharmaceutical industry spends a lot of money on advertising and has employed several promotional techniques to boost medicine sales. Medical representatives (MRs) are the essential professionals involved in product promotion in this environment. According to statistics, medical representative visits cost \$20.4 billion in the USA in 2004, accounting for 35.5% of spending on promotions [6]. Approximately 84% of pharma marketing is aimed at doctors. Free samples, advertisement in journals, and MR visits are all examples of this [7].

The most common sales strategy is still face-to-face detailing [8]. Between 1999 and 2001, worldwide sales forces increased by 21% [9]. Over the last decade, the number of pharmaceutical sales agents has increased [10]. Due to rising MR expenses, the effectiveness of face-to-face detailing is dwindling [11, 12]. In the USA, the expense of detailing can range from 45 to 70% of total marketing budget [13]. As a substitute to face-to-face detailing, pharmaceutical corporations have created electronic detailing (e-detailing) using ICT in pharma marketing [14–16]. Medical representatives (MRs) from pharmaceutical companies that visit doctors to acquaint them on pharma products have a relatively brief visit time [17, 18].

Representatives from companies arrive with their work bags, or rather detailing bags, and graphic aids, which they use to market the items to physicians. Visual aid is a traditional promotion tool for products that consists of a set of hard pages that are occasionally plastic coated or laminated and cover all the items/brands that a firm is marketing. A page is usually created for a particular product, and the entire visual aid is alphabetically separated or split [19]. Even though medical representatives have

the potential to influence their prescribing decisions, physicians' welcome MR visits and receive gifts, free samples, and other forms of provisions to be standard practice [20].

Electronic detailing (e-detailing) refers to pharmaceutical companies' use of digital technology to communicate information about their goods to physicians [21]. According to Banerjee and Dash [22], the professional service representative (PSR) visit, invitation, and detailing type all have a beneficial effect on the likelihood of e-detailing acceptance. Furthermore, it can be demonstrated that increasing the PSR activities increases the likelihood of physicians accepting e-detailing.

In India, physicians seem to view e-detailing as more of a complementary tool for ensuring a comprehensive effect of pharmaceutical marketing communication. Physicians are becoming more interested in using social media and other digital technologies because of digital marketing [23]. Pharmaceutical enterprises can benefit from e-detail, which is enabled by current ICT, to improve promotion efficiency [24]. Pharmaceutical companies' accessibility to digital marketing has yet to be addressed [25]. E-detailing will cut sales force time, lower costs, and increase physician prescription time [26].

3 Objective

The objective of the study is to determine if preference of the physicians for traditional detailing is an outcome of select benefits sought.

4 Methodology

The research is based on empirical data and is intended to be conclusive. The study's goal is to see if physicians' preference for traditional detailing is influenced by the benefits they seek. Seven benefits that are frequently requested by physicians have been identified for this purpose, and they are listed in the table along with their respective codes. One statement sought preference for traditional detailing (as opposed to e-detailing)—I prefer traditional detailing over e-detailing. All responses were implored on a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (Strongly agree). The survey's population consisted of physicians from eastern India. Non-allopathic practitioners were not included in the study. In total, 500 physicians were asked to respond. There were 81 people who did not respond, and 35 questionnaires were found to be incomplete. As a result, the effective sample size was 384. Personal responses were gathered and recorded. For the analysis, regression was used, with the preference for traditional detailing as the dependent variable and the benefits sought as the independent variable.

5 Results and Discussion

In order to accomplish the objectives of the study, seven variables were taken. Description of the variable ID given in Table 1.

ANOVA was conducted to check whether the mean of the two groups of physicians (one who preferred traditional detailing and the other who did not prefer traditional detailing) were similar or not. The results of the analysis in Table 2 show that the F value is 1155.586 which is significant (p is less than 0.05). Hence, we can conclude that there is significant difference between the preferences of physicians for detailing methods. The fitness of model, therefore, stands established.

Table 3 gives the model summary of the regression analysis that was performed. R square is the goodness-of-fit measure of regression models. A high value of adjusted

Table 1 Description of variables

Sl.no	Description	Variables
1	Appointment at convenient time	ACT
2	Appointment at convenient day	ACD
3	Social interaction with MR	SRM
4	Longer time to interact	LTI
5	Possible technology failures	PTF
6	Improved information	II
7	Personalized information	PI

Table 2 ANOVA

ANOVA						
Model		Sum of squares	df	Mean square	F	Sig
1	Regression	942.250	7	134.607	1155.586	0
	Residual	49.506	425	0.116		
	Total	991.755	432			

a. Dependent variable: Traditional

b. Predictors: (Constant), PI, ACT, SRM, II, PTF, LTI, ACD

Table 3 Model summary

Model summary				
Model	R	R square	Adjusted R square	Durbin-Watson
1	0.975	0.95	0.948	1.962

a. Predictors: (Constant), PI, ACT, SRM, II, PTF, LTI, ACD

b. Dependent variable: Traditional

Table 4 Regression coefficients

Coefficients						
Model		Unstandardized coefficients	Standardized coefficients			
		B	Std. Error	Beta	t	Sig
1	(Constant)	-0.364	0.124		-2.924	0.004
	ACT	0.02	0.045	0.021	0.434	0.664
	ACD	0	0.077	0	-0.002	0.999
	SRM	0.389	0.035	0.314	11.096	0
	LTI	-0.065	0.045	-0.064	-1.429	0.154
	PTF	0.327	0.038	0.272	8.702	0
	II	0.065	0.048	0.074	1.37	0.171
	PI	0.473	0.021	0.473	22.153	0

a. Dependent variable: Traditional

R square (0.948) indicates that the model explains 94.8% of all the variation in the response variable around its mean.

The coefficients of regression analysis are given in Table 4. The variables ACT, ACD, LTI, and II have a high p value (>0.05) hence these variables are not significant. However, the variables SRM, PTF, and PI have p value (<0.05) and are significant. Hence, the regression equation can be written as:

$$\text{Traditional} = -0.364 + (0.389 * \text{SRM}) + (0.327 * \text{PTF}) + (0.473 * \text{PI}).$$

6 Conclusion

Traditional detailing appeared under threat with the arrival of e-detailing possibility made possible through the wide penetration of internet among physicians in particular. The first wave of e-detailing through emails failed miserably as doctors reacted strongly against bombardment of emails carrying detailing messages. Traditional detailing hence remained the norm with the modification that medical representatives carried internet-backed detailing power point presentations. Arrival of COVID-19, however, put traditional detailing under tremendous constraints. As more sophisticated e-communication tools were available now to the pharmaceutical industry e-detailing flourished. This study was undertaken to find if some physicians still prefer traditional detailing. The objective of the study is to determine if preference of the physicians for traditional detailing is an outcome of select benefits sought. From the result and discussion, it is obvious that preference for traditional detailing is

strongly linked to certain benefits, important for the physicians. The important benefits in this respect being (a) social interaction with MR, (b) possibility of technology failures, and (c) personalized information. Of these, the third one, i.e., personalized information emerged as the most dominant benefit for preferring traditional detailing.

Gönül and Carter [27] discovered that increasing both e-detailing and traditional detailing benefits prescription product (Rx) manufacturers. Their findings also suggest that there is room for improvement in the synergy between the two types of detailing. According to study by Verma [28], the norm of reciprocity is followed by the pharma industry. They try to help the people who play an important role in their growth. If the company assists them in achieving their goals, no matter how small, the companies will undoubtedly benefit from their strategy [19].

As traditional detailing is an exorbitantly expensive component of pharmaceutical marketing, pharmaceutical companies are keen to enhance component of e-detailing in promotion of their products. Implications for them is that such campaigns be personalized to the maximum extent. For the medical representatives also, there are important implications. Not all are happy with virtual interactions and hence medical representatives must leverage upon the social visits and continue to provide personalized promotions. Additionally, companies can successfully identify and categorize physicians into two groups. One that is comfortable with e-detailing and the other that prefers traditional detailing. This will help in reducing the cost and in maximizing physician satisfaction.

References

1. Montoya R, Netzer O, Jedidi K (2010) Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability 29(5): 909–924
2. Kremer STM, Bijmolt THA, Leeflang PSH, Wieringa JE (2008) Generalizations on the effectiveness of pharmaceutical promotional expenditures. *Int J Res Mark* 25(4):234–246
3. Spiller LD, Wymer WW Jr (2001) Physicians' perceptions and uses of commercial drug information sources: an examination of pharmaceutical marketing to physicians. *Health Mark Q* 19(1):91–106
4. Banerjee S, Dash SK (2014) Factors influencing adoption of e-detailing as a communication tool: views of physicians of India. *Int J E-Health Med Commun* 5(3):29–39
5. Harindranath RM, Sivakumaran B (2021) Pharmaceutical promotion: a literature review and research agenda. *New Tech Brand Manag Healthc Sect* 44–63
6. Gagnon M-A, Lexchin J (2008) The cost of pushing pills: a new estimate of pharmaceutical promotion expenditures in the United States. *PLOS Med* 5(1):e1
7. Marco CA, Moskop JC, Solomon RC, Geiderman JM, Larkin GL (2006) Gifts to physicians from the pharmaceutical industry: an ethical analysis. *Ann Emerg Med* 48(5):513–521
8. Molloy W, Strang D, Guyatt G, Lexchin J, Bédard M, Dubois S, Russo R (2002) Assessing the quality of drug detailing. *J Clin Epidemiol* 55(8):825–832
9. Trucco MY, Amirkhanova S (2006) Transforming pharmaceutical marketing through e-detailing: case studies and recommendations. In: *The 8th IEEE international conference on e-commerce technology and the 3rd IEEE international conference on enterprise computing, e-commerce, and e-services (CEC/EEE'06)*, p 25
10. Boehm EW, Brown EG, Molvar K (2001) Pharma's detailing overhaul, *Forrester Rep*

11. Ventura K, Baybars M, Dedeoglu AO (2012) A new debate for Turkish physicians: e-detailing. *Health Mark Q* 29(4):362–377
12. Montoya ID (2008) E-detailing: information technology applied to pharmaceutical detailing. *Expert Opin Drug Saf* 7(6):635–641
13. Bates A, Bailey E, Rajyaguru I (2002) Navigating the e-detailing maze. *J Med Mark* 2(3):255–262
14. Davidson T, Sivadas E (2004) Details drive success. Physicians are responding to electronics sales calls. *Mark Health Serv* 24(1):20–25
15. Alkhateeb FM, Doucette WR (2009) Influences on physicians' adoption of electronic detailing (e-detailing) 34(1):39–52
16. Kwak E-S, Chang H (2016) Medical representatives' intention to use information technology in pharmaceutical marketing. *Healthc Inform Res* 22(4):342–350
17. Gleason M (2001) Internet detailing opens the doctor's door as this case study shows, the use of cyberspace technology to reach target physicians can give companies a competitive edge with a positive return on investment. *Med Mark Media* 36(1):80–87
18. McKillen D (2002) E-detailing gaining acceptance among physicians an e-detailing study conducted by Manhattan research shows that physicians are embracing this mode of detailing in increasing numbers. *Med Mark Media* 37(9):10–13
19. Nandy M, Pal B (2015) Pharmaceutical marketing and product promotion: a paradigm shift in Indian pharmaceutical industry (IPI). *Adhyayan A J Manag Sci* 5(02):27–40
20. Al-Areefi MA, Hassali MA, Ibrahim MIBM (2013) Physicians' perceptions of medical representative visits in Yemen: a qualitative study. *BMC Heal Serv Res* 13(1):1–8
21. Alkhateeb FM, Khanfar NM, Loudon D (2009) Physicians' adoption of pharmaceutical e-detailing: application of rogers' innovation-diffusion model 31(1):116–132
22. Banerjee S, Dash SK (2011) Effectiveness of e-detailing as an innovative pharmaceutical marketing tool in emerging economies: views of health care professionals of India 11(3):204–214
23. Parekh D, Kapupara P, Shah K (2016) Digital pharmaceutical marketing: a review. *Res J Pharm Technol* 9(1):108–112
24. Zhadko SV, Mostafa A (2018) The possibilities of using information and communication technologies in the work of medical representatives
25. Jawaid M, Ahmed SJ (2018) Pharmaceutical digital marketing and its impact on healthcare physicians of Pakistan: a national survey. *Cureus* 10(6)
26. Nalini R, Alamelu R, Amudha R, Badrinath V, Kumar DG (2017) E-detailing—empowering doctors in digital era. *Res J Pharm Technol* 10(8):2663–2667
27. Gönül FF, Carter FJ (2009) Impact of e-detailing on the number of new prescriptions. *Heal Care Manag Sci* 13(2):101–111
28. Verma SK (2004) Physician-pharmaceutical industry interaction: changing dimensions and ethics. *Indian Pediatr* 41(1):29–36

Perceptual Difference Between Patient and Physician on Negative Aspects of Social Media Promotion



Samrat Kumar Mukherjee , Jitendra Kumar , Vivek Pandey, Jaya Rani Rani, Abhijit Sarkar, and Ajeya K. Jha 

Abstract It is illegal in India to promote prescription drugs to patients through social media. On the other hand, some medical social media pages have misused this law. For health-related information, patients are increasingly turning to the Internet, and especially pharmaceutical pages, for help. This article presents various points of view. Differences in perception can lead to more disagreement between patients and doctors, so this is important to address. Around 1500 people and 400 doctors were interviewed. The perspectives of patients and doctors differ significantly. For this reason, patients and doctors work as a team to find health solutions, and their perceptions for the disadvantages of SMP may be so divergent that the association may possibly fail.

Keywords Social media · Physician · Doctor · Patients · Perceptual differences · Prescription drugs

1 Introduction

Currently, the efforts of the social media promotion (SMP) are limited to over-the-counter drugs for the treatment of common diseases such as colds and fevers. Social networks have emerged as an important source of knowledge in the course of the technological revolution, but there are numerous ethical and legal issues to consider before marketing prescription drugs. That's why several nations, such as India (Magic Medicines and Remedies Act, 1954), that have banned the use of social media to advertise prescription drugs, but patients can instantly access all pharmaceutical social media pages from anywhere in the world. Because of this, numerous social media sites are very active in the US and New Zealand. An Indian or other patient can easily view these pages; it is not uncommon for patients to benefit

S. K. Mukherjee (✉) · V. Pandey · J. Rani Rani · A. Sarkar
Sikkim Manipal Institute of Technology, Sikkim Manipal University, Gangtok, India
e-mail: Samrat.k@smit.smu.edu.in

J. Kumar · A. K. Jha
Sikkim Manipal University, Gangtok, India

from the information on these pages while at other times it can be detrimental to their well-being [1].

What do patients and doctors think about this development? Is there a significant difference between disadvantages and problems of this newly created phenomenon? Since the relaxation of the perception could lead to greater disputes between patients and doctors, this is a crucial problem to ask. This article examines the depth problem.

2 Literature Review

Patients can benefit or suffer from advertising on social networks for prescription drugs, so many studies. While many researchers in US informed it is a valuable source of information, others believe that it could be harmful to patients. Concern among the patients regarding their health have increased exponentially [2]. The public seek health information through social media, so they can follow the recommendations of physicians and discuss with their colleagues to confirm their decisions. India's SMP is a key factor in pharmaceutical marketing, according to [3].

It was suggested to many academics throughout the world that they study the good and negative elements of SMP by interviewing patients and physicians.

There are restrictions on the use of social media to market prescription drugs. One of the biggest drawbacks to social media is the inability to verify the validity of health and medication information [4]. It is very difficult to understand the reliability of information related to health on social media [5].

On social media, there is a large volume of prescription drug information that is difficult to verify [6]. Patients' personal information may still be accessed, and some researchers are concerned about privacy and potential issues [7]. Social media is used by many patients to acquire information on prescription drugs, but they are often ignorant of the hazards of sharing personal information online and receiving dangerous or incorrect advice from social media sources [8].

Accessible information on social media makes it difficult for patients to decide which drugs are best or how to use them [8]. A patient's health can be affected by incorrect information posted on social media sites such as: B pictures of smokers [9]. As a result of the spread of false information on Twitter, Nigerians were told to drink more salt water to prevent Ebola [1]. Social media have been used as a source of harmful behavior, such as sexual engagement [10].

Due to social media advertisements for prescription drugs, patients may delay visiting the doctor [11]. Due to their different advantages and disadvantages of SNP, a patient and a doctor work together as a team to heal patients [12]. The doctor influence patient relationships by promoting fair communication and submissive and pleasant contact between patients [13].

The time is called the main obstruction of the provision of medical care through several studies [14], and many doctors agree. It is possible that patients are endangered, even if they are aware of the risk. For doctors considerably easier for the distinction between safe and dangerous SMPs. Correction of a recent study [15],

patients and doctors vary on this topic. It is possible that the doctor-patient relationship suffers from this perspective due to this discrepancy.

Prescription medications are surveyed for their virtues and demerits, and the results of this study are based mostly on SMP of prescription drugs. For this reason, the focus has been on this critical research gap, which has been overlooked in most studies in this subject.

3 Methodology and Data

To find out how social media affects the opinions of patients and doctors (allopathic medicine) in India, we conducted exploratory research. For this study, allopathic doctors and patients were interviewed to collect data. A descriptive technique was used in conjunction with a framework of inference. None of the research indicates that there are significant differences in the perception of the cons of prescription drugs with SMP by patients and doctors. After questioning doctors and patients about changes in the SMP, various hypothesis indicators were created. In order to determine whether the results of this exploratory study were excellent or detrimental, specific reasons were sought, which were used in this study to identify differences in perception between patient and patient. As a logical consequence, we have created the hypotheses Ho1–Ho4:

Ho1: There are no significant differences between physicians and patients in their perception on impact of SMP on self-medication.

Ho2: There are no significant differences between physicians and patients in their perception on impact of SMP on unethical activities.

Ho3: There are no significant differences between physicians and patients in their perception on impact of SMP on negative patient behavior.

The study is done in India, and respondents are doctors and patients from allopathic background. Total 1500 patients and 400 physicians took on a survey participate. 91 patients and 10 doctors considered that the answers are not valid, resulting in a final sample size of 1409 patients and 390 physicians.

Sampling methods were used in the investigation; the names of the doctors are chosen at random from the list of doctors in a city; on the contrary, the patients were approached in hospitals. Additional patient knowledge and use of social networking sites, including health-related sites, were assessed. The study only included patients who said they frequently visited health-related pages on social media.

The Cronbach-Alpha was used to evaluate the reliability of the interval fragment sheet. With the comparison of the data from the doctor and the patient, the calculated alpha value of Cronbach was 0.841 and 0.812, which is considered high reliability of more than 0.6, and therefore, the results of the questionnaire should be considered reliable.

Statistical analysis was carried out using SPSS (V 25.0). A 95% level of confidence is required to test the hypothesis. Patients' and doctors' perceptions were compared using a t-test. t-tests are considered relevant when their Z value reaches 1.96, which is less than 0.05.

4 Result

Ho1: There are no significant differences between physicians and patients in their perception on impact of SMP on self-medication.

Because of SMP; patients buy prescription medicines on their own (bypassing the physician): It is evident from Table 1 that physicians are more inclined than patients to believe that social networks have an influence on self-medication, with a mean answer of 4.95. The t-value of 1.71 is substantially lower than the tabular value of 1.96, indicating a major disagreement. When it comes to self-medication, consumers may believe social media content is safe, but professionals do not.

SMPs demotivate some patients from consulting physicians: A patient's average response is 4.39, while a doctor's average answer comes in at 4.05 (Table 1). This suggests that doctors are more inclined than patients to feel that information obtained via the use of social networks demotivates people from contacting their physician.

Ho2: There are no significant differences between physicians and patients in their perception on impact of SMP on unethical activities.

Validating information on SMPs is a challenging task: It is evident from Table 2 that physicians are more inclined than patients to believe that social networks have an influence on self-medication, with a mean answer of 5.21. The t-value of 6.3 is substantially higher than the tabular value of 1.96, indicating a major disagreement. Doctors do not trust the reliability of the information available on the social media.

People with vested interest can infiltrate SMPs: According to Table 2, the patient's mean response is 5.19 while the doctor's mean response is 5.36, suggesting that physicians are more likely than patients to believe that people with vested interest can infiltrate SMPs.

There is high possibility of misuse of personal data of patients provided to SMPs: According to Table 2, the patient's mean response is 5.19 while the doctor's mean response is 5.36, suggesting that physicians are more likely than patients to

Table 1 Patient-physician differences in belief on self-medication

Patient-physician independent sample t-test (unethical)		
Category		Mean
Validating information on SMPs is a challenging task	Patient	4.54
	Physician	5.21
People with vested interest can infiltrate SMPs	Patient	5.19

Table 2 Patient-physician differences in belief on unethical activities

Patient-physician independent sample t-test (unethical)					
Category		Mean	Std. deviation	t-value	p-value
Validating information on SMPs is a challenging task	Patient	4.54	1.86	6.3	0
	Physician	5.21	1.69		
People with vested interest can infiltrate SMPs	Patient	5.19	1.67	1.98	0.071
	Physician	5.36	1.58		
There is high possibility of misuse of personal data of patients provided to SMPs	Patient	5.19	1.65	1.98	0.065
	Physician	5.36	1.6		

believe that there is high possibility of misuse of personal data of patients provided to SMPs.

Ho3: There are no significant differences between physicians and patients in their perception on impact of SMP on negative patient behavior.

SMP driven patients' waste precious time of physicians: According to Table 3, the patient's mean response is 4.18 while the doctor's mean response is 5.41, suggesting that physicians are more likely than patients to believe that SMP driven patients' waste precious time of physicians.

SMP driven patients disobey their physicians' instructions: According to Table 3, the patient's mean response is 4.21 while the doctor's mean response is 5.65, suggesting that physicians are more likely than patients to believe that SMP driven patients disobey their physicians' instructions.

SMP driven patients insist on being prescribed a specific drug/brand: According to Table 3, the patient's mean response is 4.21 while the doctor's mean response is 5.65, suggesting that physicians are more likely than patients to believe that SMP driven patients insist on being prescribed a specific drug / brand.

Table 3 Patient-physician differences in belief on negative patient behavior

Patient-physician independent sample t-test (NP)					
Category		Mean	Std. deviation	t-value	p-value
SMP driven patients' waste precious time of physicians	Patient	4.18	1.7	12.98	0
	Physician	5.41	1.55		
SMP driven patients disobey their physicians' instructions	Patient	4.21	1.7	15.07	0
	Physician	5.65	1.53		
SMP driven patients insist on being prescribed a specific drug/brand	Patient	4.21	1.69	15.3	0
	Physician	5.66	1.54		
SMP driven patients tend to ask too many questions	Patient	4.21	1.68	12.9	0
	Physician	5.43	1.59		
SMP driven patient become argumentative	Patient	4.21	1.69	12.56	0
	Physician	5.41	1.58		

SMP driven patients tend to ask too many questions: According to Table 3, the patient's mean response is 4.21 while the doctor's mean response is 5.43, suggesting that physicians are more likely than patients to believe that SMP driven patients tend to ask too many questions.

SMP driven patient become argumentative: According to Table 3, the patient's mean response is 4.21 while the doctor's mean response is 5.41, suggesting that physicians are more likely than patients to believe that SMP driven patient become argumentative.

5 Discussion

From review of literature, it is clear that one of biggest disadvantage of using social media to search health information is self-medication. However, it is crucial to evaluate the disagreements between patients and doctors. We clearly conclude (Table 1) that the null hypothesis is rejected because physicians consistently assign higher beliefs to SMP to the two variables we examined than their medical counterparts indicator. According to some study [12], social media might lead to self-medication of patients. This is where the role of social media cannot be overlooked; there are instances where people have suffered due to following the suggestion given on these media [7].

Overall, we discovered that both patients and doctors feel that SMP can lead to unethical behavior, although there is debate on the extent of these effects. These changes are less pronounced than those observed with other beneficial features of SMP. What are the potential consequences of these differences? SMP has been shown to be deceptive [16]. This might result in misconceptions and information discrepancies. Clinicians report losing time during patient visits clearing up misconceptions caused by SMP [17].

Patient conduct is highly important since patients might get argumentative and squander doctors' valuable time [6]. The null hypothesis is refuted in Table 3 because doctors consistently attribute greater beliefs to SMP than their medical colleagues to the five variables we investigated.

6 Conclusion

As a result of the growth of social media sites, patients and physicians may hold opposing views on the drawbacks of social media promotion. If this trend continues, it will have a negative impact on the whole healthcare system. Researchers and authorities were unaware of the massive disparity, and no attempt was made to correct it. As a consequence of the study's findings, authorities will be able to better comprehend patients and physicians and bridge the gap between their divergent viewpoints on prescription medicine social media promotion.

In order to provide health support, patients and experts must work together. With the growth of health blogs on the Internet, it is expected that there would be significant differences in their perspectives on the benefits and drawbacks of SMP. When this occurs, the essential context is affected, which can have serious consequences for the whole health system. Researchers and authorities were unaware of the massive disparity, and no attempt was made to correct it. Patients and doctors have different perspectives about SMP of prescription medicines, which is why this study is so important. It assists patients and physicians in better understanding one another and bridging the gap between their differing perspectives on SMP.

Patients, the third major player in this regard, must also learn to curb their enthusiasm for SMP and recognize their limited ability to understand information and the likelihood of misinformation (including withholding unfavorable information) in promotional materials. You need to learn to appreciate the multiple viewpoints of patients and the merits of certain beliefs of doctors in general or in specific circumstances. A doctor is only as good as the trust a patient has in him. A violation of this belief by SMP will have significant ramifications for everyone involved.

References

1. Carter M (2014) How twitter may have helped Nigeria contain Ebola. *BMJ: British Med J* (Online) 349
2. Schroeder CM (2010) The era of the empowered patient. *DTC Perspectives*, 21–23
3. Jha A (2005) A study on the role of information technology in marketing pharmaceutical products
4. Kukreja P, Sheehan AH, Riggins J (2011) Use of social media by pharmacy preceptors. *Am J Pharm Educ* 75(9):176
5. Tian Y (2010) Organ donation on web 2.0: content and audience analysis of organ donation videos on youtube. *Health Commun* 25(3):238–246
6. Mukherjee SK, Kumar J, Jha AK, Rani JR (2019) Role of social media promotion of prescription drugs on patient belief-system and behaviour. *Int J E-Collab (IJeC)* 15(2):23–43
7. Mukherjee SK, Kumar JK, Jha AK, Pandey JR (2021) Dynamics of social media promotion of prescription drugs and resulting patient belief systems. In: *E-collaboration technologies and strategies for competitive advantage amid challenging times*. IGI Global, pp 144–170
8. Adams SA (2010) Revisiting the online health information reliability debate in the wake of “web 2.0”: an inter-disciplinary literature and website review. *Int J Med Inform* 79(6):391–400
9. Freeman B, Chapman S (2007) Is “YouTube” telling or selling you something? tobacco content on the youtube video-sharing website. *Tob Control* 16(3):207–210
10. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA (2009) Display of health risk behaviors on MySpace by adolescents: prevalence and associations. *Arch Pediatr Adolesc Med* 163(1):27–34
11. Kim S (2009) Content analysis of cancer blog posts. *J Med Libr Assoc: JMLA* 97(4):260
12. Pandey JR, Saha SK, Mukherjee SK, Jha A (2019) Self-medication and direct-to-consumer promotion of drugs. In: *International conference on communication, devices and networking*. Springer, Singapore, pp 159–166
13. Smalhodzic E, Hooijsma W, Boonstra A, Langley DJ (2016) Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. *BMC Health Serv Res* 16(1):442

14. Volpp KG, Mohta NS (2017) Social networks to improve patient health advisor analysis. *NEJM Catal*
15. Jha A, Pandey JR (2017) An empirical note on health information digital divide: a study of Indian patients. *Int J Asian Bus Inform Manage (IJABIM)* 8(2):15–34
16. Young HN, Paterniti DA, Bell RA, Kravitz RL (2005) Do prescription drug advertisements educate the public? the consumer answers. *Drug Inf J: DIJ/Drug Inf Assoc* 39(1):25–33
17. Bozic KJ, Smith AR, Hariri S, Adeoye S, Gourville J, Maloney WJ, Rubash HE (2007) The 2007 ABJS Marshall Urist Award: the impact of direct-to-consumer advertising in orthopaedics. *Clin Orthop Related Res* 458:202–219

Impact of Ethicality and Marginalized Group Convenience on Social Media Promotion of Branded Drugs



Samrat Kumar Mukherjee , Jitendra Kumar , Abhijit Sarkar, Bibeth Sharma , Jaya Rani Rani, and Ajeya K. Jha 

Abstract The use of social media to promote prescription drugs is controlled by ethical and legal behavior and practice standards. Only two countries, the USA and New Zealand, have made it lawful to contact patients directly in order to sell medications. This reasonable approach is no longer applicable due to technological developments. The use of social media for health information, including drug marketing, is gaining popularity. This development has been lauded by a wide spectrum of organizations. This is owing to the fact that it provides a wide range of advantages. The advantages of this form of advertising are widely established. According to some studies, SMP is related to many factors such marginalized group can get lots of benefit using social media and patients' party think that it is ethical and nothing wrong can happen while sharing personal information in social media. This study has significant implications for marketers, physicians, and policymakers who play critical roles in promoting the creation of safe systems to improve health systems. Future research should use an enhanced communication model tailored to the social media environment to ensure a thorough investigation.

Keywords Social media promotion · Social media health information · Marginalized group · Ethical aspect

1 Introduction

The promotion of prescription drugs through social media is still subject to severe ethical and legal restrictions. Many nations have banned any kind of direct promotion of drugs. Actually in Asia not even a single country has given a legal nod to DTCP in the pharmaceutical business. The Drug and Cosmetic Rule, 1945 in India forbids the general public from promoting the diagnosis, healing, alleviation, and treatment

S. K. Mukherjee (✉) · J. R. Rani
Sikkim Manipal Institute of Technology, Sikkim Manipal University, Gangtok, India
e-mail: Samrat.k@smit.smu.edu.in

J. Kumar · A. Sarkar · B. Sharma · A. K. Jha
Sikkim Manipal University, Gangtok, India

described in the program, which are laid down in the regulations of this law. Violation of this legislation is a felony punishable by up to six months in prison, a fine, or both. If convicted, the sentence might be increased to one year.

The use of social media to promote prescription drugs on the market is still subject to stringent ethical and legal constraints. In certain nations, direct drug marketing is prohibited. Even in Asia, several nations do not recognize the DTCP's legal position in the pharmaceutical sector. According to Indian law, it is prohibited for the general public to advocate for the program's diagnosis, healing, or treatment, or for any other disease, affliction, or condition (by whatever name it is called) mentioned in the statutes.

However, with the emergence of social media, everything has changed. As a result, the entire globe has shrunk to the size of a hamlet. As a result, patients now have quick access to all of the information they require about their ailments and therapies. Furthermore, depending on the symptoms stated on social media sites, people may self-diagnose and then select the best appropriate prescription for the illness at hand. The government will not be able to prevent the patient party from delving into the ocean of information. There is no stopping social media promotion, as plain as day (SMP).

As a result of these advancements, today's patients are well-informed. They seek knowledge not only out of curiosity, but also so that they may discuss their treatment options with doctors. Various studies have confirmed this [1, 2]. Patient information is available on the Web sites of pharmaceutical companies [3]. Many businesses have established patient communication channels to facilitate patient interaction on their company, disease, and product (brand) Web sites [4].

2 Review of Literature

Many studies have been conducted to evaluate whether or not promoting of medicine (not OTC) via social media is actually useful to the patient party. Some researchers believe it is a valuable source of information, while others say it may be harmful to patients. According to [5], SMP empowers the patient by providing health information, healthcare alternatives, and medical growth. The general population, according to [6], is growing increasingly concerned about their medications. They undertake further research before consulting with a doctor so that they can follow their doctors' advice and consult with their colleagues to validate their judgments. SMP, according to scientists, has become a significant component of pharmaceutical marketing in India [4, 7].

Numerous researchers have reported that the personality traits of social media consumers differed significantly between different groups. Researchers discovered a huge digital gap in access to health information, putting vulnerable people at a disadvantage [8]. According to some study, there were significantly fewer guys than girls accessing social networking sites [1, 9]. It is been demonstrated in few studies that social media users tend to originate from lower-income families [10], which

leads to conclusions like education does not affect the capacity to look for health information on social media. According to some researchers, more females are using social media to search for health information in urban areas [1]. People who use social media are between 11 and 34 years old [11, 12]. Especially during this pandemic, geriatric patients and the ones suffering from the chronic disease had to use social media as a telehealth tool for their periodical consultation with their physician as they were unable to go and visit their physician due to fear of infection [1].

Information about prescription medications is routinely distributed on social media, but it is difficult to verify [13, 14]. Although some academics are concerned about privacy and other issues; still some patients' party sometime shares their personal information [1, 13]. Especially adolescents and other marginalized are not worried with the risks connected with sharing personal information in social media and receiving erroneous or harmful advice from social networking platforms [13].

3 Methodology and Data

According to the research technique of this study, social media promotion is associated with marginalized group and ethical. To collect data for this study, patients getting allopathic treatment were surveyed.

H_0 : SMP is not at all associated with marginalized group.

H_{01} : SMP is not at all associated with ethical.

Table 1 gives the values of R and R^2 . The R value (column "R") shows a simple correlation and is 0.398, indicating a high degree of connection. The number R^2 (in the column labeled "R squared") indicates how well the independent variable social media promotion describes total variance in the dependent variable marginalized group. Only 15.8% of the variance can be accounted for in this case, which is exceedingly minimal. Furthermore, the Adjusted R^2 (0.156) value is quite near to the R^2 value (0.158), indicating that the model is fit.

Table 1 Model summary: marginalized group

Model	R	R square	Adjusted R square	Std. error of the estimate
1	0.398	0.158	0.156	8.00083

Table 2 ANOVA: marginalized group

Model		Sum of squares	df	Mean square	F	Sig
1	Regression	3170.016	1	3170.016	72.696	0.000
	Residual	16,875.696	387	43.606		
	Total	20,045.712	388			

As demonstrated in Table 2, the regression model predicts the dependent variable with $p < 0.0005$ less than 0.05 accuracy and statistically predicts the outcome variable (That is, it matches the facts nicely). As a result, hypothesis H0 is rejected.

The coefficient table gives the data we need to anticipate the “marginalized group” of the “social networks” and determine whether the “social networks” are statistically significant added to the model (see column “Sig.”). Also, as shown below, we can express the regression equation using the values in column “B” under the column “Nonstandard Coefficients” as follows:

$$\text{Marginalized group} = 14.507 + 0.373(\text{Social Media}).$$

The “VIF” value in Table 3 is also crucial. If this number for all predictors is less than 10, the model is in good form.

Eigenvalues aid in the identification of multicollinearity problems (Table 4), here the value is less than 15 may imply that there are no multicollinearity difficulties (IBM 2019).

Table 5 gives the R and R² values. The R value (column “R”) shows a simple correlation and is 0.2, indicating a high degree of correlation. The number R² (in the column labeled “R squared”) indicates how well the independent variable social media marketing describes total variation in the dependent variable immoral acts. In this case, only 4% of the variance can be accounted for, which is incredibly low. Furthermore, the Adjusted R² (0.039) value is quite near to the R² value (0.040), indicating that the model is fit (Table 6).

Table 2 gives that the regression model properly predicts the dependent variable ($p = 0.0005$) as well as the outcome variable ($p = 0.05$) both conceptually and quantitatively (That is, it matches the facts nicely). As a result, hypothesis H01 is eliminated.

Table 3 Coefficients: marginalized group

Model		Unstandardized coefficients		Standardized coefficients	t	Sig	Collinearity statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	14.507	1.077		13.471	0.000		
	Socialmedia1	0.373	0.044	0.398	8.526	0.000	1	1

Table 4 Collinearity diagnostics: marginalized group

Model	Dimension	Eigenvalue
1	1	1.95

Table 5 Model summary: ethical

Model	R	R square	Adjusted R square	Std. error of the estimate
1	0.200	0.040	0.039	6.94730

Table 6 ANOVA: ethical

Model		Sum of squares	df	Mean square	F	Sig
1	Regression	2840.583	1	2840.583	58.854	0
	Residual	67,860.641	1406	48.265		
	Total	70,701.224	1407			

Table 7 Coefficients: ethical

Model		Unstandardized coefficients		Standardized coefficients	t	Sig	Collinearity statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	21.553	0.707		30.497	0.000		
	Socialmedia1	0.208	0.027	0.200	7.672	0.000	1	1

Table 8 Collinearity diagnostics: ethical

Model	Dimension	Eigenvalue
1	1	1.95

The coefficient table gives us the information we need to anticipate the... “ethical behaviors” of the “social networks” and determine whether the “social networks” are statistically significant added to the model (see column “Sig.”). Also, as shown below, we can define the regression equation using the values in column “B” under the column “Unstandardized Coefficients” as follows:

$$\text{Ethical Aspect} = 21.553 + 0.208(\text{Social Media})$$

The “VIF” value in Table 7 is also crucial. If this number for all predictors is less than 10, the model is in good form.

Eigenvalues aid in the identification of multicollinearity problems (Table 8), here the value is less than 15 may imply that there are no multicollinearity difficulties [15].

4 Conclusion

The study found that patients are attracted to the SMP because of its supports toward marginalized groups and ignorance toward unethical activity. Marginalized group uses social media for health information for different reasons. During this pandemic, many reputed organization came in-front to support people through social media. People started to perceive all of the social media are legit like them and share their personal information. They perceive it is ethical. Many other researchers also have

same opinion regarding this [1, 16–19]. In conclusion, SMP is believed to have a strong relationship with marginalized group and ethical aspect.

5 Future Scope of Study

The ramifications for policy makers are to make clear how marginalized group and ethical aspect are related to SMPs, which could help develop systems that can significantly improve the health system. In addition, to ensure a comprehensive study, future studies should use an extended communication model that is adapted to the context of social media.

References

1. Mukherjee SK, Kumar JK, Jha AK, Pandey JR (2021) Dynamics of social media promotion of prescription drugs and resulting patient belief systems. In: E-collaboration technologies and strategies for competitive advantage amid challenging times. IGI Global, pp 144–170
2. Volpp KG, Mohta NS (2017) Social networks to improve patient health advisor analysis. *NEJM Catal*
3. Mukherjee SK, Kumar J, Jha AK (2022) Analysis of Information about medicines available on facebook. *Int J E-Collab (IJeC)* 18(4)
4. Jha A (2005) A study on the role of information technology in marketing pharmaceutical products
5. Auton F (2004) The advertising of pharmaceuticals direct to consumers: a critical review of the literature and debate. *Int J Advert* 23:5–52
6. Schroeder CM (2010) The era of the empowered patient. *DTC Perspect*, 21–23
7. Pandey JR, Jha A, Saha SK (2019) Impact of manipulative character of direct-to-consumer promotion. In: Dynamic perspectives on globalization and sustainable business in Asia. IGI Global, 198–211
8. Jha A, Pandey JR (2017) An empirical note on health information digital divide: a study of Indian patients. *Int J Asian Bus Inf Manage (IJABIM)* 8(2):15–34
9. Jha A, Pandey JR, Mukherjee SK (2018) An empirical note on perceptions of patients and physicians in direct-to-consumer promotion of pharmaceutical products: study of Indian patients and physicians. In: Management strategies and technology fluidity in the Asian business sector. IGI Global, 65–87
10. Scanfeld D, Scanfeld V, Larson EL (2010) Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control* 38(3):182–188
11. Selkie EM, Benson M, Moreno M (2011) Adolescents' views regarding uses of social networking websites and text messaging for adolescent sexual health education. *Am J Health Educ* 42(4):205–212
12. Marcus M, Yasamy MT, van Ommeren MV, Chisholm D, Saxena S (2012) Depression: a global public health concern
13. Adams SA (2010) Revisiting the online health information reliability debate in the wake of “web 2.0”: an inter-disciplinary literature and website review. *Int J Med Inf* 79(6):391–400
14. Mukherjee SK, Kumar J, Jha AK, Rani JR (2019) Role of social media promotion of prescription drugs on patient belief-system and behaviour. *Int J E-Collab (IJeC)* 15(2):23–43
15. IBM (n.d.) Collinearity diagnostics., from https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/tutorials/reg_cars_collin_01.html., (2019)

16. Yoon H, Jang Y, Vaughan PW, Garcia M (2018) Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. *J Appl Gerontol*
17. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW (2018) Differences in access to and use of electronic personal health information between rural and urban residents in the United States. *J Rural Health* 34:s30–s38
18. Statt N (2020) Major tech platforms say they're 'jointly combating fraud and misinformation' about COVID-19. *The Verge*
19. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H (2020) The pandemic of social media panic travels faster than the COVID-19 outbreak

Tracking of Maximum Power of Solar PV Array Under Partial Shading Condition Using Grey Wolf Optimization Algorithm



Neeha Cintury, Soumyarupa Saha, and Chitragada Roy

Abstract In this paper, solar energy has been used to generate required maximum value of power from solar photovoltaic (PV) system under partial shading condition. A simulation study of Maximum Power Point Tracking (MPPT) controller for solar PV array under partial shading condition is shown in MATLAB using the two known techniques that is conventional Perturb and Observe (P&O) and a global search algorithm Grey Wolf Optimization (GWO). The result of both the techniques are compared to show effectiveness of GWO algorithm. The proposed GWO based MPPT controller extracts the maximum amount power from the solar photovoltaic (PV) system under partial shading condition by modifying the duty cycle of Boost Converter (DC-DC). It was observed that GWO gives oscillation free maximum value of power at a faster rate as compared to conventional P&O technique.

Keywords Solar photovoltaic (PV) system · Maximum Power Point Tracking (MPPT) · Perturb and Observe (P&O) · Grey Wolf Optimization (GWO)

1 Introduction

Conventional energy sources can neither be replaced nor can be re-energized. Burning of fossil fuels causes global warming. While fossil fuels are burnt, it emits harmful gases like nitrous oxide which ends up causing pollution, sulphur dioxide which gives rise to acid rain and greenhouse gases [1]. So, in order to avoid this disadvantageous environmental effect from these fuels and their depletion, renewable energy sources are taken into consideration knowing that they are more sustainable. As solar energy is easily available and eco-friendly, it is considered to be one of the best renewable energy sources [2]. Hence, solar energy is widely used as they have numerous health and environmental benefits.

Solar PV system is one of the most important energy system which can be used in future as they require minimum amount of maintenance and fuel cost compared

N. Cintury · S. Saha · C. Roy (✉)
Department of Electrical and Electronics Engineering, Sikkim Manipal Institute of Technology,
Sikkim Manipal University, Sikkim, India
e-mail: chitragada.r@smit.smu.edu.in

to other systems and also is productive in terms of producing huge range of output voltage and current. There are many MPPT techniques designed to track maximum amount of power. In [3], in order to draw out maximum amount of power in PV system with a dual MPPT controller, FLC-based P&O scheme has been implemented in order to design a new control strategy. In [4], simulation results of P&O-based and Fuzzy Logic Controller (FLC)-based MPPT are compared and designed to draw out amount of power from the solar PV array under non-uniform solar irradiation. It was found out by comparing both the methods that FLC-based MPPT gave better results with precise value of output response under variable irradiation than P&O-based controller, with increasing efficiency, stable output power as well as reduced fluctuations. In [5], a new 3-point Perturb and Observe method for MPPT of solar photovoltaic system has been designed by generating a controlled signal (gate pulse) for driving the DC Boost Converter which has been used to operate the load as well as the photovoltaic panel. Comparison has been carried out between the simulation result and conventional Hill Climbing (HC) method. It was found that 3-point P&O technique gave better result as compared to that of Hill Climbing-based controller under both changing as well as constant temperature and irradiation levels. In [6], a new method has been introduced which is hybridization of two well-known techniques, namely PSO and GWO algorithm. The fresh new version which is obtained is a better technique than individual GWO and PSO algorithms as it performs excellently providing solution to the search spaces.

Solar PV arrays faces different problems one of them is partial shading which is caused due to environmental conditions such passing clouds, shadow of tall buildings on one part of the PV array. In [7], bypass diodes has been used to prevent partial shading as well as it defines the effect of shading as well as the hotspot effect on solar PV arrays. This paper gives a detailed idea on the working of bypass diodes and discusses its impact on output power. In [8], the efficiency of the new hybrid method is observed where the hybridization of the two well-known algorithms are done, namely Grey Wolf Optimization and Perturb and Observe technique. As a result of hybridization, the extraction of the required amount of maximum power from the solar photovoltaic system has been enhanced with change in solar irradiation and under partial shading condition.

Thus, in this paper, MATLAB Simulink has been used so as to observe and analyse the problems faced under partial shading condition by the solar PV modules using a conventional method, Perturb and Observe (P&O), a global search algorithm, Thus, these techniques are implemented to track the maximum value of power from the PV arrays which has been discussed in details with the help of flow charts below.

2 System Modelling

The schematic diagram of the proposed MPPT Controller under partial shading condition has been shown in Fig. 1. Solar PV array is partially shaded i.e. it gets variable solar irradiation at constant temperature throughout the day. The solar PV

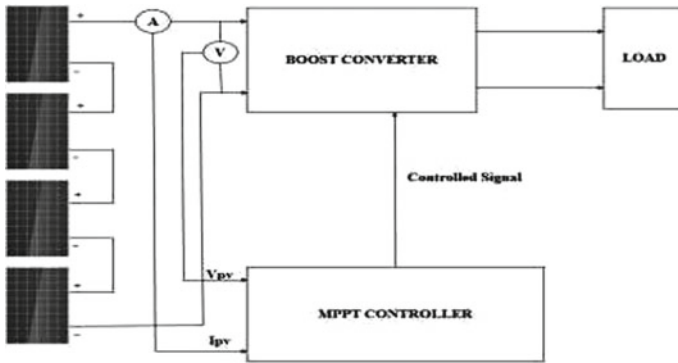


Fig. 1 Block diagram of the proposed MPPT controller for solar photovoltaic array under partial shading condition

array supplies voltage and current to the Boost Converter which has been used in order to get a stable DC voltage as output. The Boost Converter receives the controlled signal from the proposed MPPT controller. MPPT [9, 10] is a method often used in order to maximize the output power of photovoltaic PV system. Classical MPPT controller gives several local optima under partial shading condition. Hence, with the basic Maximum Power Point Tracking controller, the system is unable to track the global peak under partial shading condition due to trapping into local maximum peak point. So as to overcome the above-mentioned problems, it is suggested to use global optimization algorithms to modify the classical MPPT controller in order to get global power peak even at partial shading condition. In this work, firstly conventional P&O technique has been implemented in order to design the MPPT controller which failed to give us the global optima under partial shading condition so as to overcome this problem Grey Wolf Optimization (GWO) algorithm, which is global optimization algorithm have been implemented which generates a controlling signal for the power electronics switch used in the DC-DC Boost converter. Output of the converter has been used to drive the DC load. The objective is to get global optimal point which has given us the maximum amount of power for the solar photovoltaic array under partial shading condition.

2.1 Designing of Solar PV System

As large amount of power cannot be extracted from a single solar PV array, we have connected four solar PV arrays in series so as to get desired amount of voltage in order to satisfy our needs. Each panel is represented by a single cell diode.

In [8], the difference between the current source (I_p) and diode current (I_d) is represented by

$$I = I_p - I_o \left[\exp \left(\frac{qV + qR_{sum}I - 1}{N_s K_s T_a} \right) \right] - \frac{V + R_{sum}I}{R_l} \tag{1}$$

where

- I_p is the photovoltaic current source in Amps,
- D_p is the diode parallelly connected to the current source,
- R_{sum} is the total resistance (should be low as possible) of all the components in the current path in Ohm,
- R_l is the leakage (high as possible) over the P-N junction in Ohm,
- I_d is the diode current in Amps and I_0 is the saturation current in Amps,
- a is the ideality factor of diode,
- K_s is the Boltzmann's constant which is 1.381×10^{-23} J/K, is electron charge which is 1.602×10^{-19} C,
- T is the temperature in kelvin and N_s is the total number of cells in series.

In this paper, output power of the solar PV system is given by:

$$P_{mp} = V_{mp} * I_{mp} \tag{2}$$

where

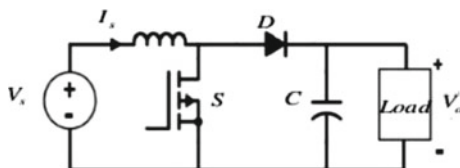
- P_{mp} is the maximum output power,
- V_{mp} and I_{mp} are the maximum values of voltage and current, respectively, at which maximum power is obtained.

2.2 Designing of the Boost Converter

In a DC-DC Boost Converter, the voltage at the output is always more than that of the voltage at input side, i.e. it steps up the DC voltage of the input side by stepping down the input current. An equivalent circuit of the Boost Converter has been shown in Fig. 2.

So, to design the Boost Converter two power electronic devices are used, first is the MOSFET whose gate gets triggered by the controlled signal sent by the MPPT controller and the other one is the diode. To decide the values of the inductance, capacitance and duty cycle, the following equations are used:

Fig. 2 Equivalent circuit of the boost converter



$$L = \frac{V_s}{F_s \cdot \Delta I_l} \quad (3)$$

$$C = \frac{I_o \cdot D}{F_s \cdot V_o} \quad (4)$$

$$D = \frac{V_o - V_s}{V_o} * 100 \quad (5)$$

The output of the boost converter is:

$$V_o = \frac{V_s}{1 - D} \quad (6)$$

where source or input voltage is V_s , F_s is the switching frequency = 50000 Hz, ΔI_l is the current ripple, V_o is output voltage and I_o is saturation current and D is duty cycle.

2.3 Designing of the MPPT Controller

2.3.1 Perturb and Observe (P&O) Based MPPT Controller

Conventional P&O technique is implemented in this paper using MATLAB for designing of MPPT controller in order to do a comparative analysis.

In [11], PV array is perturbed by making small changes until and unless maximum power point is achieved.

Working of P&O technique has been shown using a flow chart in Figs. 3 and 4.

2.3.2 Grey Wolf Optimization (GWO) Based MPPT Controller

Grey Wolf Optimization (GWO) comes under the modern heuristic optimization algorithm. In [12], Mirjalili et al. was first to recognized this algorithm in the year 2014. GWO technique has been mainly extracted from the idea of Grey wolves' lifestyle in terms of how they live, chase, hunt their preys, etc. This bunch of wild animals usually lives in a pack of 5–10 following very strict dominant hierarchy. In [13], hierarchy here signifies four different levels of leadership where in levels are denoted by α , β , δ and ω which is particularly maintained in a pyramid structure which shows α (alpha) as leader, β (beta) as sub leaders following with δ (delta) and ω (omega), ω being the lowest ranked leader. During a hunt, a prey is encircled by a pack of grey wolves and this manner of encircling can be achieved with the help of following equations:

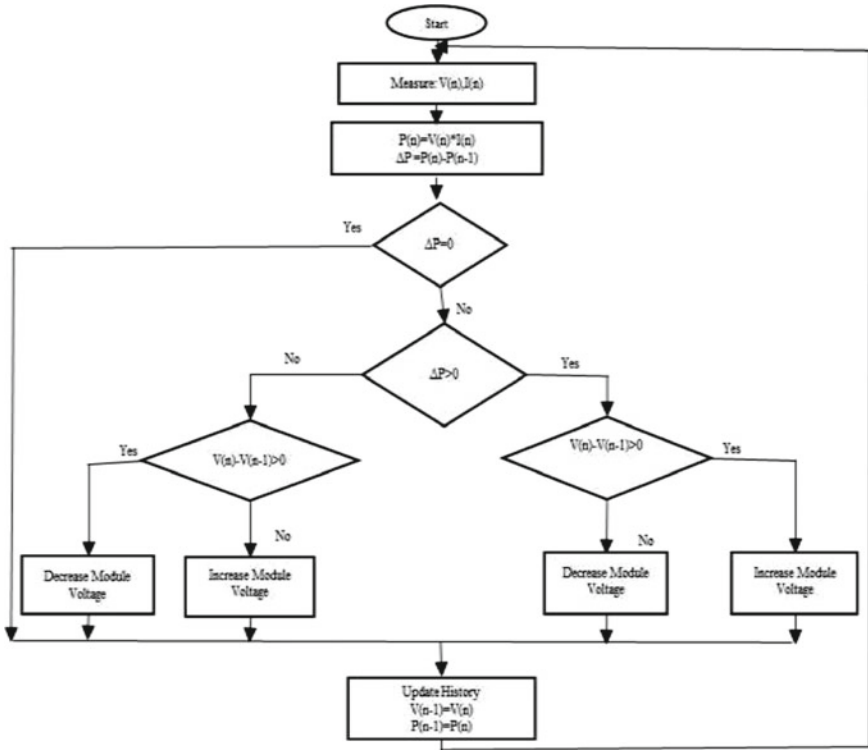


Fig. 3 Flow chart of conventional P&O technique

$$\vec{E} = \vec{G} \cdot \vec{X}_p(t) - \vec{X}_p(t) \tag{7}$$

$$\vec{X}(t + 1) = \vec{X}_p(t) - \vec{F} * \vec{E} \tag{8}$$

where current iteration is denoted by “t”, the vector coefficients are denoted by E , F , G , position vector of the prey is given by X_p and the position vector of grey wolf is denoted by X .

$$\vec{F} = 2\vec{p} \cdot \vec{r}_1 - \vec{p} \tag{9}$$

$$G = 2 \cdot \vec{r}_2 \tag{10}$$

The linear value above is denoted by “p” which decreases from two to zero and the random vectors [0, 1] are denoted using $r1$ and $r2$.

Further, Eq. (11) is modified to,

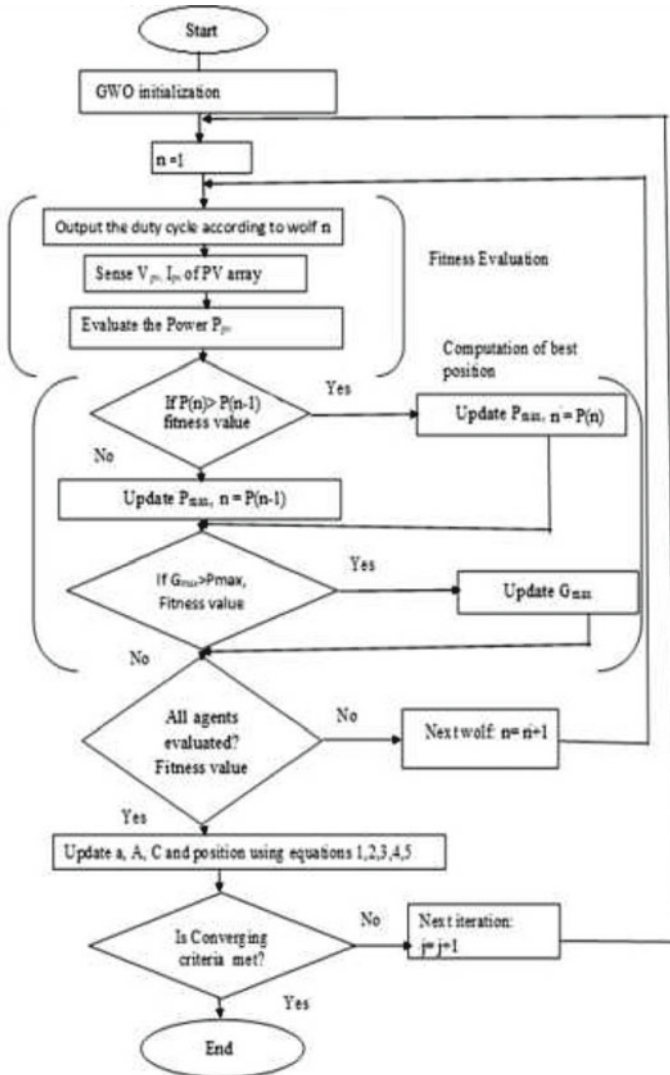


Fig. 4 Flow chart of the GWO algorithm

$$D_n(j + 1) = D_i(j) - F.D \tag{11}$$

where D here denotes the duty cycle of the grey wolf.

Therefore, the fitness function of this algorithm can be written as,

$$P(d_n^j) > P(d_n^{j-1}) \tag{12}$$

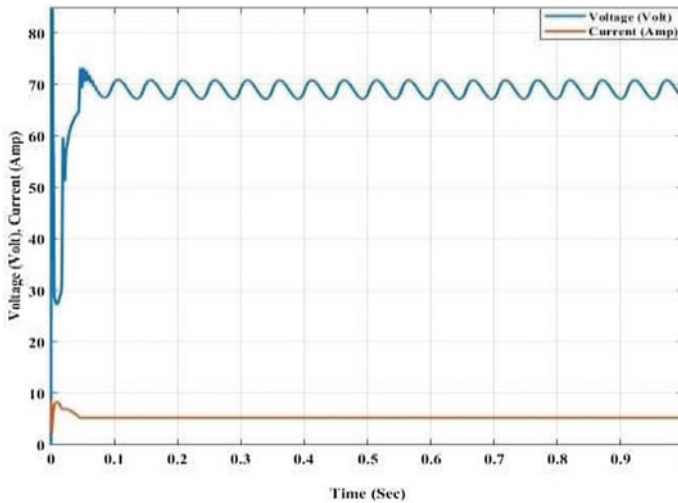


Fig. 5 Voltage output and current output—time characteristics of solar PV array

where power is denoted by “ P ”, duty cycle is denoted by “ d ”, the number of present grey wolves is denoted by “ n ” and number of iterations is given by “ j ”. Hence, the working of GWO algorithm has been explained in details using a flow chart in Fig. 5.

3 Result and Discussion

In both the methods, that is P&O and GWO-based MPPT Controller are developed in MATLAB. In order to design the P&O technique and GWO algorithm, codes have been used. The number of iteration used in GWO is 100. Output voltage and output current time characteristics for the solar PV array has been shown for 1 second which has been in shown in Fig. 6. The temperature where partial shading has been portrayed by varying the irradiation from 400 W/m^2 , 600 W/m^2 , 800 W/m^2 and 1000 W/m^2 and by keeping the temperature fixed at $25 \text{ }^\circ\text{C}$. The value of output voltage of solar photovoltaic array is found out to be 67.27 v (shown using blue line in Fig. 5) and output current is 5.29 A (shown using red line in Fig. 5) and also input to the Boost Converter.

The comparison between the output power of Boost Converter using both the optimization techniques, namely GWO and conventional Perturb & Observe for the MPPT controller has been portrayed in Fig. 6. It has been observed that using the GWO-based MPPT controller, the power reaches the global peak (G_p) of 485.31 watt ; whereas, using P&O-based MPPT controller, power reaches to 366.2 watt which is a local peak (L_p) and results in oscillation as it traps into local maximum point under partial shading condition. It is also observed that in presence of GWO-based

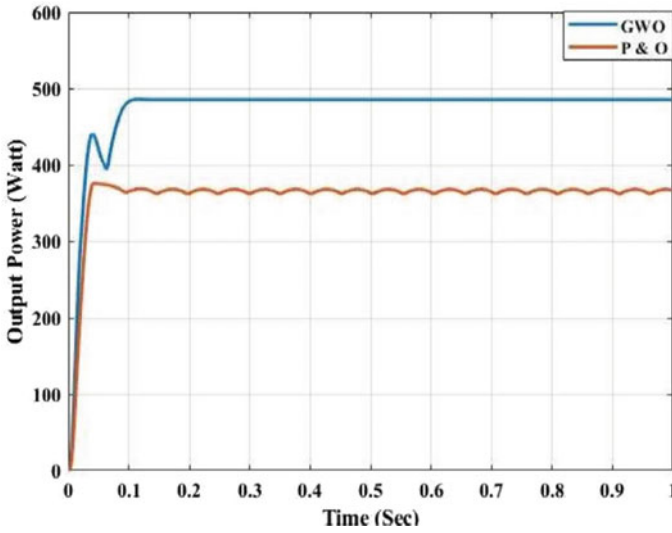


Fig. 6 Comparison of output power of DC boost converter using GWO and P&O

MPPT controller, power reached the G_p faster as compared with the conventional P&O-based MPPT controller.

As Figure 7 portrays comparison of voltage output of DC boost converter using both the techniques used to design the MPPT controller, the GWO technique which is global search algorithm and the conventional P&O technique. From the graph, it is observed that GWO algorithm gives the maximum value of output voltage that is 160.37 V for a time span of 1 second denoted by a blue line in order to reach

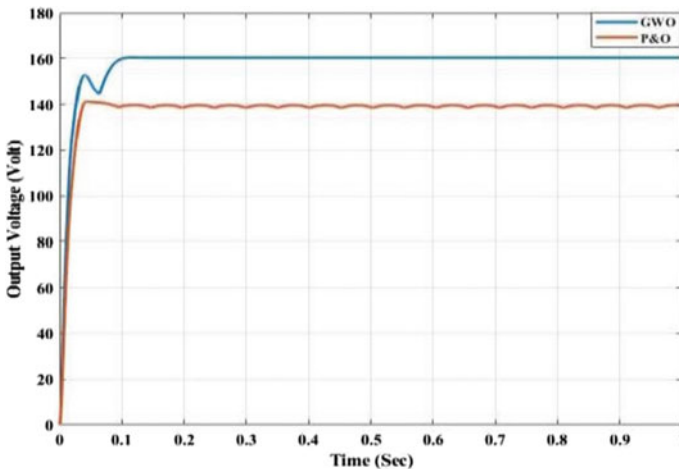


Fig. 7 Comparison of voltage output of DC boost converter using GWO and P&O

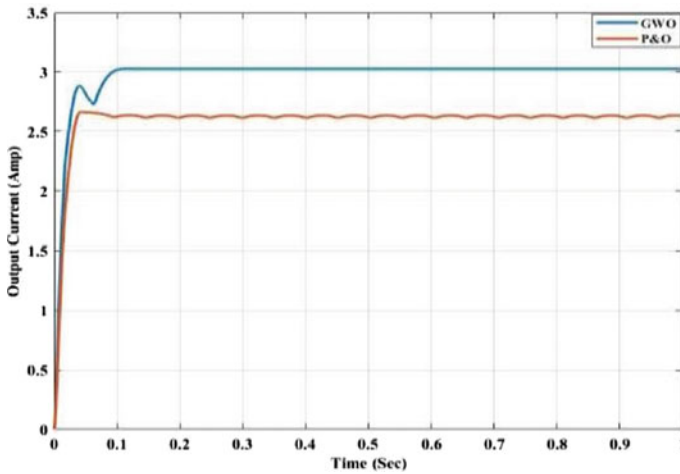


Fig. 8 Comparison of current output of DC Boost Converter of GWO and P&O

the global peak of 485.31 Watt (as seen in Fig. 7); whereas, the conventional P&O technique gives an output voltage of only 139.31 V (which is much lesser compared to that of the GWO) for 1 second denoted by a red line in order to track a GP of 366.2 Watt which also can be a local peak (LP) along with some oscillations which is not seen in case of GWO.

Figure 8 shows the comparison of current output of DC boost converter using GWO algorithm and conventional P&O technique which has been implemented in the paper to design the MPPT controller to track the maximum power point. In the graph, it is seen that the GWO-based MPPT gives a greater value of output current that is 3.02 A for 1 second (denoted by a blue line) corresponding to the maximum amount of output voltage of 160.37 V (as seen in Fig. 8) in order to track the GP (i.e. 485.31 Watt) but the conventional P&O-based MPPT gives a lesser value of output current that is 2.63 A for a time of 1 second (denoted by a red line) corresponding to a lesser amount of output voltage of 139.31 V in order to track a power of 366.2 Watt along with oscillations.

4 Conclusion

In this paper, MATLAB Simulink has been used to develop MPPT controller based on two techniques, namely conventional P&O and a global search algorithm GWO. In order to get desired result, a comparison between these two techniques has been done under partial shading condition. After the comparison, it was observed and hence concluded that GWO gives better result than P&O under partial shading condition. As it is seen in the graphs, GWO-based MPPT controller handles the partial shading

condition more efficiently and it outperforms the conventional P&O by providing faster tracking of GP, greater value of maximum power and reduced steady state oscillations as well as the GWO-based MPPT surpassed the conventional P&O-based MPPT giving greater value of output voltage with reduced steady state oscillation.

References

1. Bose BK (2010) Global warming: energy, environmental pollution, and the impact of power electronics. *IEEE Ind Electron Mag* 4(1):6–17
2. Guangul FM, Chala GT (2019) Solar energy as renewable energy source: SWOT analysis. In: 2019 4th MEC international conference on big data and smart city (ICBDSC), Muscat, Oman, pp 1–5
3. Sankar R, Velladurai S, Rajaranjan R, Thulasi JA (2017) II. PV system description: maximum power extraction in PV system using fuzzy logic and dual MPPT control. In: 2017 international conference on energy, communication, data analytics and soft computing (ICECDS), Chennai, pp 3764–3769
4. Al-Tashi Q, Kadir SJA, Rais HM, Mirjalili S, Alhussian H (2019) Binary optimization using hybrid grey wolf optimization for feature selection. *IEEE Access* 7:39 496–39, 508
5. Fatemi SM, Shadlu MS, Talebkah A (2019) Comparison of three-point P&O and hill climbing methods for maximum power point tracking in PV systems. In: 2019 international power electronics, drive systems and technologies conference (PEDSTC), Shiraz, Iran, pp 764–768
6. Khan MJ, Mathew L (2017) Artificial Intelligence based maximum power point tracking algorithm for photovoltaic system under variable environmental Conditions. In: 2017 recent developments in control, automation and power engineering (RDCAPE), Noida, pp 114–119
7. Vieira RG, de Araujo FM, Dhimish M, Guerra MI (2020) A comprehensive review on bypass diode application on photovoltaic modules. *Energies* 13(10):2472
8. Mohanty S, Subudhi B, Ray PK (2016) A grey wolf-assisted perturb and observe MPPT algorithm for a PV system. *IEEE Trans Energy Convers* 32(1):340–347
9. Bhowmik S, Pradip RP, Roy C (2019) Design of PID controller for maximum power point tracking for PV energy systems. In: Bera R, Sarkar S, Singh O, Saikia H (eds) *Advances in communication, devices and networking. Lecture notes in electrical engineering*, vol 537. Springer, Singapore
10. Pradip RP, Bhowmik S, Hazarika N, Roy C, Sengupta A (2020) Maximum power point tracking of photovoltaic systems using fuzzy-PID-based controller. In: Bera R, Pradhan PC, Liu CM, Dhar S, Sur SN (eds) *Advances in communication, devices and networking. ICCDN 2019. Lecture notes in electrical engineering*, vol 662. Springer, Singapore
11. Selmi T, Abdul-Niby M, Devis L, Davis A (2014) P&O mppt implementation using matlab/simulink. In: 2014 ninth international conference on ecological vehicles and renewable energies (EVER), IEEE, p 14
12. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Software* 69:46–61
13. Mohanty S, Subudhi B, Ray PK (2015) A new MPPT design using grey wolf optimization technique for photovoltaic system under partial shading conditions. *IEEE Trans Sustain Energy* 7(1):181–188

Yield Prediction of Indian Crops Based on Weather Data



P. Athulya and B. Mohammed Ismail

Abstract Agriculture provides employment to more than fifty percent of people in India. So, sustaining the variety of crop production is essential for the economy of India. Crop yield prediction involves an analytical study of Indian crop productivity based on different states, along with weather data of respective states. It would be beneficial if we could perform more analytical studies and produce results that are helpful for the agricultural sector to grow more. The crop prediction can be used by the sector to take necessary actions on the management of fields related to manpower, the number of crops to be taken, the type of crops to be selected, etc. In this work, a crop prediction based on linear regression, random forest, and support vector machine methods is applied to existing data to predict crop productivity. The dataset considered here has additional features added like annual precipitation, temperature, humidity, and surface pressure for more accuracy in yield prediction.

Keywords Crop productivity · Machine learning · Random forest · Machine learning · Yield prediction

1 Introduction

Agriculture has a significant role in the Indian economy. More than fifty percent of citizens are employed through agriculture and its related sectors. The huge amount of productivity data generated here can be used to analyze and produce some useful results. There is a significant amount of historical productivity data available to analyze. By using this historical data and the available weather data, it is possible to predict the future productivity of individual crops. The weather data includes rainfall, humidity, temperature, and surface pressure. Prediction of future crops can be used

P. Athulya (✉) · B. Mohammed Ismail
Department of Information Technology, Kannur University Campus Mangattuparamba, Kannur,
Kerala 670567, India
e-mail: athulyalaksh@gmail.com

B. Mohammed Ismail
e-mail: aboutismail@gmail.com

to make a better decision on the farm area. As the research is based on state-wise data, respective states can also take necessary actions to promote the most productive farm crops, and also, they can make farmers concentrate on productive crops in the respective area rather than following the same crops every year.

2 Related Works

There are so many researches that took place on crop productivity, but there are a few on Indian crops. As agriculture has a significant role in the life of every Indian citizen, taking better decisions by predicting its outcome would be of great benefit.

Kumar et al. [1] proposed a way to forecast crop yield using historical data such as rainfall, perception, pH, temperature, and production. The results are obtained using the machine learning algorithm random forest regression. The impact of surface pressure in production could be a significant feature in this study. Shah et al. [2] performed the crop yield prediction on a dataset collected from the USDA website by using multivariate polynomial regression, support vector regression, and random forest. They found the SVM model as the best model to predict the result. Through this research, they are guaranteeing the most suitable humidity, yield, temperature, and rainfall for cultivating better crop yield.

Crane-Droesch [3] performed crop yield prediction on the corn yield dataset and conducted a study on the impact of temperature, and rainfall in the crop yield. They found that timing of the heat, and moisture is important for predicting corn yield. A study conducted by Palanivel and Surianarayanan [4] used machine learning models and big data techniques to predict crop yield. They performed the prediction on different crop datasets using a variety of algorithms and resulted in a variety combination of algorithms and features as crop changes. And later, they could predict black gram yield using big data tools and techniques.

Kumar et al. [5] proposed an approach to increase yield based on biological factors such as seed quality, crop hybridization, and chemical factors like the use of fertilizers. In addition, the crop sequencing technique is also used to predict the yield. Here season of the crop plays an important role so that they could provide the result based on the order of crop cultivation in a year. Meeradevi and Salpekar [6] conducted research in which they implemented a mobile application for predicting yield. They used features such as the area covered, and crop-wise, district-wise, season-wise, year-wise production data, along with temperature, and rainfall data. They used the autoregressive integrated moving average model to train this dataset. Kalimuthu et al. [7] conducted a study on crop prediction using soil nature, seed data, and climatic conditions are considered for this research. The Naïve Bayes method is used for the prediction of yield.

3 Methodology

The proposed study uses machine learning models linear regression, random forest regressor, and support vector machine to predict crop yield. The features considered are state name, district name, crop year, season, crop, area, state-wise annual precipitation, state-wise annual temperature, state-wise annual humidity, and state-wise annual surface pressure. The historical crop data were collected from the website Kaggle.com. The climate-related data are collected from POWER Data Access Viewer by considering different states of India. The proposed research collected data from 1997 through 2015, and by analyzing these historical data, the relationship between a variety of climate factors and productivity of crops in different states of India are analyzed and used for the prediction of crop yield.

India has heterogeneity in its climatic condition across its states [8]. As most of the people in India depend on agriculture as the significant source of income, and also the entire population has to be fed, without depending on other countries, considering a systematic way to increase productivity within the country is imperative nowadays [9]. So, the proposed system tried to consider the data according to several Indian states. States from all the climatic zones are included. Exploratory analysis for finding the impact of all the selected features is carried out and analyzed the results thoroughly. Based on the findings, cross-checked the different features for selecting the training dataset using the following machine learning models.

3.1 Linear Regression

Linear regression is one of the widely used and simplest forms of regression algorithm. It assumes a linear relationship between two entities, and hence, a linear function is performed to predict the value of y entity when x entity has a value [10]. If there is any relation, the plotting on the entities can show a linear line within the graph.

In this research, the features selected do not show any linear relationship with each other, other than area and productivity. So, considering linear regression is as part of showing change in accuracy is not a good idea.

3.2 Random Forest

Random forests are a supervised learning algorithm that can be used for both classification and regression. The random forest classifier is a set of decision trees from a randomly selected subset of the training set [11]. It aggregates the votes from different decision trees to decide the final class of the test object. Random forest is best suitable for predicting the crop yield because the selected dataset is extremely

large, and the various features selected need to be analyzed properly in order to identify complex dependencies among them [12].

“Steps in random forest are:

1. Select random samples from a given dataset.
2. Construct a decision tree for each sample and get a prediction result from each decision tree.
3. Perform a vote for each predicted result.
4. Select the prediction result with the most votes as the final prediction.”

3.3 Support Vector Machine

This model is based on the support vector machine or SVM. These are supervised machine learning models that analyze data for regression analysis and classification. SVR is used when the available data is not linearly separable [13]. It is the same as SVM, using SVM for regression problems is referred to as SVR. It finds the nonlinearity within the provided data, and predict real values [14].

The workflow of the proposed research is shown in Fig. 1

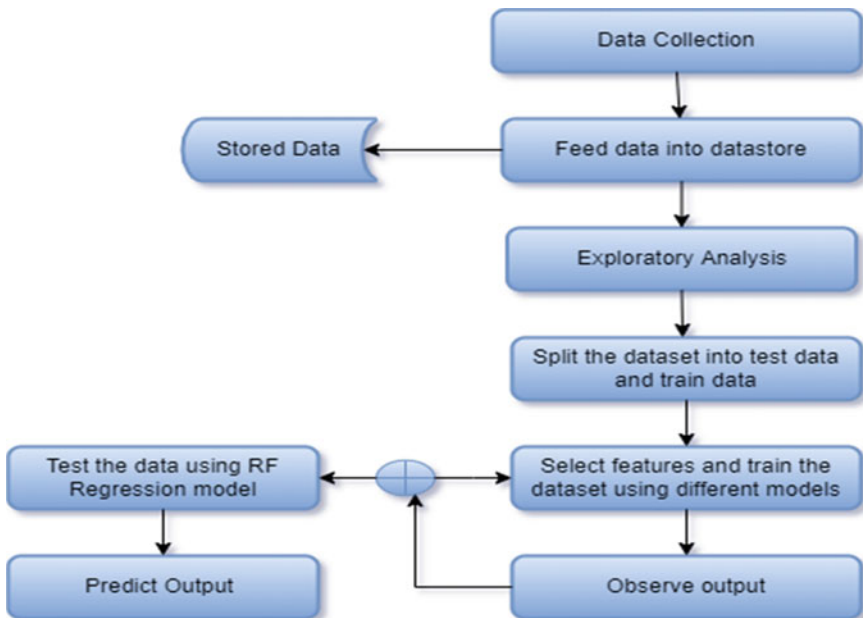


Fig. 1 Workflow

4 Result Analysis

The dataset is taken from kaggle.com. The features available on the dataset were state name, district name, crop year, season, crop, and area. In order not to use the same dataset, and in the hope of getting better accurate results, the proposed work added four more features, annual precipitation, temperature, humidity, and surface pressure of different states [15] (Figs. 2, 3, 4, 5, 6, 7, and 8).

Experimental results on the complete dataset are as follows:

From the analysis, it is understood that most of the crops depend on all the features selected, instead of the district name. As the precipitation and temperature data are an annual average of individual states, the district name becomes invalid, since all the other features are valid. There is no linear relationship between features, other than productivity and area of cultivation. Even though linear regression and support vector regressions are also considered in order to see the change in accuracy, the experiment was conducted on more than 2 lakhs of data; these are partitioned into testing and training datasets. Among them, 68% of records are taken for training the set, and the remaining 32% are taken for testing the data [16].

As the crops depend on most of the features selected, other than district name, and season all of them could result in better accuracy more than 95.75%, using random forest regression [17] (Table 1).

Fig. 2 Productivity against temperature

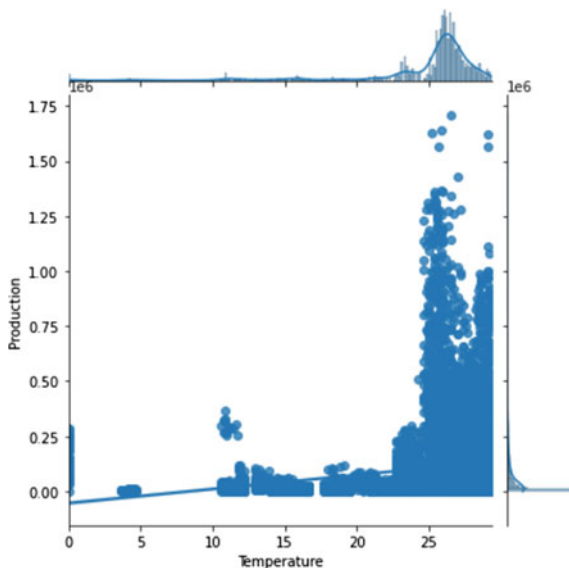


Fig. 3 Productivity against state

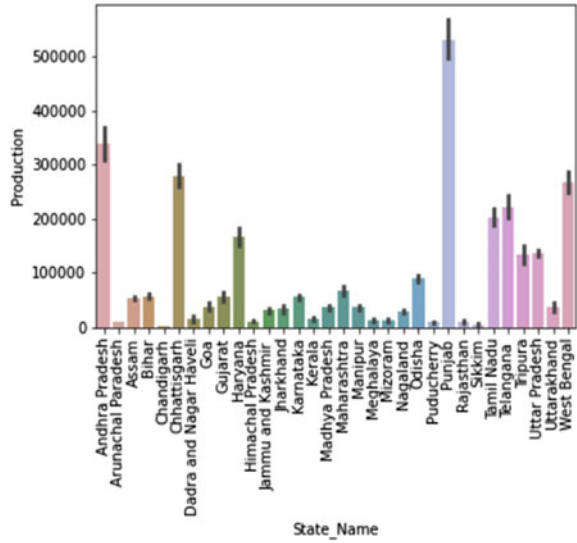
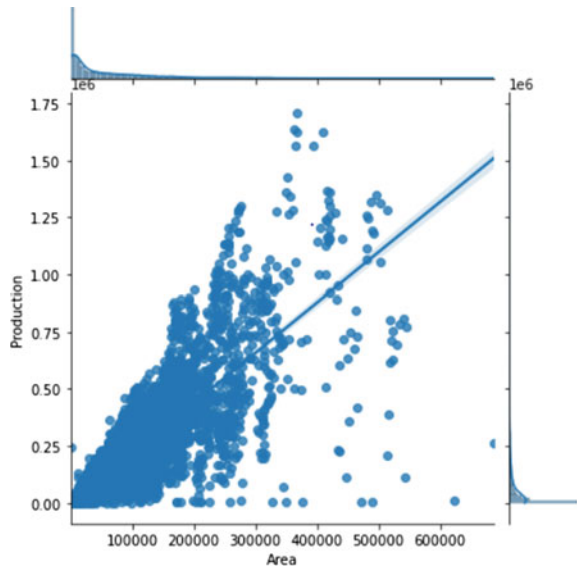


Fig. 4 Productivity against area



5 Conclusion

The analytics is performed on the entire dataset and other than a prediction, and the study was resulting in more common features of Indian crops. Most of the main crops in India produce high productivity when surface pressure is in the range of 95–100. The average farm area comes nearly 20–30 ha. There is productivity in Indian crops

Fig. 5 Productivity against crop year

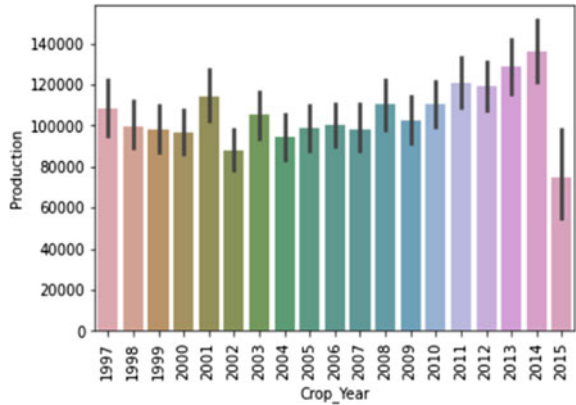
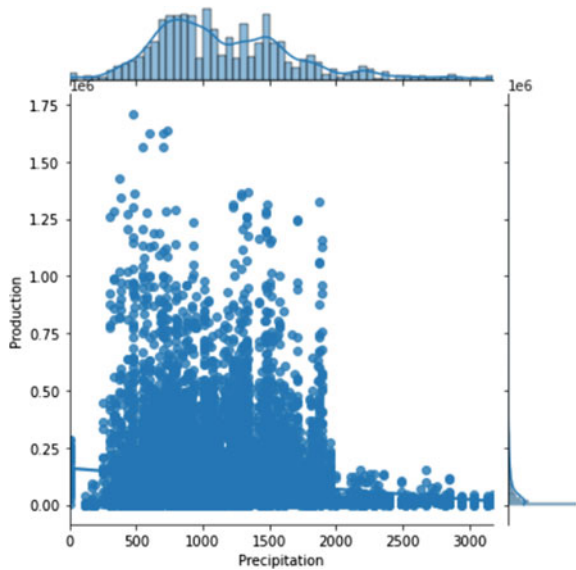


Fig. 6 Productivity against precipitation



when the humidity is in the range of 75–80. There are more crops depending on the annual average precipitation of India. According to the annual rainfall volume in 2019, it is 1284 ml. In the current research, it is found that most of the Indian crops depend on a 500–2000 ml precipitation rate. In the past 2 years, it may be increased due to low-pressure areas and heavy rainfall as its aftereffects. Exceeding 2000 ml is not a good practice. Crop productivity shows a significant increase when it possesses a temperature of nearly 25 °C.

The implementation of available data on LR, random forest, and SVM techniques has given 95% accuracy by using the random forest method. The obtained results show that the additional features of annual precipitation, temperature, humidity, and

Fig. 7 Productivity against humidity

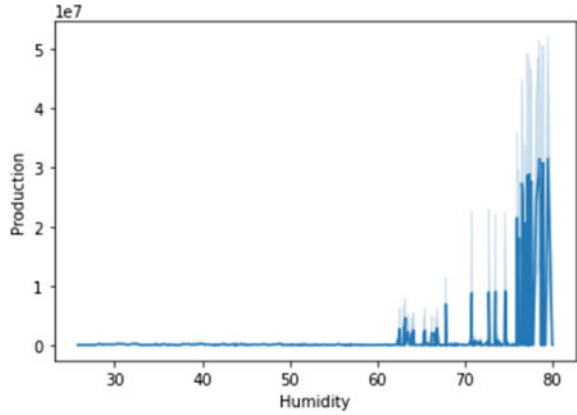


Fig. 8 Productivity against surface pressure

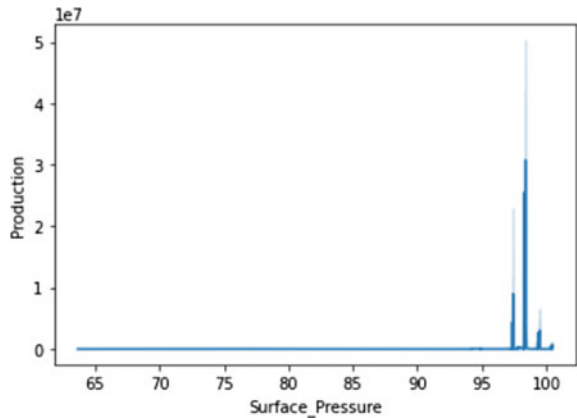


Table 1 Accuracy of models

Sl. no	Algorithm	Accuracy (%)
1	Linear regression	15.33
2	Support vector regression	0.11
3	Random forest regression	95.75

surface pressure selected are crucial in the cultivation of most of the crops in India. Changes in the selected features can result in a significant impact on productivity.

The experiments revealed that most of the Indian crops depend on the area of cultivation, precipitation, temperature, and environmental factors of different states such as humidity, surface pressure, temperature, and precipitation. There may be some other factors too such as fertilizers used. But the results here conclude that using the above features yield prediction up to 95% accuracy and are predictable.

References

1. Kumar Y, Spandana V, Vaishnavi VS, Neha K, Devi VGRR (2020) Supervised machine learning approach for crop yield prediction in agriculture sector. In: Fifth international conference on communication and electronics systems (ICCES 2020)
2. Shah A, Dubey A, Hemnani V, Gala D, Kalbande DR (2018) Smart farming system: crop yield prediction using regression techniques. In: International conference on wireless communication, Lecture notes on data engineering and communications technologies 19
3. Crane-Droesch A (2018 Sept 13) Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ Res Lett*,
4. Palanivel K, Surianarayanan C (2019) An approach for prediction of crop yield using machine learning and big data techniques. *Int J Comput Eng Technol (IJCET)* 10(03) May–June 2019
5. Kumar R, Singh MP, Kumar P, Singh JP (2015) Crop selection method to maximize crop yield rate using machine learning technique. In: 2015 International conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)
6. Meeradevi, and Salpekar H (2019) Design and implementation of mobile application for crop yield prediction using machine learning. In: 2019 global conference for advancement in technology (GCAT)
7. Kalimuthu M, Vaishnavi P, Kishore M (2020) Crop prediction using machine learning. In: Third international conference on smart systems and inventive technology (ICSSIT 2020)
8. Lakshmi KN, Reddy YK, Kireeti M, Swathi T, Ismail M (2019) Design and implementation of student chat bot using AIML and LSA. *Int J Innov Technol Exploring Eng* 8(6):1742–1746
9. Mohammed M, Kolapalli R, Golla N, Maturi SS (2020) Prediction of rainfall using machine learning techniques. *Int J Sci Technol Res* 9(2020):3236–3240
10. Ismail M, Harsha Vardhan V, Aditya Mounika V, Surya Padmini K (2019) An effective heart disease prediction method using artificial neural network. *Int J of Innov Technol Exploring Eng*
11. Mohammed Ismail B, Rajesh P, Alam M (2020) A machine learning based improved logistic regression method for prostate cancer diagnosis. *Int J Emerg Trends Eng Res* 8(9):5693–5698
12. Srinivas K, Mohammed Ismail B (2018) Test case prioritization with special emphasis on automation testing using hybrid framework. *J Theoret Appl Inf Technol* 96(13):4180–4190
13. Shahane R, Ismail Md, Prabhu CSR (2019) A survey on deep learning techniques for prognosis and diagnosis of cancer from microarray gene expression data. *J Comput Theoret Nanosci* 16(12):5078–5088
14. Mohammed Ismail B, Shaik MB, Eswara Reddy B, (2012) High rate compression based on luminance & chrominance of the image using Binary Plane Technique. *j Theor Appl Inf Technol* 42(2):191–195
15. Rajendra Prasad K, Mohammed M, Noorullah RM (2019) Visual topic models for healthcare data clustering. *Evol Intell*:1–18
16. Prasad KR, Reddy BE, Mohammed M (2021) An effective assessment of cluster tendency through sampling based multi-viewpoints visual method. *J Ambient Intell Humanized Comput* (2021):1–14
17. Ismail BM, Reddy TB, Reddy, BE (2016) Spiral architecture based hybrid fractal image compression. In: 2016 international conference on electrical, electronics, communication, computer and optimization techniques, ICECCOT 2016 pp 21–26

Categorization of Diabetic Retinopathy Applying Ensemble Model



R. Shekhar and T. Sridhar

Abstract Regardless of the advancement in the science and medical field, diabetes still emerges as a big threat to humanity. The only medicament that will help in reducing the effects of this disease to minimum is early detection and taking prophylactic steps toward it. One of the complications that affect eyes is diabetic retinopathy (DR). A diabetic causes impairment in the blood vessels of the tissue that is light sensitive at the retina. As per survey conducted by International Diabetes Federation in the year 2015, it was mentioned that it affects approximately 410 million people, worldwide. India is a commorancy to approximately 70 million people with this disease. Diabetic eye disease or DR is the most common impediment of diabetes. Round about 2.6% of global blindness is caused by this disease. In unmitigated terms, just about 3–4.5 million individuals in India are anticipated to fall victim to vision threatening diabetic retinopathy (VTDR). The choices for treatment of VTDR demand costly devices and medications, and a regular follow-up from diagnosis is required to the last day of your life. In deliberation of the fact that 70% of the Indian citizenry depend on necessitous expenses for their healthcare benefit, one person with VTDR in a household is enough to drive a menage to below poverty line. Therefore, all undertakings should be initiated exigently to prevent individuals with diabetes to move into the vicious cycle of diabetes, blindness, and poverty. This proposed work will classify the stages of diabetic retinopathy with an approximate precision of 97.8% which is an improvement from the previous model with a margin of around 10%. We have used pretrained models in our ensemble learning process.

Keywords Diabetic retinopathy · Ensemble model · VTDR · Image processing

R. Shekhar

Department of CSE, Alliance College of Engineering and Design, Alliance University, Bagalore, Karnataka, India

e-mail: shekhar.r@alliance.edu.in

T. Sridhar (✉)

Department of ECE, Alliance College of Engineering and Design, Alliance University, Bagalore, Karnataka, India

e-mail: sridhar.t@alliance.edu.in

1 Introduction

Diabetic retinopathy is one of the major reasons of blindness in humans throughout the world. Predictions of affected cases are estimated to exceed 370 million patients by 2030. The main effect of DR is the closing of the blood vessels in the eyes, followed by the eyes responding to this in two ways:

Firstly, new blood vessels are created, and bleeding occurs above the main area (vitreous) which must be clear to allow light to be transmitted to the most sensitive part of eyes (retina) through the cornea, pupil, and lens. The main role of the retina is to convert light into impulses which are disseminated through the optic nerve to the brain where processing occurs to be able to see the image and understand it.

Secondly, the blood vessels leak, and this will influence and harm the retina, specifically in the macula (core part of the retina), whose role is detailed vision, as shown below. Diabetic retinopathy can be categorized into five classes:

1. Normal DR
2. Mild non-proliferative DR
3. Moderate non-proliferative DR
4. Severe non-proliferative DR and
5. Proliferative DR (Figs. 1 and 2).

In this work, the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 dataset which had 3662 images of the fundus and with highly imbalanced data used. The data was classified by clinicians into five stages as mentioned above. The end goal of our project is to use different pretrained deep learning convolutional neural network models varying in width (number of parameters) and depth (numbers of layers) to design a program that can classify input images into diabetic retinopathy levels and with high precision. This is challenging due to the difference between the domain that used ImageNet as a dataset to build a deep learning model and a medical image domain that used diabetic retinopathy images as a dataset and fine-tuning a pretrained convolutional neural networks (CNN) model could give similar results in DR images.

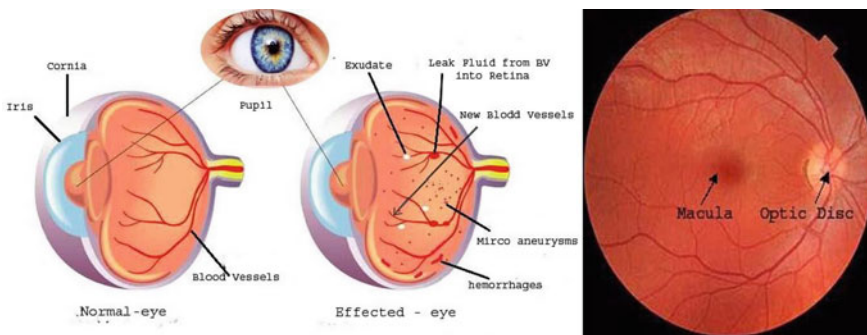


Fig. 1 Normal eye versus effected eye

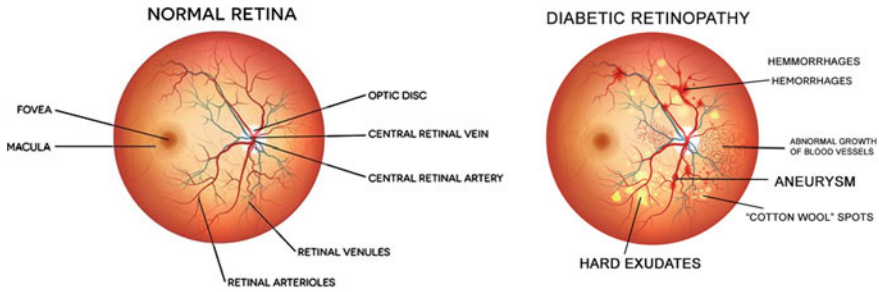


Fig. 2 Diabetic retinopathy effected retina versus normal retina

Therefore, we would like to meet the challenge by selecting an appropriate algorithm to build a model for DR image classification. The pretrained models VGG16, ResNet50, DenseNet201, InceptionV3, InceptionResNetV2, and Xception used all of these in the ensemble learning process and produced a higher prediction rate.

1.1 Existing Research Analysis

For these work, twenty-two different papers analyzed. *Qummar* et al. offered and introduced a CNN ensemble-based framework in color fundus images to detect and classify the various categories of DR [1]. They used the database from Kaggle, which is the largest publicly available database to train and assess their representation. The outcomes show that their representation performed better than other futuristic methods and was successful in detecting all the stages of DR with 80.8% accuracy. *N. Sikder* et al. used a light CNN model [2]. The overall accuracy of this experiment conducted by us was 89.07%.

Mobeen-ur-Rehman et al. in this paper, it is suggested to use automated tools for detecting diabetes using DR images [3]. Implementation of pretrained VGG-16 and SqueezeNet, CNN models AlexNet; it applies a customized CNN model proposed specifically for DR image classification. The pretrained model shows better results than the customized model in the past papers. *Masood* et al. presented a model of computer vision that detected the diseases in diabetic eye in eye fluorescein angiography photographs. The size of the input was the key drive for the performance [4]. To overcome the problem of overfitting and to obtain final improved results, the final test accuracy achieved from 37.6% to 48.2%, the model of transfer learning eradication was used. To highlight the features and remove the class imbalance, subtracting the local average color of the image from the complete image is used. (*Krizhevsky* et al.) The results show that the deep CNN model, when implemented can achieve high result on a high exigent dataset using purely superintendence learning [5]. Degeneration in the network execution occurs if a single convolutional layer is removed. The results have improved as the network is made bulkier and more

rigorously trained, but even after that, it must go through several orders of magnitude to be at par with the infero-temporal pathway of the human visual system. Dutta et al. proposed an optimal model for diabetic retinopathy detection [6]. In term of image, both DNN and CNN models were effective as CPU training time of CNN was getting affected in the study. The DNN surpassed CNN for training accuracy and validation accuracy. CHen et al. in their paper presented a deep convolutional neural networks (DCNNs)-based framework for the DR detection [7]. This proposed framework achieves 0.959 and 0.965 AUC for DR and RDR cases on the Messidor dataset among six useful methods which surpassed futuristic (0.921 and 0.957). Agung et al. in this research, they put forward a magnificent new method using CLAHE in G channel to improve the color retinal image quality [8]. The algorithm conclusion is that, when applied to 30 retinal images of which 21 contained exudates and nine without pathology [9]. In this sensitivity and specificity for exudate detection were 88.5% and 99.7%, respectively. Islam et al. [10] in their work, the OCT images were used deep learning models to perform diabetic retinopathy classification. Mirza Mohd Shahriar Maywood et al. [11] have attained accuracy of 0.9402 on the training set and achieved 0.9333 on the testing set. Hassan et al. [12] said that their process of detecting DR disease could be made significantly accurate with an image processing and machine learning technique. Niemeyer et al. [13] demonstrated their results in an online competition. (Swathi et al.) Retinal fundus images were considered techniques for preprocessing in their paper [14]. The noises like Gaussian, salt and pepper removed and enhance the data of the retinal images with this method. Tripathi et al. [15] presented a work with differential morphological profile (DMP) method. This is to exudate detection from color images. Two other methods were used to compare the specificity, sensitivity, and PPV. (Raju et al.) The observation showed that from fundoscopic retinal images of the eye detection of high sensitivity and specificity of DR staging, implementing a model that utilizes the powers of a CNN algorithm [16]. CNN's massive capability to be an automatic classifier of the laterality and severity of fundus images in real time is exemplified. To improve the prediction accuracy, demographical parameters of the patient are also a necessity. To filter out images with poor quality criteria must be incorporated. (Tanuki O. Oladele et al.) In order to predict DR disease, the same algorithm was utilized [17]. A positive influence on predictions could has been shown by implementing on support vector machines but that would have a burden on time consumption in design and functioning of the model. The performance of the algorithm was improved while using the same algorithm for feature selection and classification than using the same algorithm for classification without feature selection. In most algorithms considered, the improvement was most noticeable for support vector machines. Gadekallu et al. [18] in their work used UCI machine learning repositor, using DNN-PCA-GWO model. The standard scaler method was utilized to remove the outliers, transforming it, and normalizing the data for the model implemented. The enhancement of the model performance has increased by the number of records and randomly duplicating them. There was a significant rise in accuracy of DNN noted when the number of records was increased. Doshi et al. [19] used DR automatic detection and classification for color fundus. The best score of 0.3996 is obtained by the ensemble of those three models. Geetha

Ramani et al. [20] used data mining techniques based on image processing and the focus was placed on the techniques for feature relevance and classification to categorize accurately. Arcadu et al. [21] used a method of binary, the input data are sets of seven CFP images obtained for a selected eye at baseline was presented the overall model. Costa et al. [22] presented a new weakly supervised DR detection system based on the multiple-instance learning (MIL) framework.

2 Methodology

See Fig. 3.

2.1 Dataset

The dataset used in this work downloaded from APTOS 2019 from Kaggle. In this work, a colossal set of retina images obtained with fundus photography under various imaging environments. Medical practitioner has graded every single image for the severity of using a scale of zero to four. The scale zero as no diabetic retinopathy, one as mild, two in middle treated as moderate, whereas three is severe and finally, four as proliferative DR.

Congruent to any real-world dataset, we too encountered noise in both images as well as the labels. Images contained artifacts, were not in focus, overexposed, or underexposed. The images were accumulated from several clinics using a sundry of cameras over a prolonged period, which brought in additional variations.

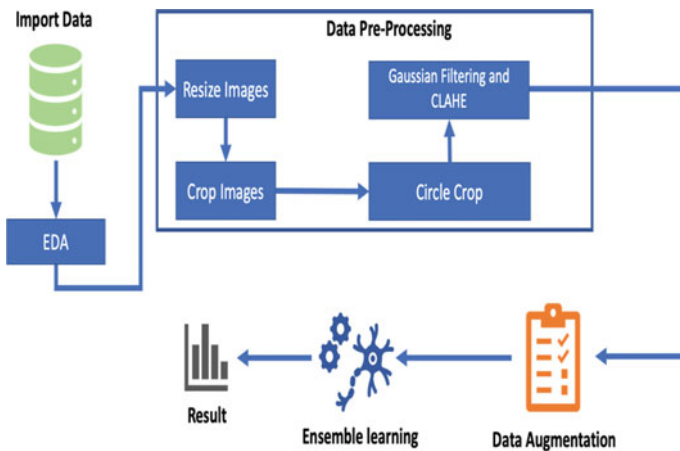


Fig. 3 Data flow architecture

Number of images in train dataset: 3662
Number of images in Test dataset: 1928

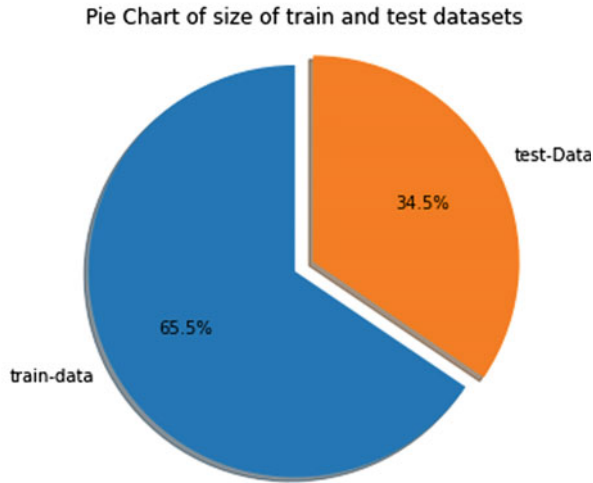


Fig. 4 Dataset overview

2.2 Exploratory Data Analysis

The number of images in both the train and test datasets used in this work is shown here (Fig. 4 Dataset overview).

We chose pie chart to show us the pictorial representation of the size of the train and test dataset. Then, we saw the class wise distribution of diabetic retinopathy in the dataset (Fig. 5 DR-class distribution, Fig. 6 class wise distribution in train dataset).

Then, we plotted a histogram showing range of height and width of images in both train and test datasets. (Fig. 7 Histogram showing height and width range) Fig. 6 Class wise distribution of images in train dataset.

2.3 Data Preprocessing

Here, in the dataset, we have imbalanced data. One way of addressing the problem of imbalance of class is to desultorily resample the training dataset. Random resampling of an imbalanced dataset can be achieved through two main approaches that are to delete items from the majority class, called under sampling, and to replicate items from the minority class, called oversampling (Fig. 8). Here, we started with optimizing the train dataset by oversampling the data so that we can have more number of images to train of each class on our model (Figs. 9 and 10). Then, we

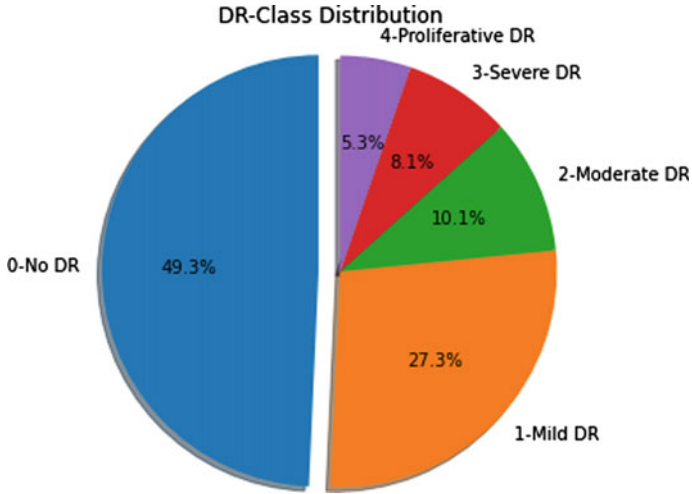
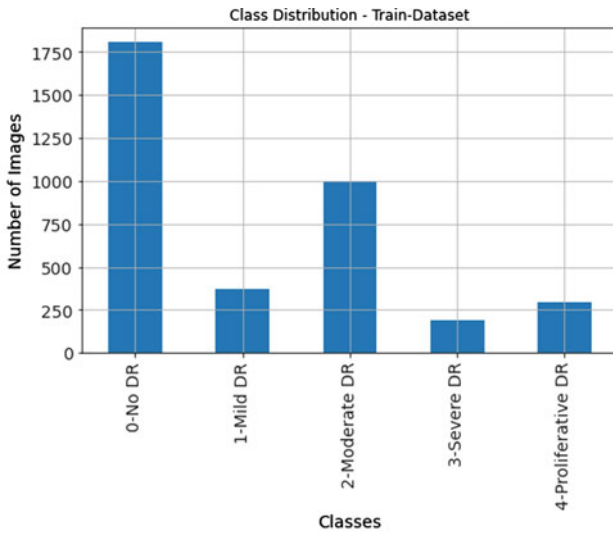


Fig. 5 DR-class distribution



Number of images in class 0-No DR --> 1805 (49.29% of total data)
 Number of images in class 1-Mild DR --> 370 (10.1% of total data)
 Number of images in class 2-Moderate DR --> 999 (27.28% of total data)
 Number of images in class 3-Severe DR --> 193 (5.27% of total data)
 Number of images in class 4-Proliferative DR --> 295 (8.06% of total data)

Fig. 6 Class wise distribution

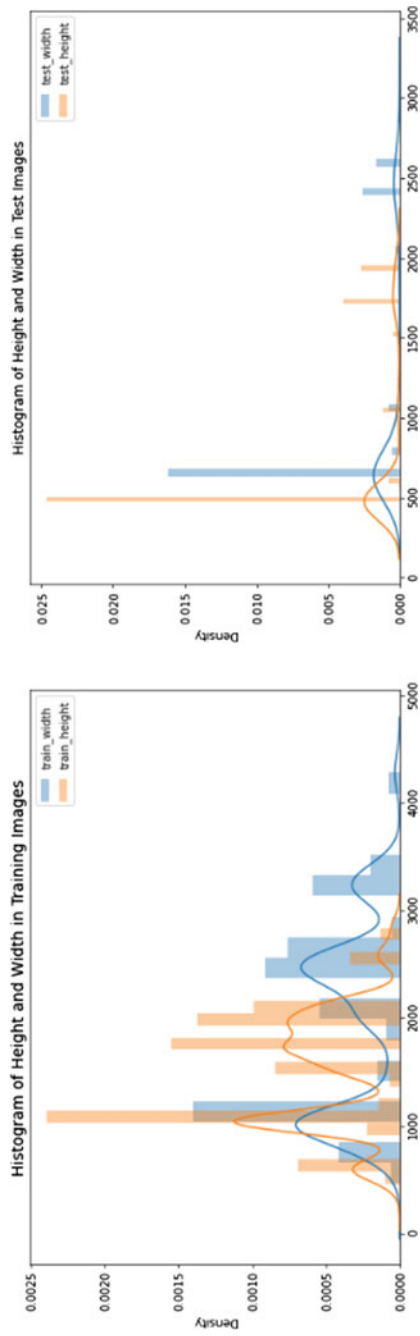


Fig. 7 Histogram showing height and width range

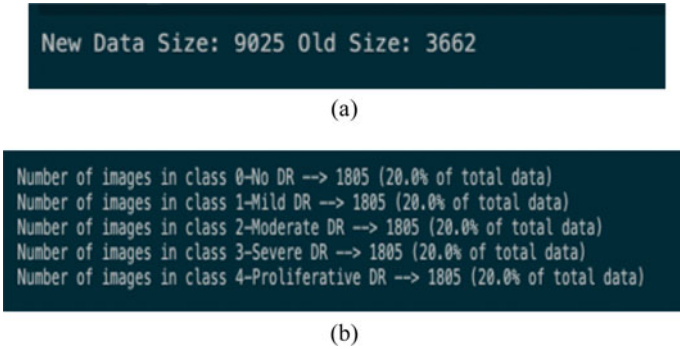


Fig. 8 a, b represents new balanced data

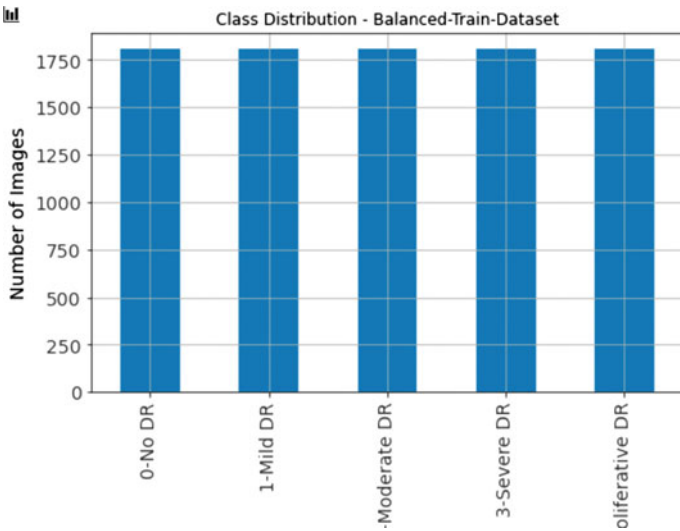


Fig. 9 Balanced train dataset

resized the images to 256×256 size as it is faster for the architecture to process the smaller-sized data and for most of our pretrained models, this size is the default one. After that, we cropped the images, cutting the sides and then, we applied circle crop to focus mainly on the retina removing the black sides, hence removing the noise. Then, we applied Gaussian blur and CLAHE for optimizing the color levels of the images resulting in more optimized images and focusing on the abnormalities in the retina.

Gaussian Blurring: Gaussian blur is the product of image blurring by a Gaussian function. It is a broadly used effect in graphics software, ordinarily to diminish image noise and minimize detail. Gaussian blurring is also used as a preprocessing stage hitherto applying our machine learning or deep learning models. **CLAHE:** apply

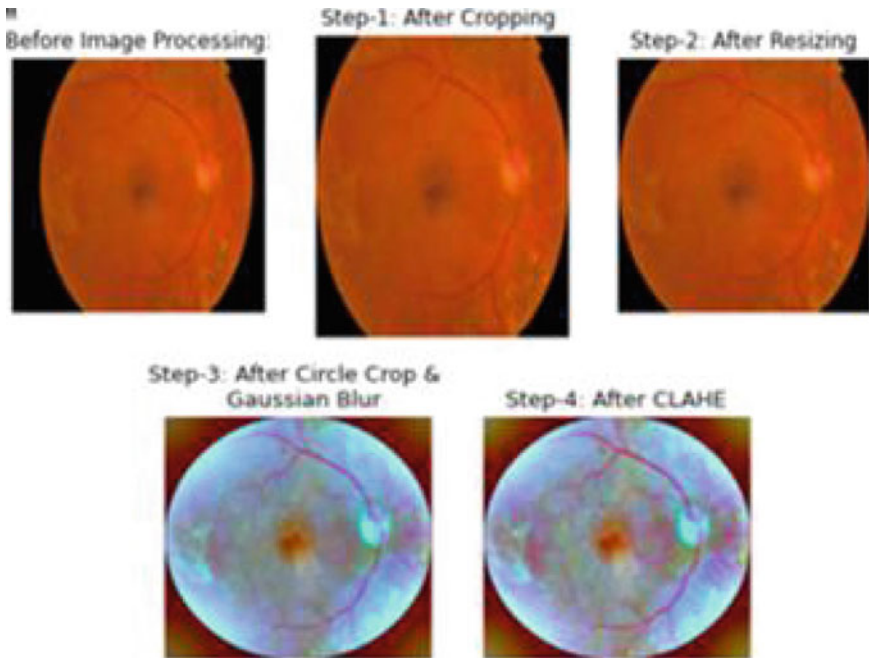


Fig. 10 Stages of image preprocessing

contrast-limited adaptive histogram equalization (CLAHE) to regularize images. CLAHE is a modified version of *adaptive histogram equalization (AHE)* which looks after contrast's over-amplification. CLAHE enacts on small patches in the image, called tiles, in preference over the entire image. Bilinear interpolation is used on the immediate tiles and then pooled to remove the artificial boundaries. This algorithm can be applied to enhance the contrast of images.

In Fig. 10, you can see the pictorial representation of step-by-step stages of image preprocessing.

2.4 Data Augmentation

After the data processing step, our next plan was to increase the accuracy of the model and for that we augmented the training data and added the variations of the data to it by flipping the images, rotating it, etc. In data analysis, data augmentation is a technique that increases the sum of data by slightly increasing the modified copies of data already existing or synthetic newly created data from existing data. This acts as to regularize that helps to reduce over fitting during training a machine learning model (Fig. 11).

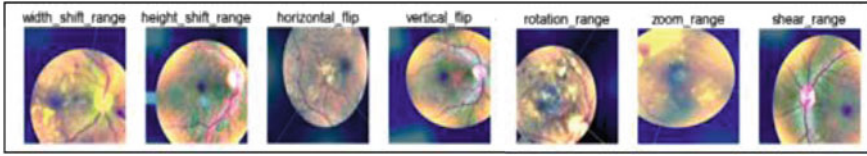
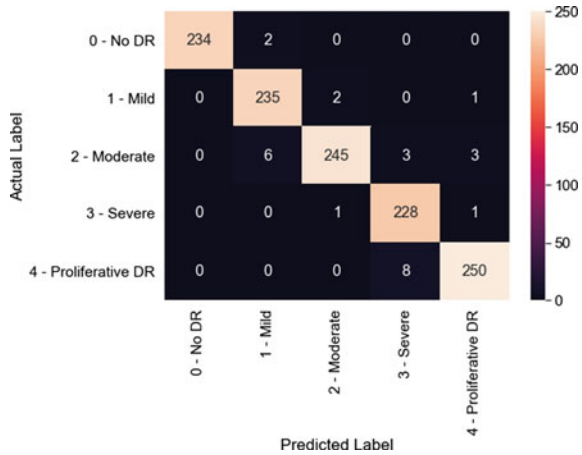


Fig. 11 Data augmentation steps

Fig. 12 Confusion matrix of ensemble model



2.5 Ensemble Model

Ensemble modeling is the method of amalgamation of several machine learning techniques in a meta algorithm into a single predictive model. In this ensemble model method, various objectives are used to decrease variance (bagging), bias (boosting), or improve predictions (stacking). Stacking is the creation of new models by combining multiple predictive models. Because of the relaxing nature, the busy approach generally outperforms individual models.

Stacking puts the best performing models in featurite and poorly performing ones aside. For the same reason, stacking is most effectual when the base models are notably different. That is the reason why we used to stack to improve the prediction of our model, which is evident from our results. The proposed approach ensembles the six deep CNN models VGG16, ResNet50, DenseNet201, InceptionV3, InceptionResNetV2, and Xception. For each model, the output layer uses *SoftMax* as an activation function which culminates the probabilities that how much the input fits to the set of different classes {*Normal, Mild, Moderate, Severe, PDR*}. To predict the class label of the unseen example, in the case of testing, we use stacking to integrate the results of all different models and beget a unified output. The ensemble approach integrates the prerequisites of individual models and leads to better performance.

There are seven steps in creating the base model first ‘load data’, second ‘define Keras model’, then third one ‘compile Keras model’, fourth ‘fit Keras model’, fifth one ‘evaluate Keras model’, sixth ‘tie it all together,’ and finally last step one ‘make predictions’.

The fine-tuning requires, in addition to update the CNN architecture, to learn new objective classes in this work.

1. Initialize with removing the fully connected nodes at the end of the network (i.e., where the actual class label predictions are made).
2. Then, replace the fully connected nodes with freshly initialized ones.
3. Ensuring that any previous robust features learned by the CNN are not destroyed, freeze earlier CONV layers earlier in the network.
4. Initialize training, but train only the FC layer heads.
5. Unfreeze some of/all the CONV layers in the network facultatively and execute a second pass of training.

3 Results

3.1 Performance Parameters

To quantifiably assess the proposed model, we use some performance metrics such as accuracy, sensitivity, specificity, precision, and F-1 measures.

Accuracy: Accuracy is computed both positive and negative in terms of classes:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Recall/Sensitivity: Sensitivity can be calculated as

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Specificity: Specificity also can be found as

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Precision: Precision can be found as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

F1-score: It shows the weighted harmonic average of both precision and recall

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

This returns a score of zero and one, best score considered as one and worst score is zero.

QWK Score: This metric ordinarily vacillates random agreement considered as 0 and as complete agreement which is 1. If there is less agreement than the metric may go below zero.

Quadratic weighted kappa, calculated in the following way. First, an $N \times N$ histogram matrix O is built, then $O_{i,j}$ corresponds to the number of its upgrades. ‘i’ considered as actual that received a predicted value ‘j’. A $N \times N$ matrix of weights, w , is calculated based on the difference between actual and predicted values

$$w_{i,j} = \frac{(i - j)^2}{(N - 1)^2}$$

An N -by- N histogram matrix of foreseeable outcomes, E , is calculated expecting that there is *no correlation* between values. This is quantified as the outer product between the actual histogram vector of outcomes and the predicted histogram vector, normalized such that E and O have the same sum.

Finally, from these three matrices, the quadratic weighted kappa is calculated as

$$\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}$$

where w = weighted matrix, O = histogram matrix, and E = expected matrix.

Confusion Matrix

See Fig. 12 and Tables 1 and 2.

Table 1 Ensemble model classification report

Classes	Precision
0	0.99
1	0.97
2	0.97
3	0.98
4	0.96

Table 2 Ensemble model predicted label

Accuracy	Precision	Recall	F1-score	QWK score
0.977850697292863	0.9781879741573148	0.977850697292863	0.9778555993857755	0.9909831548602301

4 Conclusion

Diabetic eye disease, DR is the foremost reason of visual impairment causing blindness in diabetic patients at a global level. Images of retina are vital for ophthalmologists to diagnose diseases. Good performance can be achieved by several techniques on retinal feature that are clearly visible. The accuracy that we get from the model matters finally. However, the biggest contention with diagnosing DR is the total number of cases. Country like India has more than 70 million individuals suffering from diabetes. Hence, proper screening with good medication, early diagnosis with healthy diet prevents further damage. The main objective of our model is to detect diabetic retinopathy and classify its severity on the following basis: 0—No DR, 1—Mild, 2—Moderate, 3—Severe, 4—proliferative DR. We performed prediction on different pretrained models and achieved the highest accuracy when we used ensemble model, i.e., of 97.8% which is better than the existing models. The existing models used different methods that includes light CNN model, using only VGG16 as their base model and even ensemble model but none of them could gain accuracy more than 92%, thus we tried many different methods to get better than these existing models, and hence, we landed up on our ensemble model approach with VGG16, ResNet50, DenseNet201, InceptionV3, InceptionResNetV2, and Xception Models as their layers.

In the future, we would be optimizing our current model to take less processing time and then creating a web app and running our ensemble model on it and will be creating a platform where doctors just must upload the retinal image and in a nick of time, they will be getting the classification report and can treat the patient as soon as possible.

References

1. Qummar S, Khan F, Shah S, Khan A, Band S, Rehman Z, Khan I, Jadoon W (2019) A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access*. 7:1–1. <https://doi.org/10.1109/ACCESS.2019.2947484>
2. Chowdhury MS, Taimy FR, Sikder N, Nahid AA (2019) Diabetic retinopathy classification with a light convolutional neural network. In: 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), pp 1–4. <https://doi.org/10.1109/IC4ME247184.2019.9036614>
3. Mobeen-ur-Rehman, Khan SH, Abbas Z, Danish Rizvi SM (2019) Classification of diabetic retinopathy images based on customised CNN architecture. In: 2019 amity international conference on artificial intelligence (AICAI), Dubai, United Arab Emirates, pp 244–248. <https://doi.org/10.1109/AICAI.2019.8701231>
4. Masood S, Luthra T, Sundriyal H, Ahmed M (2017) Identification of diabetic retinopathy in eye images using transfer learning. In: 2017 international conference on computing, communication and automation (ICCCA), pp 1183–1187. <https://doi.org/10.1109/CCAA.2017.8229977>
5. Krizhevsky, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
6. Dutta S, Manideep BC, Basha SM, Caytiles RD, Iyengar NCSN (2018) Classification of diabetic retinopathy images by using deep learning models. *Int J Grid Distribut Comput* 11(1):99–106

7. Chen Y, Wu T, Wong W, Lee C (2018) Diabetic retinopathy detection based on deep convolutional neural networks. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 1030–1034. <https://doi.org/10.1109/ICASSP.2018.8461427>
8. Setiawan W, Mengko TR, Santoso OS, Suksmono AB (2013) Color retinal image enhancement using CLAHE. In: International conference on ICT for smart society, pp 1–3. <https://doi.org/10.1109/ICTSS.2013.6588092>
9. Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, Lal S, Usher D (2002) Automated detection of diabetic retinopathy on digital fundus images. *Diabet Med* 19(2):105–112. <https://doi.org/10.1046/j.1464-5491.2002.00613.x> PMID: 11874425
10. Islam KT, Wijewickrema S, O’Leary S (2019) Identifying diabetic retinopathy from OCT images using deep transfer learning with artificial neural networks. In: 2019 IEEE 32nd international symposium on computer-based medical systems (CBMS), pp 281–286. <https://doi.org/10.1109/CBMS.2019.00066>
11. Shahriar Maswood MM, Hussain T, Khan MB, Islam MT, Alharbi AG (2020) CNN based detection of the severity of diabetic retinopathy from the fundus photography using efficientnet-B5. In: 2020 11th IEEE annual information technology, electronics and mobile communication conference (IEMCON), pp 0147–0150. <https://doi.org/10.1109/IEMCON51383.2020.9284944>
12. Hassan MA, Ullah RI, Hamdan Alenezi A, Rassem TH (2020) Identifying the level of diabetic retinopathy using deep convolution neural network. In: 2020 emerging technology in computing, communication and electronics (ETCCE), pp 1–6. <https://doi.org/10.1109/ETCCE51779.2020.9350905>
13. Niemeijer M et al (2010) Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Trans Med Imaging* 29(1):185–195. <https://doi.org/10.1109/TMI.2009.2033909>
14. Swathi C, Anoop BK, Dhas DAS, Sanker SP (2017) Comparison of different image pre-processing methods used for retinal fundus images. *Conf Emerg Devices Smart Syst (ICEDSS) 2017*:175–179. <https://doi.org/10.1109/ICEDSS.2017.8073677>
15. Tripathi S, Singh K, Singh B, Mehrotra A (2013) Automatic detection of exudates in retinal fundus images using differential morphological profile. *Int J Eng Technol* 5:2024–2029
16. Raju M, Pagidimarri V, Barreto R, Kadam A, Kasivajjala V, Aswath A (2017) Development of a deep learning algorithm for automatic diagnosis of diabetic retinopathy. *Stud Health Techno Inform* 245:559–563 PMID: 29295157
17. Oladele TO, Ogundokun RO, Kayode AA, Adegun AA, Adebisi MO (2019) Application of data mining algorithms for feature selection and prediction of diabetic retinopathy. In: Misra S et al. (eds) *Computational science and its applications—ICCSA 2019*. ICCSA 2019. Lecture notes in computer science, vol 11623. Springer, Cham. https://doi.org/10.1007/978-3-030-24308-1_56
18. Gadekallu TR, Khare N, Bhattacharya S et al (2020) Deep neural networks to predict diabetic retinopathy. *J Ambient Intel Human Comput*. <https://doi.org/10.1007/s12652-020-01963-7>
19. Doshi D, Shenoy A, Sidhpura D, Gharpure P (2016) Diabetic retinopathy detection using deep convolutional neural networks. In: 2016 international conference on computing, analytics and security trends (CAST), Pune, India, pp 261–266. <https://doi.org/10.1109/CAST.2016.7914977>
20. Geetha Ramani R, Balasubramanian L, Jacob SG (2012) Automatic prediction of diabetic retinopathy and Glaucoma through retinal image analysis and data mining techniques. In: 2012 international conference on machine vision and image processing (MVIP), Coimbatore, India, pp 149–152. <https://doi.org/10.1109/MVIP.2012.6428782>
21. Arcadu F, Benmansour F, Mauuz A et al (2019) Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med* 2:92. <https://doi.org/10.1038/s41746-019-0172-3>
22. Costa P, Galdran A, Smailagic A, Campilho A (2018) A weakly-supervised framework for interpretable diabetic retinopathy detection on retinal images. *IEEE Access* 6:18747–18758. <https://doi.org/10.1109/ACCESS.2018.2816003>

The Utility of Simulink Subsystems in Handling and Processing of Biomedical Signals and Images



S. Sivaarunagirinathan, D. Jeyakumari, M. Sundar Prakash Balaji, P. Thanapal, and V. Elamaran

Abstract Simulink, a graphical programming framework based on MATLAB, can be used to model, simulate, and study multi-domain dynamical systems. Because of its drag-and-drop capabilities, easy graphic user interface features, and zero-coding settings, Simulink has become the most extensively used tool in industry and academy. In comparison with alternative software solutions, Simulink can minimize the development cycle time of any real-time system. The value of subsystems in biomedical signal and image processing, such as the enabled subsystem, the triggered subsystem, the triggered and enabled subsystem, and the control flow subsystem, is highlighted in this paper. Breast cancer picture and human voice signal are used to implement image segmentation using enabled subsystem, voiced/unvoiced classification using triggered subsystem, and calculating of root mean square (RMS) amplitude using If Action subsystem. In an experimental environment, the MATLAB 9.4 tool is used to simulate biomedical signals and images.

Keywords Classification · Image segmentation · Root mean square · Simulink · Voiced/unvoiced · Zero-crossings detection

1 Introduction

Simulink is a simulation and graphical modeling tool that works with the MATLAB software suite. A system is represented for a specific application by linking graphical elements (block diagrams), making it simple to understand and implement. Figure 1

S. Sivaarunagirinathan · V. Elamaran (✉)

Department of ECE, School of EEE, SASTRA Deemed University, Thanjavur, India
e-mail: elamaran@ece.sastra.edu

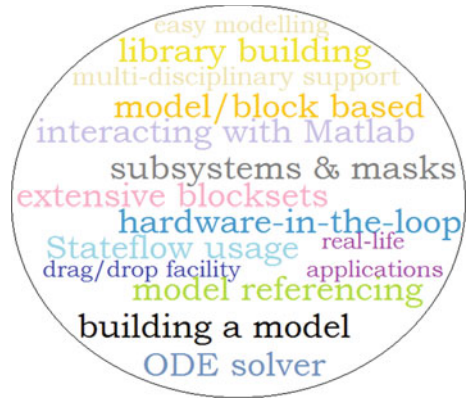
D. Jeyakumari · M. Sundar Prakash Balaji

Department of ECE, RVS College of Engineering and Technology, Coimbatore, India

P. Thanapal

School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

e-mail: thanapal.p@vit.ac.in

Fig. 1 Features of Simulink

depicts a few Simulink features. The following are three key components of any Simulink model [1–3]:

- Collection of Simulink blocks
- Inter-connections of the blocks
- Parameters of each block

For real-time applications, Simulink is frequently compatible with communicating (supporting) a variety of hardware boards. The following are some of the hardware boards that the Simulink package supports [4–6]:

- Universal Software Radio Peripheral (USRP)
- Arduino
- Raspberry Pi
- Texas Instruments (TI) DSP boards
- Xilinx FPGA boards (using Xilinx System Generator)
- Altera FPGA boards (using Altera DSP Builder)

A subsystem in Simulink is a single block that replaces a group of blocks. Simulink allows you to create a variety of subsystems, including conditional subsystems (see Fig. 2). Figure 3 depicts one such application of the subsystem. Figure 3 shows how to resample using an up-sampling followed by a down-sampling, which may be concatenated using “Ctrl + G” to construct the subsystem [7–9].

The following is a breakdown of how this article is structured. The second section contains the implementation details of the Simulink subsystems in elaborate with examples, as well as the materials and techniques. The third section contains simulation findings focusing on voiced/unvoiced categorization and RMS amplitude calculation. Finally, the fourth section discusses the conclusions.

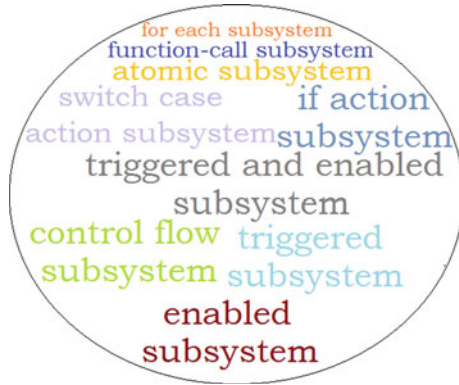


Fig. 2 Various subsystems available in Simulink

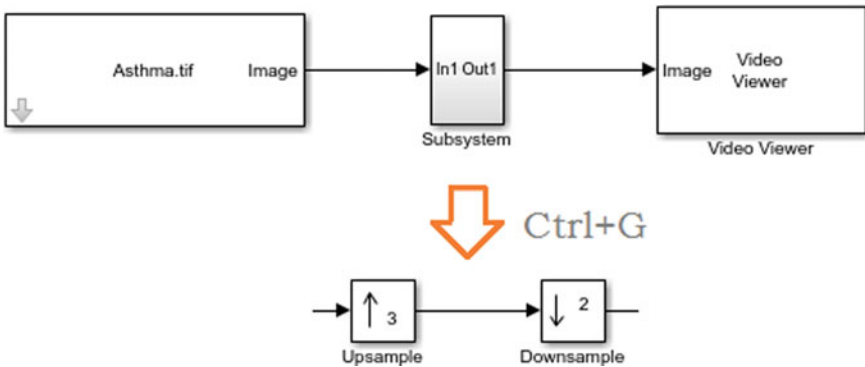


Fig. 3 Resampling “asthma” image using the subsystem

2 Materials and Methods

2.1 The Enabled Subsystem

The execution of a conditional subsystem is dependent on the value of the input signal. A conditional subsystem that functions when the control signal is affirmative is known as an enabled subsystem. Figure 4 shows an image segmentation using an enabled subsystem as an example. If the control signal crosses zero and becomes positive, the enabled subsystem begins to operate and continues to operate until the control signal remains positive.

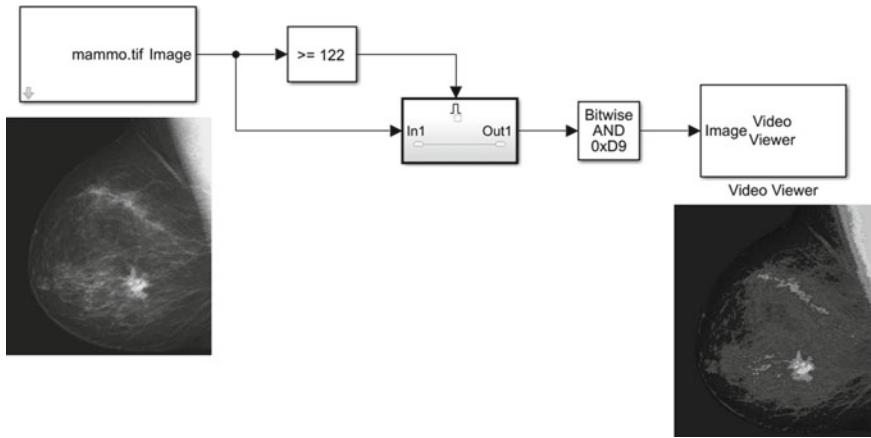


Fig. 4 Image segmentation using a conditional subsystem

2.2 The Triggered Subsystem

If a trigger event happens, the triggered subsystem is activated; the triggered event can be continuous or discrete. A trigger event can also occur on a trigger signal's rising or falling edge. Figure 5 depicts one such experiment: zero-crossing detection on a voice stream utilizing a triggered subsystem. This experiment demonstrates how to use a triggered subsystem to count the number of zero-crossings in a spoken signal. The zero-crossing rate characteristic is frequently used in speech processing applications such as pitch estimation and vocal activity recognition [10–12]. This functionality is also important for retrieving music data and speech recognition. Figure 6 depicts the input voice signal. Figure 7 shows a fragment of the spoken signal (top) and the output of a zero-crossing detector (bottom).

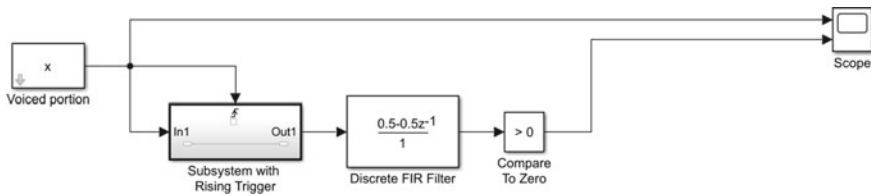


Fig. 5 Zero-crossing detection on a speech signal

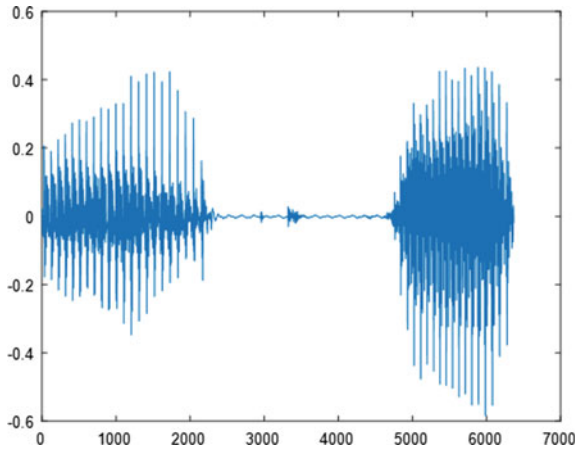


Fig. 6 Input voice signal

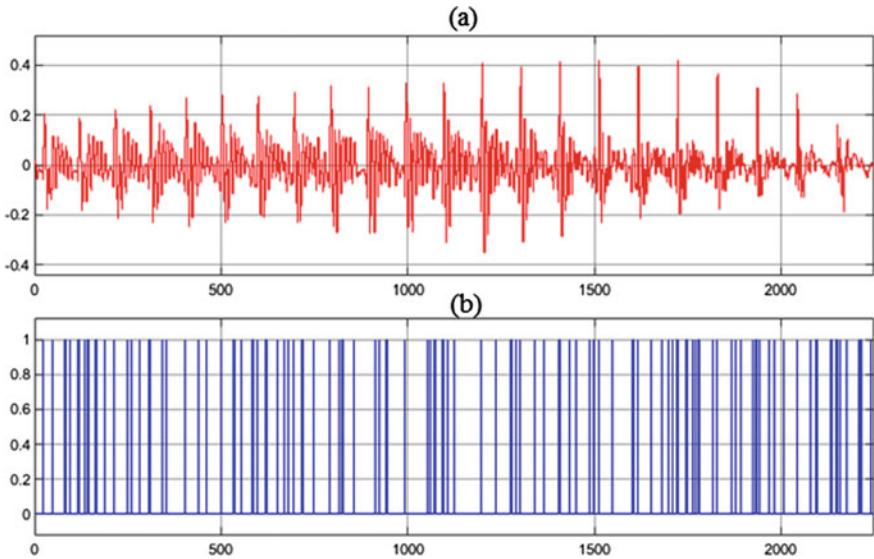


Fig. 7 Zero-crossing detection. a Input speech signal and b result of zero-crossing detection

2.3 The Control Flow Subsystem

During the enabling time of control flow, a control subsystem runs, which can be done with blocks like for, do, while, and if-then. Figure 8 shows one such scenario employing the control flow subsystem. Using If Action subsystems, this model is used to differentiate positive (top in Fig. 9) and negative (bottom in Fig. 9) magnitudes. The recognition of positive and negative magnitudes could aid in the development

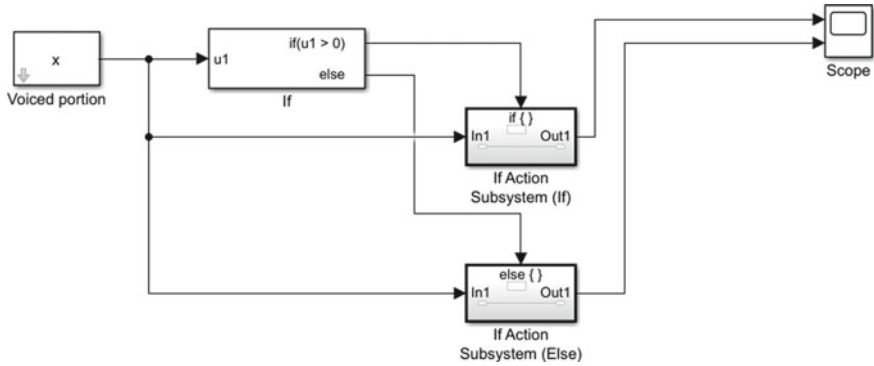


Fig. 8 Deployment of a control flow system for positive/negative magnitudes

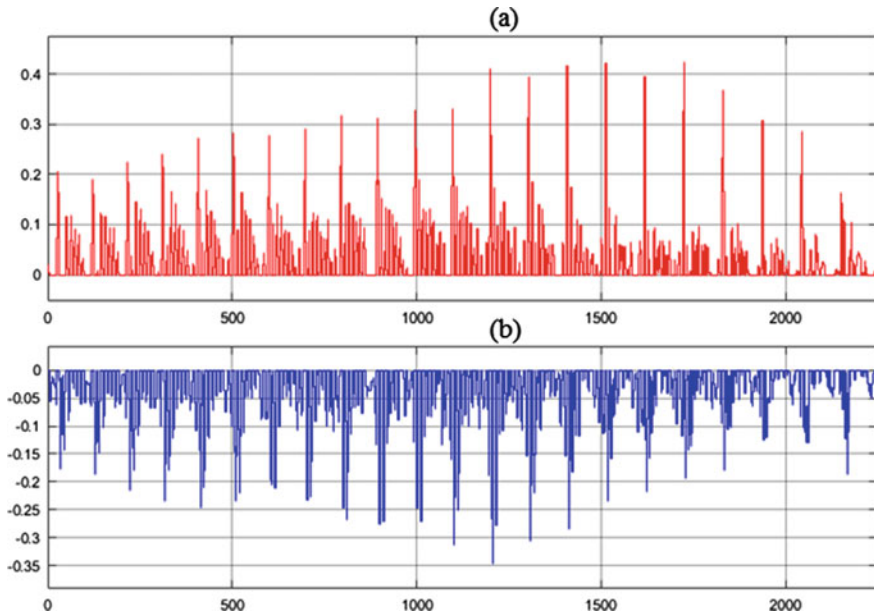


Fig. 9 Results of a control flow system. a Positive magnitudes and b negative magnitudes

of a model for calculating the root mean square (RMS) amplitude of a speech signal utilizing a control flow subsystem. The RMS value is frequently used to describe the signal's intensity.

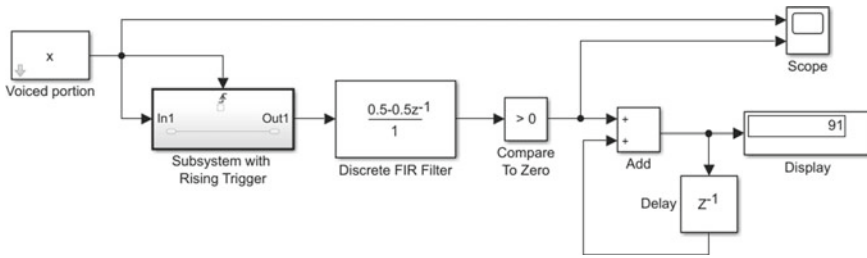


Fig. 10 Number of zero-crossings detection using a triggered subsystem

3 Results and Discussion

The results of the studies, such as voiced/unvoiced classification and RMS amplitude computation, are described in this section.

3.1 Voiced/Unvoiced Classification Using Triggered Subsystem

The number of zero-crossing detections in the speech input can be used to determine whether the speech is voiced or unvoiced. The steps to count the ones (zero-crossings) are as follows (concerning Fig. 5):

- Add an Add block from the Simulink Math Operations library to your model.
- Add a Delay block from the DSP System Toolbox Signal Operations library to your model.
- Add a Display block from the Simulink Sinks library to your model.
- Consider a model that connects Add, Delay, and Display blocks.

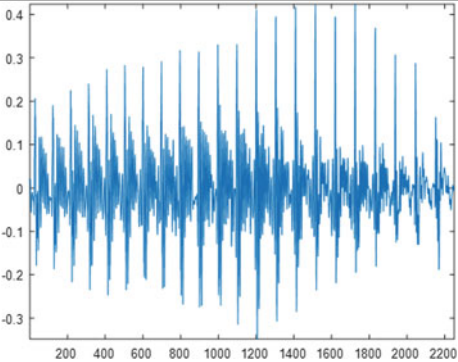
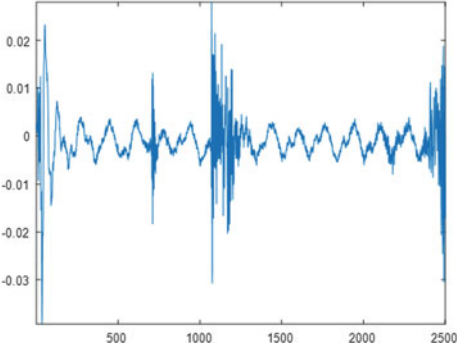
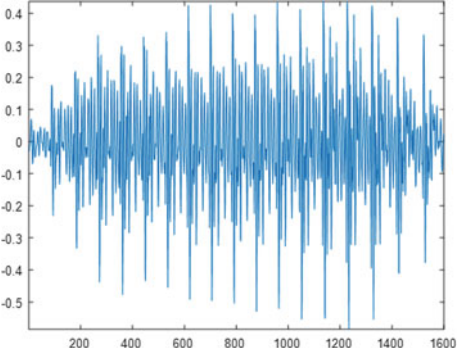
Figure 10 shows the resulting model for classifying the voiced/unvoiced portion. Table 1 shows the results of the classification (voiced or unvoiced). The value of the threshold is assumed to be 60 in this case.

3.2 Root Mean Square (RMS) Amplitude Calculation

Take the discrete-time signal x , which has N samples. Then, using the following formula, determine the root mean square (RMS) amplitude of the signal x

$$RMS = \sqrt{\frac{\sum x^2}{N}} \tag{1}$$

Table 1 Results of voiced/unvoiced classification

A portion of the voice signal	One's count	Classification result
	91	Voiced (since 91 is higher than the threshold)
	58	Unvoiced (since 58 is lower than the threshold)
	75	Voiced (since 91 is higher than the threshold)

The square root of the average of the squares of all the samples is the RMS amplitude. The steps are as follows (regarding Fig. 8) to compute and display the RMS amplitude:

- Add an Add block from the Simulink Math Operations library to your model.

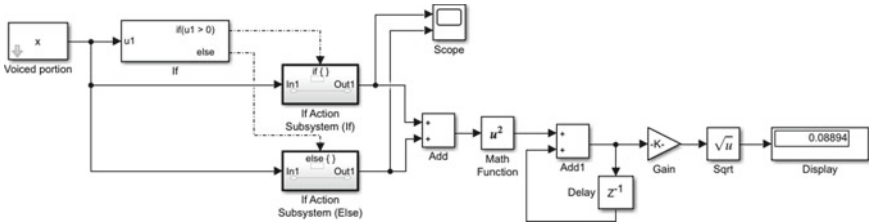


Fig. 11 RMS calculation using if action subsystems

- To perform a squaring operation, add a Math Function block from the Simulink Math Operations library to your model. Open a block parameter user interface by double-clicking the block. Then, set the function parameter to a square.
- To perform addition, again add an Add block from the Simulink Math Operations library to your model. Name this block as Add1. Also, add a Delay block from the DSP System Toolbox Signal Operations library to your model.
- To perform the division, add a Gain block from the Simulink Math Operations library to your model. Set the Gain block parameter to $1/\text{length}(x)$.
- To perform the square root, the Sqrt block from the Simulink Math Operations library to your model.
- Add a Display block Simulink Sinks library to your model.
- Consider a model that connects Add, Math Function, Add1, Delay, Gain, Sqrt, and Display blocks.

Figure 11 shows the resulting model for computing and displaying the RMS amplitude of the voiced part. The RMS amplitude (0.08894) obtained by a Simulink model can be confirmed using MATLAB in the following way:

```
>> sqrt(sum((x.^2))/2250)
ans = 0.0889
```

3.3 Advantages and Disadvantages

The advantages of this study are listed as below:

- The subsystem model is simple to read and comprehend.
- It aids in reducing the number of blocks visible in your model.
- It reduces the simulation model’s complexity.
- You can group functionally relevant blocks in the subsystem model.

A major disadvantage of Simulink utility is that the models for real-world power semiconductors cannot be directly incorporated. Also, the number of semiconductors

in the circuit can dramatically increase the complexity of the block diagram used to model the power circuit.

4 Hypothesis and Limitations

The main purpose of this study is to use the multiple subsystems of the Simulink software suite for biomedical signal and image processing. Simulink is not the best option for small projects on a tight budget: a MATLAB license is expensive, as are any additional packages needed for testing on dedicated hardware. It is best suited for huge projects with a large team of developers. An experienced C developer can write code for simple to trivial applications in a fraction of the time it takes to drag Simulink blocks. MATLAB and Simulink are not general-purpose programming tools; they are designed for engineering and scientific purposes. Developing desktop apps or web services with Simulink is tough, but not unachievable.

5 Conclusion

The major goal of this research is to employ the Simulink software package's many subsystems for biomedical signal and image processing. The enabled subsystem, triggered subsystem, and If Action subsystems are used to conduct three key experiments: picture segmentation, voiced/unvoiced classification, and RMS amplitude computation. This work can also be used to calculate other important performance measures including signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR, for pictures), image enhancement factor (IEF), structural similarity index (SSIM), and so on [13–16].

References

1. Cetinkunt S (2015) *Mechatronics with experiments. Essentials of electronic testing for digital, memory and mixed-signal VLSI circuits*, 2nd edn. Wiley
2. Naveena R, Rabecka D, Rajkumar G, Elamaran V (2015) Understanding digital filters-from theory to practice using matlab and simulink. *Int J Pharm Technol* 7:9923–9934
3. Elamaran V, Praveen A, Reddy MS, Aditya LV, Suman K (2012) FPGA implementation of spatial image filters using Xilinx system generator. *Procedia Eng* 38:2244–2249
4. Dabney JB, Harman TL (2003) *Mastering simulink*. Pearson Prentice Hall
5. Durga GAS, Divyasree N, Narasimhan K, Elamaran V, Upadhyay HN (2016) Reliability improvement by hardware redundancy using Altera DSP builder. *Int J Pharm Technol* 8:23253–23271
6. Elamaran V, Vaishnavi R, Rozario AM, Joseph SM, Cherian A (2013) CIC for decimation and interpolation using Xilinx system generator. In: *Proceedings of the 2013 IEEE international conference on communication and signal processing (ICCSP 2013)*, pp 622–626

7. Reddy BM, Subbareddy TV, Reddy SO, Elamaran V (2014) A tutorial review on data compression with detection of fetal heart beat from noisy ECG. In: Proceedings of the 2014 IEEE international conference on control, instrumentation, communication and computational technologies (ICCICCT), pp 1310–1314
8. Wang W, Zhang G, Yang Y, Balaji VS, Elamaran V, Arunkumar N (2019) Revisiting signal processing with spectrogram analysis on EEG, ECG and speech signals. *Futur Gener Comput Syst* 98:227–232
9. Cooper T (2013) *Advanced mathematics for FPGA and DSP programmers: conquering fixed-point pitfalls*. Amches Inc., USA
10. Sundar Prakash Balaji M, Jayabharathy R, Jegadeesan N, Devasena L, Venkat Babu G, Elamaran V, Venkatraman V (2020) Analysis of energy concentration of the speech, EEG, and ECG signals in healthcare applications—a survey. *J Med Imaging Health Inform* 10:49–53
11. Sundar Prakash Balaji M, Jayabharathy R, Betty Martin A, Parvathy A, Arvind Shiram RK, Elamaran V (2020) Exploring modern digital signal processing techniques on physiological signals in day-to-day life applications. *J Med Imaging Health Inform* 10:93–98
12. Newbold R (2012) *Practical applications in digital signal processing*. Prentice Hall
13. Sasi G, Vimala R, Sivaaraju SS, Ramya R, Elamaran V (2021) FPGA implementation of DC bias removal filters—a case study with an ECG signal. In: Proceedings of the 3rd international conference on signal processing and communication (ICPSC), pp 1–5
14. Shoba S, Rajavel R (2018) Image processing techniques for segments grouping in monaural speech separation. *Circ Syst Sig Process* 37:3651–3670
15. Athilakshmi R, Rajavel R, Jacob SG (2019) Fusion feature selection: new insights into feature subset detection in biological data mining. *Stud Inform Control* 28:327–336
16. Mohan SB, Raghavendiran TA, Rajavel R (2019) Patch based fast noise level estimation using DCT and standard deviation. *Cluster Comput J Netw Softw Tools Appl* 22:14495–14504

Hum Noise and Breathing Interference Removal from an ECG Signal with Finite Word Length Effects



B. Ajith Bala, D. Jeyakumari, M. Sundar Prakash Balaji, G. Sasi,
and V. Elamaran

Abstract When it comes to discrete-time systems, whether it is an FIR or an IIR filter, it is important to understand the complexities of data representation, data operation, and so on. The system should be able to handle problems such as arithmetic overflow, truncation, rounding, and so on. With actual examples, this study demonstrates the impact of quantization on the filter coefficients. To investigate the relevance of finite word length effects, application examples such as hum noise reduction and breathing interference removal from an ECG signal are investigated. With the use of biomedical signal processing examples, the frequency responses, impulse responses, and pole-zero plots of the original and quantized systems are discussed. In coding with 5-bits, 6-bits, 7-bits, and 8-bits, quantization effects are used. Furthermore, an elliptic filter is taken into account while comparing the simulated outcomes. For experimental simulations with an ECG signal, the MATLAB 9.4 tool is employed.

Keywords ECG signal · Hum noise removal · Respiratory effects · MATLAB · Quantization

1 Introduction

If the physical quantities are expressed numerically, then the quantization or round-off occurs in such a scenario. The real measurements (analog quantities) in a digital system design can be represented by integers, which are the nearest numbers of units. The values of round-off noise range from minus one-half unit to plus one-half unit.

B. Ajith Bala · V. Elamaran (✉)

Department of ECE, School of EEE, SASTRA Deemed University, Thanjavur, India

e-mail: elamaran@ece.sastra.edu

D. Jeyakumari · M. Sundar Prakash Balaji

Department of ECE, RVS College of Engineering and Technology, Coimbatore, India

G. Sasi

Department of BME, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India

The following are a few examples of analog values that are represented by discrete numbers [1]:

- Time display by a digital watch
- Weight display by a weighing machine
- Temperature indication by a thermometer
- Distance signals on a map

In their respective disciplines, people working in digital signal processing, numerical analysis and computation, digital communications, and control systems must be aware of quantization noise or round-off error. Signal quantization is a nonlinear procedure that is difficult to directly examine. Rather than focusing on the signal being quantized, we will look at its probability density function.

Discrete signals are time sampled and amplitude quantized. Sampling theory can be used to study the granularity of such signals, which is created by both sampling and quantization. Figure 1 illustrates the quantization noise for a 2-bits ADC operating at an infinite sample rate. The quantization error, which is “added” to the quantized signal and is the source of noise in the upper graph, is the difference between the blue and red signals (see Fig. 2). Figure 2 depicts the five probable sources of quantization effects; these effects are almost non-existent in analog circuit designs [2–4]. Figure 3

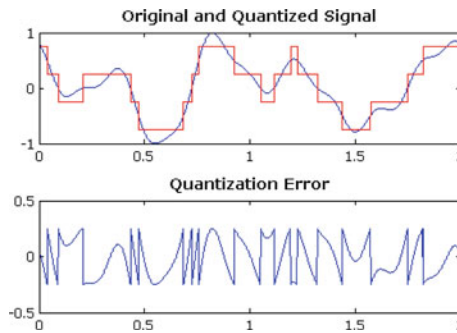


Fig. 1 Presence of quantization noise for a 2-bits ADC with an unlimited sample rate

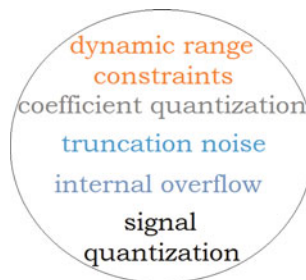


Fig. 2 Sources of quantization effects in digital filters

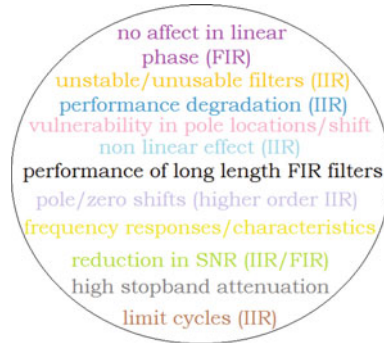


Fig. 3 Finite word length effects on FIR and IIR filters

depicts the important implications of limiting word length effects on digital FIR and IIR filters. Each extra bit allows for a 6 dB increase in dynamic range and a 6 dB reduction in quantization noise during quantization. Audio CD players, for example, have a dynamic range of 96 dB and a signal-to-noise ratio (SNR) of 96 dB since they are 16-bits wide [5, 6]. The quantization effects on the 2D image are shown in Fig. 4.

In FIR filters, the effect is not as strong; the FIR filters are always stable, regardless of quantization noise. In the case of notch filters, for example, unit circle zeros are preserved due to limited word length effects, infinite notches occur without problems, but notch frequency fluctuates slightly. Because of the symmetry property, the linear phase FIR filters preserve their characteristic (linear phase) despite the finite word length effects [3].

There is a possibility of altering pole placements—from inside the unit circle to outside the unit circle—in IIR filters due to quantization errors in the recursive feedback stage. As a result of the finite word length effects, stability is a major concern in IIR filters. Poles that are close to the unit circle are particularly vulnerable in this regard. Because the reduced sampling rate moves poles back to within the unit circle, decimation can be employed to mitigate this stability issue. If this is not possible in



Fig. 4 Quantization effects on the 2D image. a Original 8-bits image and b 3-bits quantized image

specific applications, double precision arithmetic could be used to solve the problem [7, 8].

The following is a breakdown of how this article is structured. The second section contains the materials and methods, which include practical digital filtering experiments such as removing hum noise and breathing interference from an ECG signal. The third section presents simulation results concentrating on quantized and original filtered coefficients. Finally, the fourth section discusses the conclusions.

2 Materials and Methods

2.1 Quantization Effect on Hum Noise Removal

Figure 5a depicts the ECG signal, which is recorded at 250 Hz, and Fig. 5b shows the simulated 60 Hz tone (a hum noise). To reject hum noise, a simple notch filter with a band of frequencies between 55 and 65 Hz is used. To reduce the hum noise, a 25-tap FIR filter with a Hamming window is utilized, as reported in MATLAB [9–11]:

```
>> fir1(24, [0.44 0.52], 'stop')
```

where the 0.44 and 0.52 indicate normalized lower and upper cutoff frequencies of a band-stop filter, respectively. Figure 6 depicts the filter's impulse responses, both the actual and quantized using 5-bits. The input noisy and filtered ECG signals are shown in Fig. 7. Coding 5-bits, 6-bits, 7-bits, and 8-bits results in quantization effects. Figure 8 shows the original and equivalent frequency responses, as well as the pole-zero plots of the filter for the hum noise removal task. The quantized filter

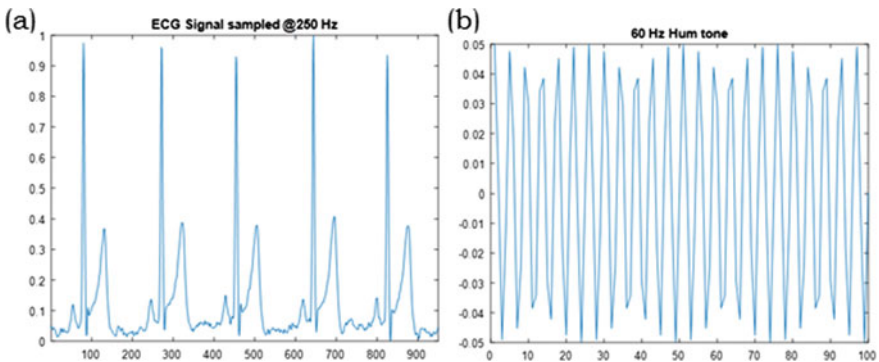


Fig. 5 ECG signal and a simulated hum noise

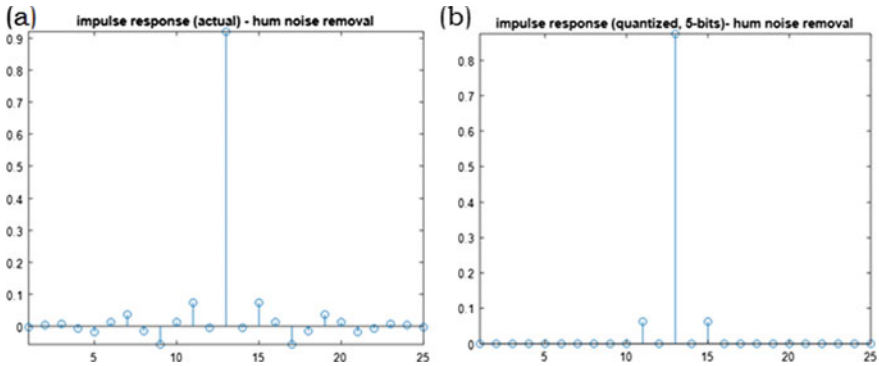


Fig. 6 Impulse responses of the filter

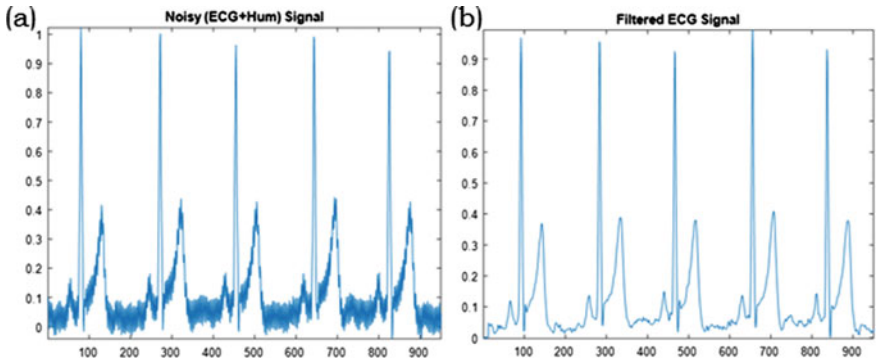


Fig. 7 Noisy and filtered ECG signals

frequency responses provide poor performance due to finite-length word effects, as shown in the frequency response diagram.

2.2 Quantization Effect on Respiratory Effects Removal

The respiratory effects (breath signal) play a critical part in physicians’ decision-making during ECG signal recording. In that instance, the breath signal interference (about 8 Hz) should be removed from an ECG signal for better diagnosis. The respiratory effects are rejected using a simple high-pass filter with an 8-Hz cutoff frequency. To remove the breath signal, a 9-tap FIR filter with a Hamming window is employed, as stated in MATLAB.

```
>> fir1(8, 0.0640, 'high')
```

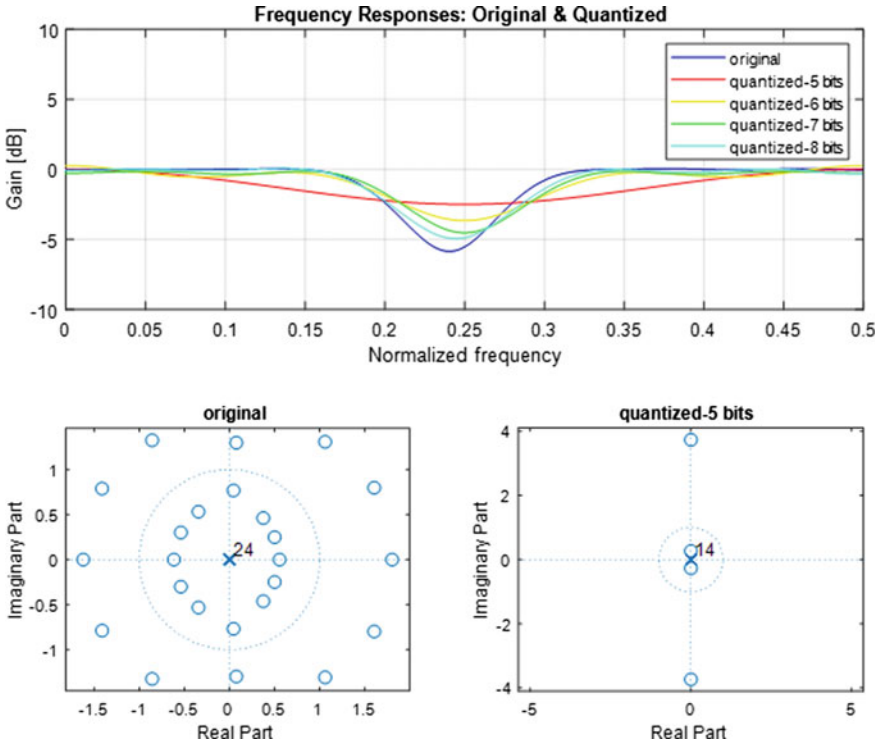


Fig. 8 Hum noise removal: frequency responses and pole-zero plots

where 0.0640 indicates the normalized cutoff frequency of a high-pass filter. Figure 9 depicts the filter's impulse responses, both the actual and quantized using 5-bits. The input noisy (breathing) and filtered ECG data are shown in Fig. 10. Coding 5-bits, 6-bits, 7-bits, and 8-bits results in quantization effects. Figure 11 shows the original

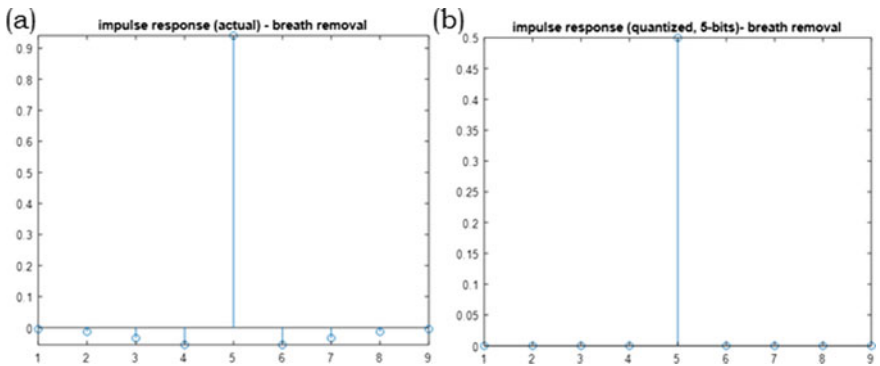


Fig. 9 Impulse responses of the filter

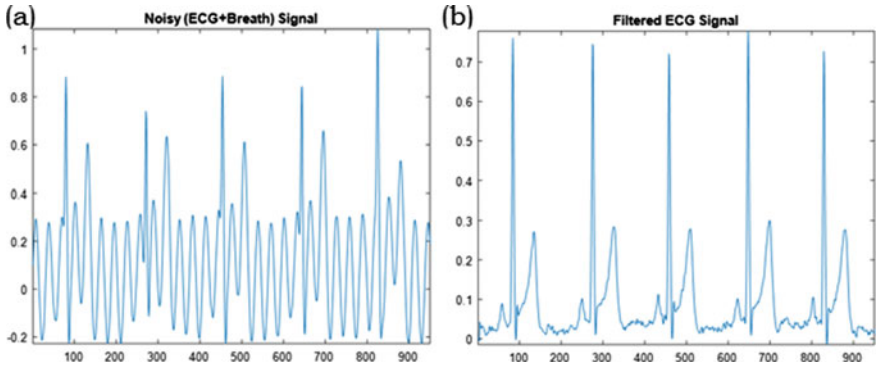


Fig. 10 Noisy and filtered ECG signals (breath removal)

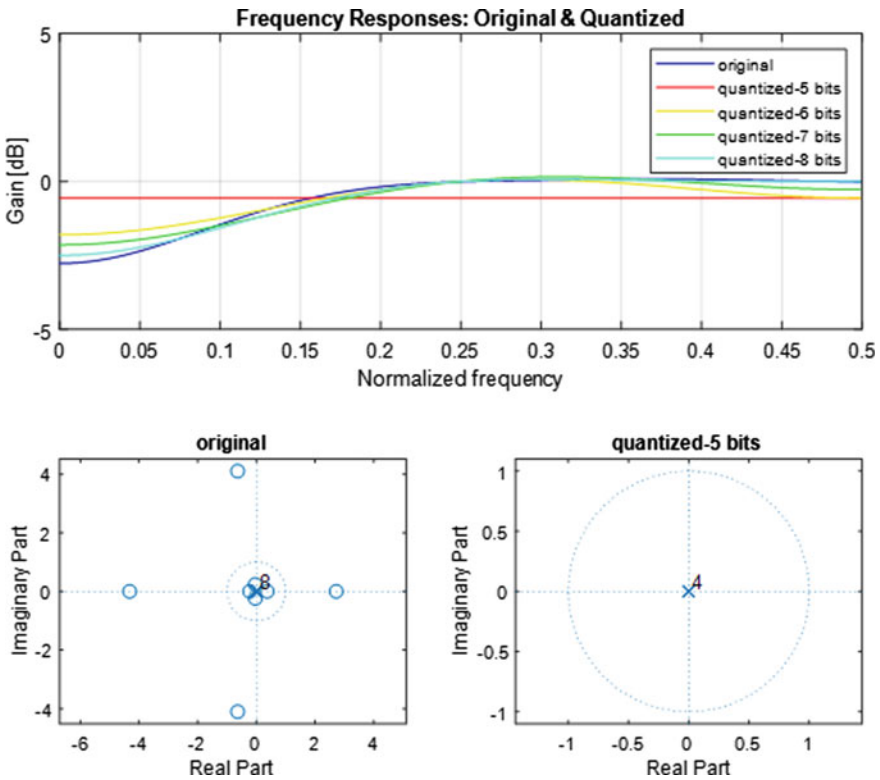


Fig. 11 Breathing interference removal: frequency responses and pole-zero plots

and equivalent frequency responses, as well as the pole-zero plots of the filter for the task of breathing removal. The quantized filter's performance is decreased, as evidenced by the frequency response data [9–11].

2.3 Hum Noise Removal Using an Elliptic Filter

To reduce hum noise from an ECG signal, an elliptic filter is used. The following is a description of a sixth order band-pass filter in MATLAB [9–11]:

```
>> [b, a] = ellip(6, 5, 80, [.44, .52], 'stop')
```

Figure 12 shows the numerator and denominator coefficients (both actual and quantized). The filtered ECG signal employing an elliptic filter is shown in Fig. 13. Figure 14 shows the frequency responses and pole-zero plot of an elliptic filter. As expected, the trade-off holds for elliptic filters as well; that is, filters with fewer quantized digits perform worse than those with higher precise representation.

3 Results and Discussion

The actual and quantization coefficients for both hum noise removal and breathing interference removal are discussed in this section.

```
>> b
b =
    0.3628 -0.2746  2.2485 -1.3799  5.7284 -2.7667  7.6855 -2.7667  5.7284 -1.3799  2.2485 -0.2746  0.3628
>> a
a =
    1.0000 -0.6913  5.1225 -2.9096 11.1155 -5.0463 13.1393 -4.5324  8.9878 -2.1249  3.4010 -0.4200  0.5613
>> bq
bq =
    0.3125 -0.2500  2.1875 -1.3750  5.6875 -2.7500  7.6250 -2.7500  5.6875 -1.3750  2.1875 -0.2500  0.3125
>> aq
aq =
    1.0000 -0.6875  5.0625 -2.8750 11.0625 -5.0000 13.1250 -4.5000  8.9375 -2.0625  3.3750 -0.3750  0.5000
```

Fig. 12 Numerator and denominator coefficients of an elliptic filter

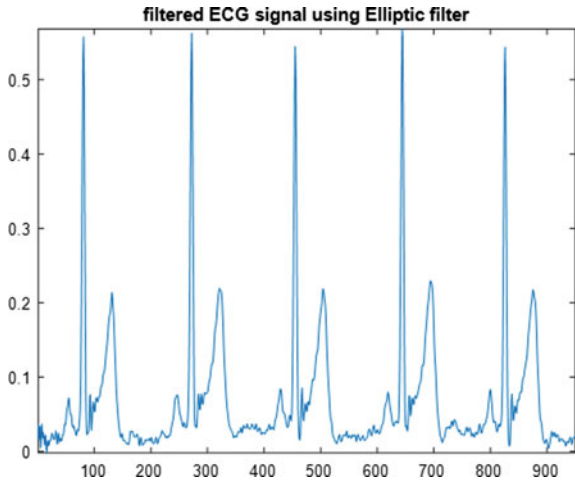


Fig. 13 Filtered ECG signal using an elliptic filter

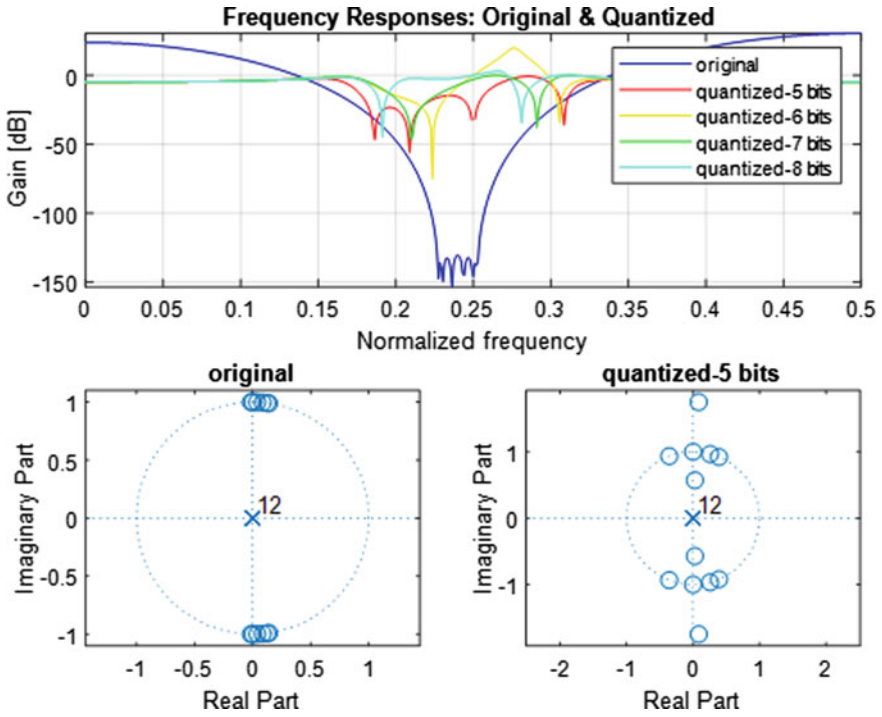


Fig. 14 Hum noise removal using an elliptic filter: frequency responses and pole-zero plots

Table 1 Portion of the original and quantized filter coefficients (hum noise removal)

Type	Filter coefficient position							
	11th	12th	13th	14th	15th	16th	17th	18th
Original	0.0737	-0.0049	0.9205	-0.0049	0.0737	0.0127	-0.0572	0.0152
Quantized (5-bits)	0.0625	0	0.8750	0	0.0625	0	0	0
Quantized (6-bits)	0.0625	0	0.9063	0	0.0625	0	-0.0313	0
Quantized (7-bits)	0.0625	0	0.9063	0	0.0625	0	-0.0469	0
Quantized (8-bits)	0.0703	0	0.9141	0	0.0703	0.0078	-0.0547	-0.0078

3.1 Quantization Effects on Hum Noise Removal

Table 1 shows the actual and quantized filter coefficients (just the fraction) used to remove hum noise from an ECG signal during the hum noise removal process. As expected, as precision grows, the discrepancies, or quantization errors, get less. The quantization with 8-bits provides higher performance than the others in the last row, but at the cost of additional processing resources.

3.2 Quantization Effects on Hum Noise Removal

The actual and quantized filter coefficients of an 8-tap FIR filter during the elimination of hum noise from an ECG signal are shown in Table 2. Only the first five filter coefficients are shown due to the symmetry in the linear phase characteristic. The quantization sounds are reduced with higher precision arithmetic, as expected.

Table 2 Portion of the original and quantized filter coefficients (breath removal)

Type	Filter coefficient position				
	1st	2nd	3rd	4th	5th
Original	-0.0046	-0.0130	-0.0338	-0.0553	0.9403
Quantized (5-bits)	0	0	0	0	0.9375
Quantized (6-bits)	0	0	-0.0313	-0.0313	0.9375
Quantized (7-bits)	0	0	-0.0313	-0.0469	0.9375
Quantized (8-bits)	0	-0.0078	-0.0313	-0.0547	0.9375

4 Conclusion

This article's main purpose is to highlight the significance of finite word length effects. Any system that spends more computational resources and so occupies a greater area and dissipates more electricity will get better performance results. On the other hand, fewer quantization levels can be used at the expense of performance to produce a smaller area and lower power dissipation. This in-depth investigation investigates the effect of quantization in both FIR and IIR filters. This approach can be expanded to various quantization applications, such as digital dithering, a quantization noise spectrum, and whitening criteria [12–15].

References

1. Widrow B, Kollar I (2008) Quantization noise: roundoff error in digital computation, signal processing, control, and communications. Cambridge University Press, Cambridge
2. Reddy BM, Subbareddy TV, Reddy SO, Elamaran V (2014) A tutorial review on data compression with detection of fetal heart beat from noisy ECG. In: Proceedings of the 2014 IEEE international conference on control, instrumentation, communication and computational technologies (ICCICCT), pp 1310–1314
3. Bateman A, Paterson-Stephens I (2001) The DSP handbook: algorithms, applications, and design techniques. Prentice Hall, Hoboken
4. Cooper T (2013) Advanced mathematics for FPGA and DSP programmers: conquering fixed-point pitfalls. Amches Inc., USA
5. Naveena R, Rabecka D, Rajkumar G, Elamaran V (2015) Understanding digital filters—from theory to practice using matlab and simulink. *Int J Pharm Technol* 7:9923–9934
6. Elamaran V, Vaishnavi R, Rozario AM, Joseph SM, Cherian A (2013) CIC for decimation and interpolation using Xilinx system generator. In: Proceedings of the 2013 IEEE international conference on communication and signal processing (ICCSIP 2013), pp 622–626
7. Sundar Prakash Balaji M, Jayabharathy R, Jegadeesan N, Devasena L, Venkat Babu G, Elamaran V, Venkatraman V (2020) Analysis of energy concentration of the speech, EEG, and ECG signals in healthcare applications—a survey. *J Med Imaging Health Inform* 10:49–53
8. Sundar Prakash Balaji M, Jayabharathy R, Betty Martin A, Parvathy A, Arvind Shriram RK, Elamaran V (2020) Exploring modern digital signal processing techniques on physiological signals in day-to-day life applications. *J Med Imaging Health Inform* 10:93–98
9. Elamaran V, Praveen A, Reddy MS, Aditya LV, Suman K (2012) FPGA implementation of spatial image filters using Xilinx system generator. *Procedia Eng* 38:2244–2249
10. Durga GAS, Divyasree N, Narasimhan K, Elamaran V, Upadhyay HN (2016) Reliability improvement by hardware redundancy using Altera DSP builder. *Int J Pharm Technol* 8:23253–23271
11. Wang W, Zhang G, Yang Y, Balaji VS, Elamaran V, Arunkumar N (2019) Revisiting signal processing with spectrogram analysis on EEG, ECG and speech signals. *Futur Gener Comput Syst* 98:227–232
12. Sasi G, Vimala R, Sivaaraju SS, Ramya R, Elamaran V (2021) FPGA Implementation of DC bias removal filters—a case study with an ECG signal. In: Proceedings of the 3rd international conference on signal processing and communication (ICSPSC), pp 1–5

13. Shoba S, Rajavel R (2018) Image processing techniques for segments grouping in monaural speech separation. *Circ Syst Sig Process* 37:3651–3670
14. Athilakshmi R, Rajavel R, Jacob SG (2019) Fusion feature selection: new insights into feature subset detection in biological data mining. *Stud Inform Control* 28:327–336
15. Mohan SB, Raghavendiran TA, Rajavel R (2019) Patch based fast noise level estimation using DCT and standard deviation. *Cluster Comput J Netw Softw Tools Appl* 22:14495–14504

Online Affinity of Instructing Methods for Using Personal-Efficacy and Reading Capacious During Covid-19



V. PremaLatha, E. Sreedevi, S. Sivakumar , Soumya Ranjan Nayak , Akash Kumar Bhoi , and Uttam Ghosh 

Abstract Now a days in EFL procedure of education the ability of reading became as significant belief and personal-efficacy reading as a basic understanding for students. By monitoring the acknowledged participates under the ballpark figure of large studying and methods of understanding, the impact of their observation is pre-meditated on reading of each one's personal-efficacy. On a daily routine all these things are comparatively considered which are put into effect by teachers of handful in number. Approach towards exhibiting Extensive reading (ER) is inspected to be "more expensive, difficult, and time-consuming". Method of recognition of elements in a various way for effective impact in putting its efforts to utilize for its empowerment. Paper has been segregated into two contexts: Association with attitude is considered as primary one and attitude is considered as secondary one. Whether knowledge work is understood by student or not is considered as the impact of ER by the first review. Procedure which are convenient is taken as the observations of

V. PremaLatha · E. Sreedevi

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, K L University, Vaddeswaram, Guntur 522502, India
e-mail: premaWilliams@kluniversity.in

E. Sreedevi

e-mail: sridevifed@kluniversity.in

S. Sivakumar

Department of Computer Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur 603203, India

S. R. Nayak (✉)

Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India
e-mail: nayak.soumya17@gmail.com

A. K. Bhoi

KIET Group of Institutions, Delhi-NCR, Ghaziabad 201206, India

Directorate of Research, Sikkim Manipal University, Gangtok, Sikkim 737102, India

U. Ghosh

Department of Computer Science and Data Science Meharry School of Computational Sciences Nashville, Nashville, TN, USA
e-mail: ghosh.uttam@ieee.org

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
S. Dhar et al. (eds.), *Advances in Communication, Devices and Networking*, Lecture Notes in Electrical Engineering 902, https://doi.org/10.1007/978-981-19-2004-2_20

student and is analysed as second one. The examinations are quantifiable to utilize the observations as information in terms of subjective way taken from students who belong to first academic year of reading course in a systematic way and 603 details of undergraduate students from KLEF of Guntur were chosen as participants for extant examination. In ER programme of includes and excludes “comprehension reading work” is treated as fundamental in the proposal of disclosures. In case of any, “the programme appeared to positively affect contributing students”. Techniques of classification like “decision tree and Mixed Model Database Miner (MMDBM)” are employed in this paper which leads to improvements of post-test to pre-test in ER group. Observations of students in ER results as optimistic and algorithm of MMDBM which leads to accuracy in higher rate in pre-test and post-test detection.

Keywords Personal-efficacy · Affinity · MMDBM · Foreign Language(EFL) · Gini index · Chi-square

1 Introduction

Impetus of doing paper is to examine the personal-efficacy reading of student’s that can be revised by the usage of “extensive reading materials” which are Internet-based. Control Cluster Model was utilized for pre-test and post-test. Picturizing English language as a Foreign Language (EFL) for quite a whilst but analysts of Iran seem issues for English learners which are expected not to represent the content in English language; this may promptly a major issue arising a problem for learning the language.

When it comes in terms of reading reliable English texts for example like “books, newspapers, online English texts, and so on” became as an open source to gain English language in real. Hence to improve and build up in language, learners need to contribute for extensive language reading in English. Riding a bicycle which can be knowledgeable for cycle riding can be just analysed by reading. Validity of Extensive reading in which basic belief and significant in best are measured [1]. Drafted for equitably for learners of primary and second language, Extensive reading has been examined mostly and all of the welfares are well-known [2]. By its constant and positive usage there are reductions in many interruptions and problems [3].

In the midst of several disadvantages, a lecturer need to overthrow exactly in undeveloped or poor countries like : “(a) scarcity of understanding materials; (b) shortage of sufficient training of lecturers; (c) burdens to cover the whole curriculum and text books and no time to perform programme such as ER; (d) inflexibility and assessment burdens of necessary assessment deeds; and finally (e) the detail that part of entire guidance, as the means to achieve knowledge and lifetime skills are still held belief. However, reading has been the utmost harassed skill in traditional EFL teaching. Currently, there is vast pressure on EFL reading lecturers to move after teaching texts to teaching [4]”. Without the support of anyone else students (who has the ability in level of specific to English) can ascertain read and learn which has

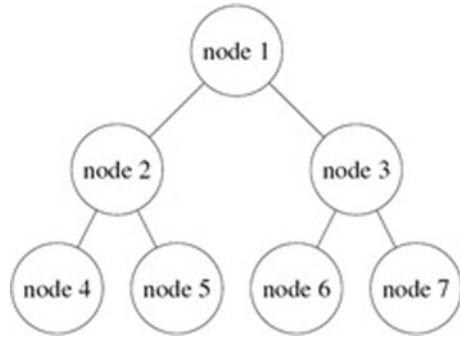
just enough proof [3]. To attain an objective for tool of predictable practically as ER turns into a treasured. Conveying data adequately about its angles in education for its handiness are the marks that are missed in the ER system study and in its decapitation. Having a great correlation with one another for the effects and personal-efficacy and settings for EFL and English leads to be motivating in giving clear explanations for the literature Revising [5, 6]. There is hefty association for each other when reading personal-efficacy and inspiration in understanding which has been demonstrated. The impact of personal-efficacy for inspiring students in comprehensive reading prompts to exercise interpretation for the inspiration activities of students for their degrees of dissimilar for their EFL classes and study for comments in casual. To be in a reasonable way and to maintain acting in psyche way which are variations in behavioural scheme the merges personal-efficacy that was estimated by Rwze Raissi [5]. Certain behaviours impact on few issues which reflects the level of personal-efficacy.

2 Methodology

A quantifiable way for dealing with questions and answers for researchers that are utilized in the existing study. Not changing in the treatment and the control of groups when taking a semi concentrate test. Based on specific tests the data were attained in two ways A comprehensive test in examining and survey of personal-efficiency examining. In a single semester for the study of the campus of Vijayawada and Guntur of KL University students were given tests to for reading courses in English. To concentrate and to be in distinct for controlling and experimenting in groups understudying is done by segregating into two notable groups considering from 60 nonidentical colleges of applicants.

When tested half of the colleges were obtained by procedures understanding as a guidance to express resulting as a solution. Whereas another half of the colleges were good in group control as a guidance of perusing in broad without indubitable for system understanding. The integrated data of techniques for assortment crop up inside in a semester for an academic about 4 months. To carry out the examination for our method proposed by utilizing tools of Weka 3.6 form all the collected data that is been evaluated. Classification methods are used here by the researchers for the study. In Data Mining when it comes to topics like “decision trees, Bayesian classification, Regression models, Best first-(BF) Tree, Functional Trees-(FT) tree, Random Forest-(RF) tree, Random tree-(RT), Alternative Decision-(AD) tree” [7, 8] in which classification method is essential techniques that is used. In recent days for research essential decision tree includes certain algorithms like “SLIQ, SPRINT, and MMDBM which was extensively utilized”. “Precision is calculated, recall, F-measure, and accuracy” [9–12] were equalled with algorithm of MMDBM and Decision tree.

Fig. 1 Algorithm for decision tree



2.1 Decision Tree

For a tree structure procedure, models of classification or regression are constructed by Decision Tree-(DT). The problem is described by using a DT for demonstration that includes class label representing leaf node each and internal nodes representing attributes in the tree which is referred in Fig. 1. Attributes of discrete help for decision tree [13, 14] for representing any functions in Boolean. Conclusions of initial reading and at end of semester for each group extensive approaches of reading are verified. To verify the level of knowledge at the begin of semester and at the end of semester in personal-efficacy and comprehensive reading of all the participants in this work, whether it is improved in personal-efficacy reading or not is detailed in the next section.

3 Splitting Point Methods in Decision Tree

3.1 Variance

Algorithm of regression in which technique of variance for node in midpoint when applied for variables of object that are in continuous. The feature of determining its quantity on a particular node is carved into child node or terminal node [12] where the variance is utilized.

$$\text{Variance} = \frac{\sum(X - \mu)^2}{N} \tag{1}$$

Value of zero is utilized in variance for calculation node for similarity for certain nodes that are equal.

3.2 Information Gain

Reduction of variation never cuts, if attribute in a categorical variable of objective appears in case of? By using following entropy final node when categorical for points of node when splitting is done and information gain is the solution for it.

$$\text{Information Gain} = 1 - \text{Entropy} \tag{2}$$

By utilizing entropy computation of transparency of node is done. “Node will be high in purity if entropy value is low”. Zero value of entropy is given for nodes that are equal. Node of purity having entropy of 1 is subtracted from nodes having maximum of 1 as value [12–14] in information gain. The below formula is utilized for calculating entropy:

$$\text{Entropy} = - \sum_{i=1}^n p_i \log_2 p_i \tag{3}$$

3.3 Gini Index

In case of Categorical variable Gini index is used for splitting node. In a DT, Gini index is one of the best for node splitting and is simple. The calculation for Gini index is:

$$\text{Gain Index} = 1 - \text{Gain} \tag{4}$$

Nodes that are considered which are randomly selected to classified elements for appropriate probability in distribution of labels. Gini index can be learned by:

$$\text{Gain Index} = 1 - \sum_{i=1}^n p_i^2 \tag{5}$$

“Node homogeneity will be greater when the Gini index is less.” Value of Zero is considered as the node of pure.

3.4 Chi-Square

Values of objects that are in nature of categorical in a DT of a node in which Chi-Square method is utilized for two-splitting-point. Difference of statistical by using a formula for node of parent and child is calculated for splitting point [10, 12–14].

$$\text{Chi-Square value is: Chi-Square} = \sqrt{\frac{(\text{Actual} - \text{Expected})^2}{\text{Expected}}} \quad (6)$$

In a parent node the established classes and distribution is done by class predicted which is expected from a node of child. Value of real is stored in nodes of terminal with value of actual finally.

4 MMDBM Algorithm

Datasets of nodes of each for splitting point and decision tree is focused in algorithm of MMDBM. SPRINT and SLIQ are two models of categorically are used in this method where in datasets of attributes whilst dealing with numeric are depended of algorithm of Decision Tree. Perfect DT of tree model is obtained by using MMDBM. To reduce cost of evaluation in attributes for numeric for a tree in a stage of construction was developed for arrangements that are improved. Classification of disk-resident datasets for construction uses technique of evolving breadth-first tree that is synchronized for the approach here. Categorical features for attributes of a SLIQ evolves a agile algorithm for split point classification. Another Datasets massively known (MMDBM) [12] in a DT is utilized in terms of data in categorical way and numerical control.

4.1 Evaluation Results

Students of under graduation of Vaddeswaram KLEF in which certain first year students are manually utilized for this course of reading collected as datasets in this study. Algorithm of Decision tree and algorithm of MMDBM are equalled to “evaluate precision, recall, F-Measure, and accuracy for the purpose of examinations”. Techniques of evaluation are needed for approaches of Data mining. Algorithm of MMDBM produced model is used to authenticate the technique that is used. Metrics like F-Measure, recall, and precision, performance and matrix confusion [10–12] using distinct models of classification to evaluate the labels of data in Supervised learning. P-Pre-test control, Q-Post-test control, R-Pre-test exp, S-Post-test exp Data classification of incorrect represents features that are left out and elements that are diagonal represent the appropriate classified ones. “Ratio between both true positive and false positive, true positive are represented for precision”. Precision: Searching is done from applicable dataset that are connected from data that was retrieved representing precision which is a retrieval of data. Table 1 indicates the matrix confusion in which the correctly classified instances that are indicated based on their occurrence. Incorrect classification of data represents error positive and correct data represents true positive.

Table 1 Matrix of confusion

	P	Q	R	S
P	tpP	ePQ	ePR	ePS
Q	eQP	tpQ	eQR	eQS
R	eRP	eRQ	tpR	eRS
S	eSP	eSQ	eSR	tpS

$$\text{Precision} = \frac{\text{tpP}}{\text{tpP} + \text{eQP} + \text{eRP} + \text{ePS}} \tag{7}$$

“where tpP is represents true positive for the class P and eQP, eRP, and ePS indicates as false negative”. Queries which are successful when applicable from the datasets that are connected from data retrieving portion implies the recall which is a retrieval of data. Ratio for true positive, error negative and true positive and correct classification are the instance of numbers for total representing recall.

$$\text{Recall} = \frac{\text{tpP}}{\text{tpP} + \text{ePQ} + \text{ePR} + \text{ePS}} \tag{8}$$

“where tpP is indicates true positive for the class P and ePQ, ePR and ePS are implies error positive”. Precision and recall for harmonic mean are evaluated by this.

$$\text{F-Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

Accuracy: Outcomes of true, true positive and true negative of all amounts are calculated from the total gathered data.

$$\text{Accuracy} = \frac{(\text{tpP} + \text{tpQ} + \text{tpR} + \text{tpS})}{(\text{tpP} + \text{eQP} + \text{eRP} + \text{eSP} + \text{ePQ} + \text{tpP} + \text{eRQ} + \text{eSQ} + \text{ePR} + \text{eQR} + \text{tpR} + \text{eSR} + \text{ePS} + \text{eQS} + \text{eRS} + \text{tpS})} \tag{10}$$

Algorithm of Classification of decision tree in which, “it runs of test datasets and records of each are processed.” The records are analysed into “pre-test control and experiment and post-test control and experiment”. “Classifier that was proposed are verified for obtaining results”. “Tables 2 and 3 will be representing the results”. Tables represents the records of total for proposed model that are employed. “Figure 4 represents the results that are precision, recall, and F-Measure”. Tables 2 and 3 represents the nodes that are classified which has been able to do by the method (MMDBM) Proposed one. “Table 2 and Fig. 2 represents the data sets that are utilizing the pre-test, post-test control and experiments that are compared with classes and algorithm of decision tree for matrix of confusion” (Fig. 3).

Table 2 Algorithm of MMDBM for confusion matrices, precision, recall, and accuracy

Data set		Confusion matrices				Results			Accuracy %
		Pre-test control	Post-test control	Pre-test exp	Post-test exp	Precision %	Recall %	F-measure %	
Class A	Pre-test control	114	30	5	20	86.7	88.5	92.3	89.4
	Post-test control	31	120	34	23	87.3	86.3	87.2	
	Pre-test exp	22	12	200	6	89	90	85.3	
	Post-test exp	34	40	9	300	90.2	91.1	89.3	

Table 3 Algorithm of MMDBM for confusion matrices, precision, recall, and accuracy

Data set		Confusion matrices				Results			Accuracy %
		Pre-test control	Post-test control	Pre-test exp	Post-test exp	Precision %	Recall %	F-Measure %	
Class A	Pre-test control	220	28	18	7	88.7	85.5	86.5	90.2%
	Post-test control	30	200	12	5	89.2	90.1	88.4	
	Pre-test exp	27	8	202	6	88.1	92.1	88.1	
	Post-test exp	49	7	9	172	93.2	96.6	95	

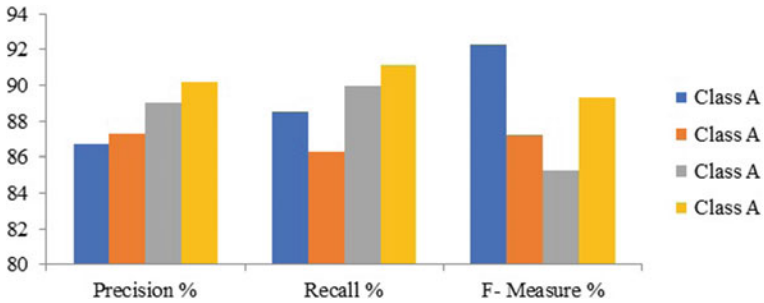


Fig. 2 Decision tree utilized for pre-test and post-test using decision tree and equalled with precision, recall, F-Measure and accuracy

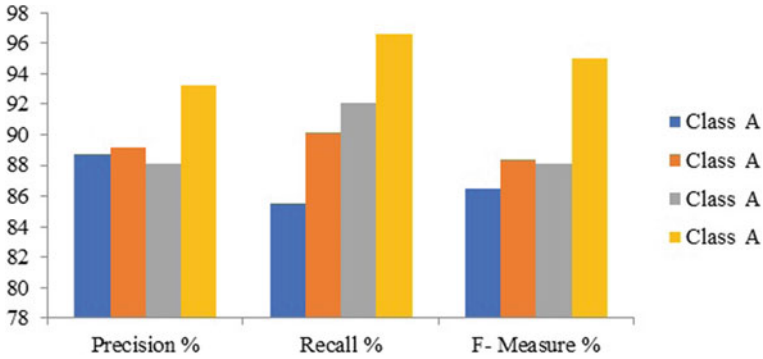


Fig. 3 Algorithm of MMDBM utilizing pre-test and post-test using MMDBM algorithm and equalled with precision, recall, F-Measure and accuracy

5 Conclusion

Algorithm of MMDBM and algorithm of decision tree obtains the test of pre and post outcomes in group of experiments and control of group are compared and discussed in the final outcomes. “Recall, precision, F-Measure, and Accuracy” are calculated and generated for the two algorithms. Algorithm of DT for method of classification in which the MMDBM was proposed. Based on the pre-test, post-test control with comparative analysis, the MMDBM model obtained accuracy 90.2% higher than the DT algorithm.

References

1. Palai G, Nayak B, Sahoo SK, Nayak SR, Tripathy SK (2018) Metamaterial based photonic crystal ber memory for optical computer. *Int J Light Electron Opt* 171:393–396
2. Elley WB, Mangubhai F (1983) The impact of reading on second language learning. *Read Res Q* 19:53–67
3. Sivakumar S, Muthukumaran M (2020) A performance study of image quality attributes on smoothed image obtained by anisotropic diffusion based models. In: *Examining fractal image processing and analysis*, book chapter-5, pp 100–120
4. Sivakumar S, Musthafa Baig M, Nayak SR (2020) Optimizing performance of text searching using CPU and GPUs. *Progr Comput Anal Network Adv Intell Syst Comput Book Ser* 1119:141–150
5. Nayak Soumya Ranjan, Jena Pyari Mohan, Mishra Jibitesh (2018) Fractal analysis of image sets using differential box counting techniques. *Int J Inf Technol* 10(1):39–47
6. Sivakumar S, Zeelan Basha CMAK, Haritha D (2018) Classification of mining techniques in multiclass dataset using wavelets. *Int J Pure Appl Math* 118(10):217–222
7. Rajesh Kumar E, Rama Rao KVS N, Nayak Soumya Ranjan, Chandra Ramesh (2020) Suicidal ideation prediction in twitter data using machine learning techniques. *J Interdisc Math* 23(1):117–125

8. Rajesh Kumar E, Rama Rao KVSN (2020) Suicide prediction in twitter data using mining techniques: a survey. In: International conference on intelligent sustainable systems, pp 122–131
9. Pradeepini G, Pradeepa G, Tejanagasri B, Gorrepati SH (2018) Data classification and personal care management system by machine learning approach. *Int J Eng Technol (UAE)* 7(32):219–223
10. Sivakumar S, Ganesan P, Sundar S (2017) A MMDBM classifier with CPU and CUDA GPU computing in various sorting procedures. *Int Arab J Inf Technol* 14(7):897–906
11. Sivakumar S, Anjali Devi S (2020) A hybrid document features extraction with clustering based classification framework on large document sets. *Int J Adv Comput Sci Appl (IJACSA)* 11(7):364–374
12. Sivakumar S, Nayak Soumya Ranjan, Kumar Ashok, Vidyanandini S (2018) An empirical study of supervised learning methods for breast cancer diseases. *Optik* 175:105–114
13. Sreedevi E, Prasanth Y (2017) A novel class balance ensemble classification model for application and object oriented defect database. *J Adv Res Dyn Control Syst* 9:702–726
14. Sreedevi E, PremaLatha V, Sivakumar S (2019) A comparative study on new classification algorithm using NASA MDP datasets for software defect detection. In: 2nd international conference on intelligent sustainable systems (ICISS 2019), pp 312–317
15. Yalla Prasanth, Mandhala Venkata Naresh, Abhishiktha Valavala, Saisree Chitturi, Manogna Kandepi (2019) Machine learning techniques to predict defects by using testing parameters. *Int J Recent Technol Eng (IJRTE)* 8(4):7829–7834
16. Prasanth Y, Sreedevi E, Gayathri N, Rahul AS (2017) Analysis and implementation of ensemble feature selection to improve accuracy of software defect detection model. *J Adv Res Dyn Control Syst* 9(18):601–613

Reliable Biometric Authentication with Privacy Protection



Shilpi Barman Sharma, Ishika Dhall, Soumya Ranjan Nayak,
and Pushpita Chatterjee

Abstract *Objectives* To study and compare different types of biometric systems and to obtain the result for the same, this paper has discussed the need of biometrics, ways of performing it and the best type of biometrics that should be implemented. *Methods and findings:* We used various statistics, graphical and tabular representations for showing data studied from different papers and reviewing them. Instances of infringements of right to privacy and protection of personal data are proliferating. Biometric systems are identity authentication systems for information assurance and protection via biological traits since they are unique. Technology is rapidly advancing in every field and choosing passwords are not enough. Biometric systems are the best approach to keep data safe, and it also makes the accessing of data simple. Everything has its own pros and cons so; we will be discussing the pros and cons of various biometric technologies in order to determine the best among them.

Keywords Iris recognition · Authentication · Enrolment · Recognition · Feature extraction · Template matching

1 Introduction

Data theft and breach of security are major threats that loom large in areas concerning data management. Technological advancement towards maintenance of stringent security is a prime requisite for any successful business. Entrepreneurs, collaborations and corporations resort to the latest upgrading in security measures for a venture to remain immune to intrusion. Data storage and its security cannot be compromised

S. B. Sharma · I. Dhall · S. R. Nayak (✉)
Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India
e-mail: nayak.soumya17@gmail.com

S. B. Sharma
e-mail: ssharma22@amity.edu

P. Chatterjee
Tennessee State University, Nashville, TN, USA
e-mail: pchatter@tnstate.edu

upon. In eventuality of any classified information leaking, there can be serious repercussions that will affect a business, adversely. Data leaking is a predecessor to business sabotage; cybercrime and can even pose a major threat to a nation's security. Since it is such a non-negotiable component for long sustaining success and security of an enterprise, we try to combat the vulnerabilities; however, we are still grappling at reaching 100% success. Every human being has unique anatomical characteristics that demarcate them from their contemporaries. Human body is astonishing; everything in our body is generated with different codes in terms of fingerprints, hand geometry and even unique iris pattern, to name the most prominent ones. These different features are decrypted as discussed in [1]. We can secure our information by making that decrypted code a password itself. 'BIOMETRIC' is a Greek term inspired from the words 'bio' (life) and 'metrics' (measure something) [1]. Biometric systems can be segregated in two different segments—enrolment and recognition.

Challenges those biometric systems face in maintaining security are inadequate authentication technology, unethical means of cracking through the biometrics, unforeseen aberrations like body injuries and age-related changes in physical attributes like nerve patterns and cataract.

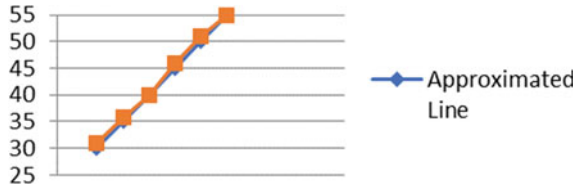
This paper will be majorly focusing on reviewing some of the best papers from the last 10 years that leveraged various techniques. The review will also be covering the functionality [2] of those techniques, their success rate [3], impediments, personal observations and conclusion.

2 Related Work

The future perspectives of biometrics and access control topologies have been discussed in reference [4]. The main concern shown here is related to situational awareness and risk management. The case study on facial biometrics, RGB and depth facial biometrics is explained. The head pose estimation was mentioned which further helped in proper face detection and recognition. The infrared thermograph diagnoses information regarding normal and abnormal working of nervous system. Infrared image processing, evaluation and infrared video imaging along with the heat conduction formulas at the skin's surface have been explained in [5]. The head pose algorithm defined in this paper basically filtered the image, and it is positioning and performs the recognition on just the front view instead of opting 3D modelling techniques. The research on facial recognition was specified using expression ($y = (0.2087 * x) + 22.28(1)$) taking x : The averaged grey-level intensity of bounding box and y : The estimated temperature. A graphical representation using vectors of temperature with respect to intensity was also given in Fig. 1.

The technologies that measure individual patterns (physiological and behavioural) are named as biometric-based technologies. We cannot extenuate the security issues and vulnerabilities that we face especially in this field, so [6] showed the way of how the unique identity of the user is spoofed and hence the countermeasure that should

Fig. 1 A graphical representation using vectors of temperature with respect to intensity



be taken. Also, it explains how we can access the concept of biometrics in our day-to-day life. Spoofing attacks are done using hardware- and software-based techniques. It explained about the good characteristics which a biometric method should have. Access control system and different paradigms under it are also explained briefly in [6] which tell about upbringing of biometrics in everything we are surrounded with, not just our smart phones and laptops but every appliance we are using so far. Whenever we are talking about data control and data access, the first and foremost thing which is required is data storage. So in the end, cloud computing and data storage is discussed. It also includes all the future perspectives but it certainly did not explain about how to implement the techniques and no algorithm or practical implication is done. In a nutshell, [6] is analysing about current advancements being made in access control systems, identifying their strong points so they can be enhanced and overall performance is improved.

Methods of biometrics and how iris recognition is better than all of them in a way have been discussed in [7]. It explained the basics, merits and demerits of all types of different recognition methods. A brief explanation about fingerprint scanning, DNA pattern, face recognition, voice recognition, palm or hand geometry, signatures and iris recognition is presented in [7]. It explained about how every type of biometric technique fails at some level. Fingerprint scanning is one of the most effective and cheap types of biometric techniques is comparatively easier to implement than others, so it is widely used. Iris recognition is a bit expensive one, but one of the most accurate type of recognition in biometrics system. A formula explaining how we can differentiate different people using DNA pattern recognition (Euclidean distance = $\sum i - \sqrt{(e - 1i - e2i) 2(1)}$) was explained but since it is a very tedious job, it is not used commonly. The aim is mostly to cluster the genes of similar levels of expression. Then it emphasized on iris recognition and explained a formula given by Prof. John Daugman. It is based on iris codes, and usually, the integral and differential operators are used to detect the centre and diameter of the iris ($\max(r, \times 0, y_0) || |G\sigma(r) * \partial \partial r \phi r, \times 0, y_0 I(x, y) 2\pi r ds| || (2)$). Iris not only differs between identical twins but also from left to right eye. So according to the author, iris recognition was the most reliable method of biometrics. Further, advantages and disadvantages of the iris recognition are explained. Iris databases are discussed in end too. In conclusion, [7] states and explains the need of iris recognition in future.

3 Methodology

Biometrics authentication is the concept of identifying oneself by something that you know, something that you have, or something that you are. In reference [8], a review is given on recent news and advancements in the field of biometrics. Then the different advantages and disadvantages were discussed in which advantages were explained very briefly but covered the main points. On the other hand in disadvantages, very apt reasons were given. News telling a good alternative of using radio frequency identification devices (RFID) for security purposes was explained which is expected to be used by US government. Further different approaches of protecting and managing the biometric authentication systems were explained like securing enrolment locations and keeping backups. Then different biometric authentications were compared on the basis of how strong the method is and how tough its spoofing can be. Iris recognition was considered and proofed to be the safest biometric authentication among many because spoofing is very tough in this due to the unique pattern of iris. But a test was performed by a German company for testing the strength of an iris recognition system, and the attack was successfully rendered. Voice recognition and palm vein pattern biometrics, the latest type of biometrics is also discussed briefly. Finally, future of biometrics is predicted by the author considering all the points of feasibility and security but since many countries are taking the help of biometric systems already, it is expected to be one of the major ways of securing data in future. So in conclusion large part of biometrics authentication future will be on dependent on user acceptance.

The size and pattern of iris vary from person to person, also pattern varies from eye to eye even if they belong to same person. After extracting the unique iris features, it is stored as template in database. These features are called as Iris code for template creation. Further in [9], the existing researches done in iris recognition have been discussed. Cascaded classifiers to increase accuracy were proposed in first reviewed paper in which a way of decreasing the noise in iris image is given. Some researchers found an iris blob matching algorithm represented by variants in order to solve it. Then after the noise, they considered environment as one of the reasons why the correct image is not captured. The concept of false rejection rate (FRR) and ways to reduce it was discussed using fusion algorithms. The third review explains about how to parallelize the iris recognition system. The sequential steps were divided into processes parallel to each other and run on field-programmable gate arrays (FPGAs). An analysis of iris code was done in next paper in which D. Daugman developed an algorithm about iris recognition called 'iris code' which was further improved. It described the experimental results of Face Recognition Vendor Test 2006 (FRVT) and Iris Challenge Evaluation 2006 (ICE). They found improvement in their work, and so they continued making progress. Step by step much advancements were made keeping all the circumstances under control like selecting optimum wavelength band for pigmented iris, occlusion estimation method to differentiate between noise and useful texture of iris, quality of biometric image, iris segmentation approach by using 'random walker algorithm', user-friendly and secure approach, and revocability

Table 1 Distance between input error and validation error [3]

ICA	PCA	Daugman	Hybrid
2.18E-10	3.26E-10	8.26E-07	9.69E-06
5.14E-07	5.14E-07	3.11E-07	8.93E-06
2.09E-10	2.09E-10	2.79E-10	9.02E-06

Note Distance between input error and validation error was calculated and compared to find the best method opted

management, etc. All of these topics are surveyed and in conclusion the authors wish on enhancing the performance of iris recognition system.

As discussed in [3], image of iris is captured and processed so that if we want to check someone for identification, we can compare the two. So the capturing and processing should be correct. In this paper, iris recognition is compared with radio frequency identification in order to check which system is more accurate and feasible. For that, both the systems were worked upon and observations and results were compared. In this experiment, they took the readings of three different images for each method (ICA, PCA, Daugman & Hybrid) and compared the accuracy. They calculated distance between input error and validation error for each image and the result showed that the sequence of performance (Descending) was as follows—ICA, PCA were nearly same, then hybrid and Daugman’s. Iris authentication is the most feasible by the use of three algorithms altogether (ICA, PCA, Daugman’s rubber sheet model) & RFID. Recorded observations and graphs that were plotted showed that the iris recognition was better in every way, and moreover, we all thought of Daugman’s theory to be the best one as of now but according to their experimentations, ICA turned out to be a better theory (Table 1).

Biometric authentication is one of the best ways of providing sufficient information entropy for good security. It discussed how using a combination of biometrics like gaze tracking, iris recognition or eye trackers can prevent the fake iris attacks. The very first technique which was used for authentication was Keystroke dynamics which was proposed by D. Spillane. The second technique in iris recognition was first proposed in 1936, but the procedure was patented in 1987 by D. Frank Burch. It was first implemented in 1990 when D. John Daugman created the algorithms for it. Then eye-tracking technologies were developed in the 1970s. They were much less intrusive for the user than using an iris recognition camera, so a methodology is shown on it in [10]. Eye tracker manufacturers usually use infrared ray cameras to capture images of iris but the eye tracker uses two infrared type LEDs which operate at an 880 nm wavelength to generate two bright-point reflections called the glint spots on the eye surface. Three parameters determine the iris recognition model: the Hamming threshold, angular sampling rate and radial sampling rate for the iris image. Different people were observed, and according to the results, conclusion was given in favour of commercial eye tracker, as it provides good performance for authentication. It has resolution 2X-5X lower than that of normal iris recognition system; moreover, users can move freely while using device, which raises considerable problems in quality of the iris image.

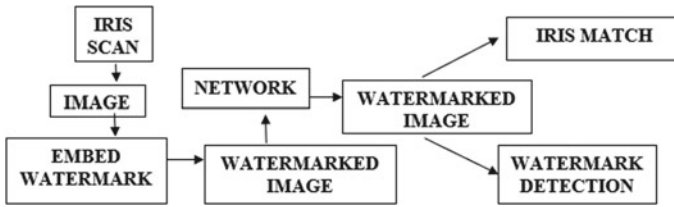


Fig. 2 Flow chart representing process of biometric system [10]

A watermark is a form, text or image that is impressed into a paper for providing an evidence of its authenticity or copyright. A digital watermarking is a pattern of bits inserted into digital images, videos and audio files which do the same job of providing authenticity as a normal watermark does. In reference [11], biometric watermarking has been explained. The impact of watermarking on the recognition systems of biometrics has been investigated mostly in the context of iris recognition only. Many different application scenarios are explained for robust water makings in biometrics like complementing live detection in biometric systems, an additional security token, feature extraction in image processing, embedding technique, enabling a multi-biometric approach, etc. Experimentally, first we take a scan of iris, watermark its raw image and then transfer it to the network then at receiver side, the watermarked image is processed by iris feature extraction and then it is matched. However, it uses robust watermarking it is more secure. After this, various popular algorithms related to it are explained and graphs are given related to them. Experiments are done using Barni's algorithm, Dugad's algorithm, Chen and Xie approximations and many more. As it was expected, watermarking actually affects the performance of iris recognition. Iris recognition algorithms which were considered showed that the observed trends are almost identical (Fig. 2).

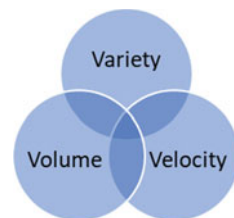
Any iris recognition method comprises of four stages, standardization, division, highlight extraction and characterization. As mentioned in [12], these stages are explained by giving examples and description. History and development of biometrics were given, telling the proposal by an ophthalmologist Frank Burch in 1936.

A variety of business frameworks are being created to improve eye images and implement verification methods. One kind of image-capturing technology was created in 1996 by a company called Sensor, and it uses a camera that can capture an image of a person's eye from a distance of up to one metre. Ticketless air travel, registrations and security system by iris acknowledgement booths are created by eye ticket in air terminals. Classification of biometrics is done on the basis of physiological and behavioural qualities of a person. Physiological qualities include the physical element like a unique identification, an iris design, the texture of retina or facial elements. Behavioural qualities are a recognizable proof of Man's mark, console lighting design or discourse design like a signature, speech or keystroke. Different physiological qualities were explained as follows: fingerprint recognition is usually

used as fingerprints are one of a kind for each finger of a person even for the indistinguishable twins, face recognition shows key elements from the focal bit of a facial picture, iris recognition is also at a boom since iris designs are extraordinary and the iris designs are received through video-based picture finding framework, hand and finger geometry. Behavioural qualities include speech recognition, signature verification, etc. There is a computerized iris recognition option for identification. The authors of this research wanted to find out about the potential of adventure biometrics. The objective of the research is to create advanced algorithms that will help to create better biometric security systems. The strategy they used was to get pictures from the database, convert them to dark scale, cut the iris and smooth out the histogram. Normalization, utilizing vigilant edge indicator, vertical and flat projection and hence finally coordinating. In conclusion, they have explained about how further they will continue it by developing a graphical user interface framework for the IRIS acknowledgement in MATLAB. It will help to identify the appropriate calculations that are needed.

Big data means extremely large datasets which are analysed computationally to unveil new patterns, and trends, especially those which are related to human interactions. Problems that are mostly faced by big data include privacy, accessibility and reliability, preventing node omission from clients during node breakdown and duplicating accounts. According to Lai et al. [13], big data have three components: variety, velocity and volume. Fingerprints, palm print, face recognition, iris recognition are the main methods which are briefly explained, and further, the recent innovations that were done were explained like airport and borders. Control Biometrics. It is getting popular and will soon be implemented in many countries since it reduces human efforts to a bigger extent. The next innovation consisted of consumer biometrics which has its main concern in mobility and connectivity. Biometrics has its importance in banking sector as well. Economic transactions and banking are very unsafe, so protection is also a big concern. In medical field also, biometrics brings protection and convenience. Whenever we are talking about data protection, we are obviously talking about cybercrimes too. So to prevent cybercrimes, biometrics is a good way of securing data. Finally, attendance system takes use of biometrics too and so it is getting implemented nowadays in corporate world. In conclusion, the authors want to convey that for privacy and security biometrics is a good approach and it has a wide scope. Biometrics is taking over traditional ways because it prevents frauds and improves security (Fig. 3).

Fig. 3 Venn diagram representing big data



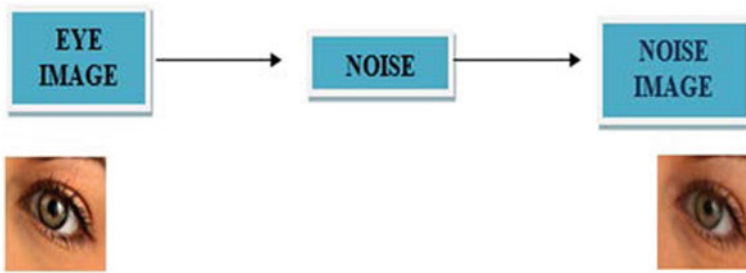


Fig. 4 Eye image without noise. Eye image with noise

Reference [14] explains the effect of noise in eye images by comparing different observations and tests that have been performed earlier. A detailed analysis of how well an iris recognition system performs with different levels of noise was performed. (0.01 , 0.05 , 0.09 or 0.1) applied to the images present in the database. Two broad terms that define the type of analysis that have been observed are: global analysis and local analysis (Fig. 4).

In [14], the reliability of iris recognition is analysed. The quality of images is checked by characterizing motion blur, defocus and occlusion problems. Chen et al., as mentioned in [15] measured the concentric energy bands of the iris, which they achieved using 2D wavelets to ensure the quality of the iris image. Ma et al., as mentioned in [16] analysed the Fourier spectrum of the local iris regions. The study by Salganicaff and Zhang given in [16] analysed the sharpness of the region found between the pupil and the iris. Kang [17], Park [18] and Daugman [19] quantified how many high spatial frequencies were in the iris region of the eye images using four traditional segmentation methods including Daugman's method. The methods like converting image into binary age map, edge point vote, Kovesis edge detector (mask's method), and Liam and Chekima method (darkness of iris, pupil and sclera being compared) were used. In iris recognition, the noise-free region is first extracted and then compared with the information present in database for accurate recognition. Peihua and Hongweis [20] gave a research. It has been found that one image can be split into several smaller sub-images to help fix various problems. Research [14] mainly utilized CASIA-IrisV3-Interval (as it was readily available) for generating the receiver operating characteristic (ROC) curve for the database. A comparison between ROC curve containing noise and the one not containing noise was done. During the process, the main problem that they faced was the mask segmentation problem. After the process, the iris was compared to see how it looked like using the Hamming distant (HD) method. The results obtained from this comparison by TPR and FPR values were used to construct the ROC curves based on the following two equations:

$$\text{TPR} = \text{TP}/P = \text{TP}/(\text{TP} + \text{FN}) \quad (1)$$

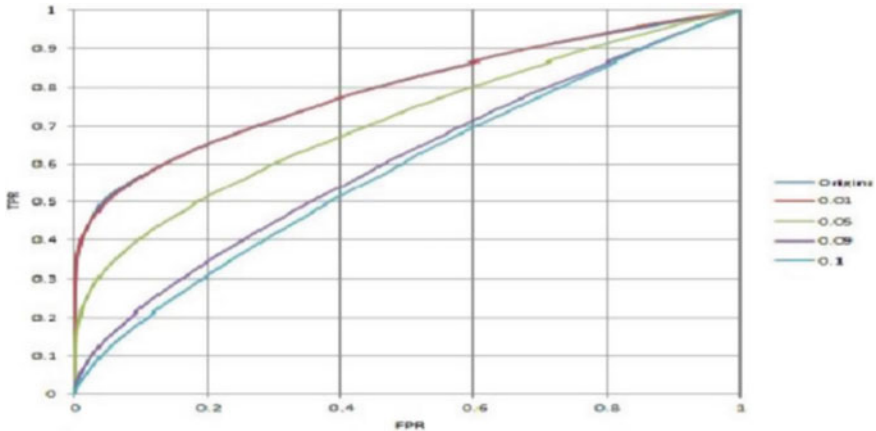


Fig. 5 Graph between TPR and FPR [14]

$$FPR = FP/N = FP/(FP + TN) \tag{2}$$

The study finally said that when the level of noise in an image was high, the image quality would be low. They also found that different ROC curves can show different noise levels that are affected by the quality of the iris image. But even if there is noise, it is possible to obtain an easy-to-recognize iris image, as shown in the figure showing the variation of TPR and FPR using CASIA dataset. Therefore, comparatively to other biometric systems, iris is the most reliable one (Fig. 5).

The concept of iris recognition was first proposed by D. Frank Burch in 1939. It was first implemented in 1990 when D. John Daugman created the algorithms for it. Iris is a part of human eye which is a thin and circular structure. Its main function is to control the size and diameter of the pupil. The colour of our eyes is also determined by iris only. Iris pattern of different people varies and no two people can have same iris pattern, which is one of the major reasons of using iris as a biometric tool of recognition (Fig. 6).

The different features of iris are explained in [2] like image acquisition, iris localization, feature extraction and template matching. Image acquisition is required to capture the images of iris for recognition. It concentrates upon capturing the correct portion of iris with sufficient resolution and sharpness. This is usually done using near-infrared cameras or LEDs. Iris localization is used to detach the iris from rest of the image captured. This is done using two circles: one of them dividing the boundary of iris from sclera and other from pupil. Features that make iris of different eyes different from each other are extracted and then converted into a code acceptable by the computer in terms of '0' and '1'. Recognition, identification and verification are three important terms that are followed in all biometrics. So, for verification part, template matching is done in which extracted part of iris is verified by comparing the different templates. CASIA database is described in the next section of [2] in brief.

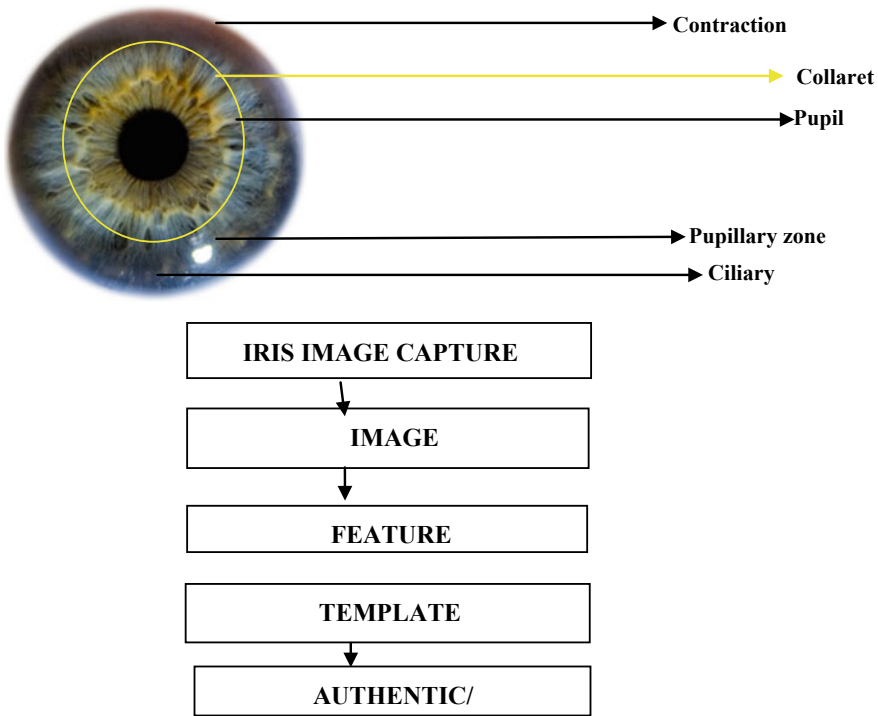


Fig. 6 Block diagram of stages in iris recognition [2]

In Sect. 5 of [2], an algorithm given by D. Daugman has been explained. Implementation of two processes is done: histogram equalization and binarization in which contrast between each eye’s regions is increased and the distance between iris region and other regions is maximized.

An integral–differential equation was used for finding the contour of iris and pupil given by:

$$\max(r, xp,)|G\sigma r * \partial/\partial r|(r, x_0, y_0) * (I(x,)/2\pi r)ds \tag{3}$$

where x_0, y_0 and r : the centre and radius of coarse circle (for each of pupil and iris). $G*\sigma(r)$: Gaussian function which is a smoothing function. $\Delta(r)$: the radius range. $I(X, Y)$: the original iris image. After giving this theorem, an optimized version of the same was released in which the miscalculations were considered and altered. Ultimately, the result shows that theory given by D. Daugman was proved to be a successful one, as it made iris recognition much easier. All the methods and equations proposed were checked and afterwards used worldwide. It was practised in Unique Identification Authority of India (UIDAI) for registering all 1.2 billion citizens of India, border crossing controls in United Arab Emirates, etc.

According to recent report in US Congress, approximately 2% of population does not have legible fingerprints and their enrolment is not possible for fingerprint biometric systems making it one of the demerits of this technique. A concept of dual biometrics was introduced and explained in [21], with a layered approach. Multimodal biometrics is basically a fusion of various biometrics indicators for recognition. It is usually used for big-scale applications like crossing borders. One of the best features of this type of technique is that it is a blend of various biometric modals; hence, it is much accurate and less vulnerable. This type of multimodal is limited to small population (hundred people at most), so commercial off-the-shelf (COTS) biometric is a better option. In [21], the experiments related to COTS are shown. Following section explained about the three methods of normalization and one new method called as adaptive normalization.

- **Min–Max method:** It maps raw scores to range of [0,1], (min(S) and max(S) being the end points) is given as:

$$n = \{S - \min(S)\} / \{\max(S) - \min(S)\}.$$
- **Z-Score method:** It transforms the score to distribution with mean = 0 and standard deviation = 1.

$$n = \{S - \text{mean}()\} / \text{std}()$$
- **Tan-h method:** This method is robustly statistical technique. This technique also maps the raw scores to range of (0, 1)

$$n = 1/2[\tan(0.01(s - \text{mean}(s)) / \text{std}(s)) + 1]$$
- **Adaptive method:** This method aims to increase the separation of the genuine and impostor distributions, keeping the range of mapping the scores in [0, 1].

$$nAD = f(nMM)$$

Function $f()$ has two parameters of overlapping regions: c and w . There are three different functions explained which can be used for adaptive method:

- **Two Quadratics (QQ):** $nAD = \{1/c (nMM^2), c + \text{root} [(1 - c)(nMM - c)]\}$
- **Logistic (LG):** $nAD = 1/[1 + A.e^{(-B.nMM)}]$
- **Quadric-Line-Quadric (QLQ):** $nAD = \{(1/(c - w/2))*nMM^2, nMM, (c + w/2) + \text{root}[(1 - c - w/2)(nMM - c - w/2)]\}$

There are five fusion methods: Simple-Sum, Min-Score, Max-Score, Matcher-Waiting, and User-Waiting. All normalization and fusion methods combining one by one were compared and the following results were extracted (Table 2):

Table 2 Comparison between normalization and fusion methods [21]

Normalization method	Fusion method				
	SS	MIS	MAS	MW	UW
MM	0.99	5.43	0.86	1.16	*0.63
ZS	*1.71	5.28	1.79	1.72	1.86
TH	1.73	4.65	1.82	*1.50	1.62
QLQ	0.94	5.43	*0.63	1.16	*0.63

Afterwards, normalization and fusion were performed, and finally, it was concluded that if specific characteristics are not known in advance (open populations, e.g. airports), Min–Max normalization and Simple-Sum fusion methods are better to be employed but for applications that deal with closed populations (e.g. an office environment), where repeated user samples and their statistics can be accumulated, the proposed UW and QLQ adaptive normalization methods can be used.

In [22], biometric identification was explained, telling how advanced biometric systems are and why it is used in all fields these days. Identity theft and fraud in disbursements, credit card transactions and ATM withdrawals take over \$6 billion every year. Considering these facts, identity security has become a matter of concern. So, all the conglomerates and companies tend to use automated identity authentication systems for customer satisfaction, efficient operations and better resource management (mostly critical). Before biometrics, two traditional ways were mainly used: knowledge-based and token-based (Fig. 7).

Due to many disadvantages, knowledge-based and token-based approaches are lying behind to differentiate between an authorized user and an impostor (who fraudulently acquired the token or knowledge of the authorized person (Fig. 8).

Various biometric techniques including fingerprints, hand geometry, facial thermogram, iris, retina, voice print, signature, etc., are explained, and a comparison is shown on the basis of their performance as follows (Fig. 9):

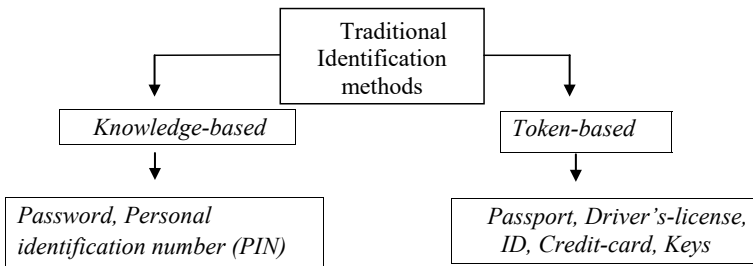


Fig. 7 Representation of traditional identification method

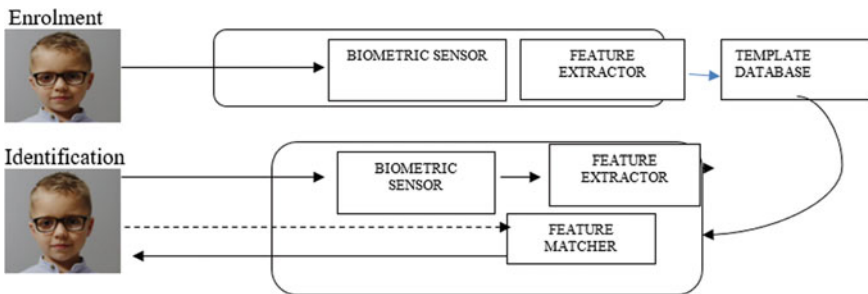


Fig. 8 Generic biometric system

Biometrics	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	high	low	medium	high	low	high	low
Fingerprint	medium	high	high	medium	high	medium	high
Hand Geometry	medium	medium	medium	high	medium	medium	medium
Iris	high	high	high	medium	high	low	high
Retinal Scan	high	high	medium	low	high	low	high
Signature	low	low	low	high	low	high	low
Voice Print	medium	low	low	medium	low	high	low
F.Thermogram	high	high	low	high	medium	high	high

Fig. 9 Comparison of biometrics on the basis of their performance [22]

Biometrics proved to be the best alternative for satisfying all types of requirements in authentication field, so it is used worldwide now and in conclusion biometric is likely to be used in every transaction which needs authentication of personal identity.

4 Observations

In reviews above, mostly the basic reasons and way of performing biometric systems have been explained. Some of the authors discussed the pros and cons of different techniques, and hence, telling the best type of technique that can be used for efficient authorization [22]. The basic reason of switching to biometrics is its efficiency and security. Different theories given by professors and researchers were distinguished by comparing input and validation errors, and it was found that ICA was the best hypothesis [3]. Comparison was done between the features and functionalities of different techniques, and the results were found [9]. Among many biometric systems, iris recognition was proved to be the best technique. In a nutshell, from different experiments we observed that biometrics is the best identification method and iris recognition is optimal technique of performing biometric identification and is used in many developing countries like Mexico and India but its set-up is extortionate.

5 Results

On the basis of the above reviews, we can come to a result that biometric systems are considered to be the most secure and convenient authentication tools for identification. It cannot be stolen, forgotten, borrowed or misplaced and forging one is also practically impossible. Figure 10 presents the percentage of biometric systems in the market and also shows the major characteristics used in biometric authentication of identity:

Also, iris recognition is considered to be the best way of recognizing since it is secure and accurate, and better than other recognition techniques overall. It is not

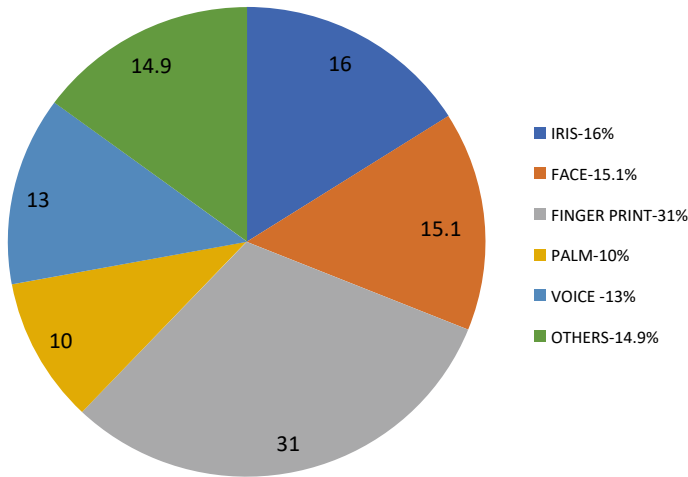


Fig. 10 Biometric modality

accepted much in the market yet but will soon be worldwidely accepted. According to Table 3, iris recognition is highly accurate, easy and safe. It is a bit obtrusive and inconvenient method but yet if we compare the features altogether, then iris recognition is the best type of biometric, and it has a lot of scope in future.

6 Conclusion

Biometrics is an automated identity security system which we are learning in order to improvise and expand it. Earlier we used traditional methods which were accepted everywhere but it failed to secure our identity to a greater extent. It mostly included security systems which were token-based and knowledge-based like driver licence, passport, personal identification number and ID. According to facts, over \$6 billion fraud has been observed every year due to these traditional methods. Identity theft [23] is a major concern nowadays which is gaining attention all over the world. To overcome this problem, automated identification of a person on the basis of their physiological and behavioural characters is adopted. It is much secure, cheap and easier to implement. The working of biometric systems is similar to traditional methods; i.e. the enrolment is done to store information in template database, and afterwards while performing identification the same database is checked and information is retrieved. Among various biometric techniques, iris recognition as stated in [24] is accepted as the best technique everywhere because of the uniqueness of characteristics of iris and its secure functioning. As the world is discerning about biometric techniques being actually effective for the protection of identity and privacy, people prefer these techniques over others in almost every field requiring authentication of identity.

Table 3 Result: representing the comparison between different techniques of biometrics

	Strengths	Weakness	Threats	Opportunities	Accuracy	Ease	Acceptance	Single nodes cost
Face recognition	Fast and one of the least expensive	Spoofing possible	Reliability	General	Low	High	High	Low
Fingerprint recognition	Inexpensive, Easy to capture, Unique	Any kind of aberration can affect the reading	Matching accuracy validation	Databases	High	Medium	Low	Low
Palm scanner	Tiny storage	Slow	Privacy or confidentiality	Manufacturing	High	High	Medium	Moderate
Iris scanner	Very secure, cannot be fooled	Inconvenient and obtrusive	Lack of liveliness testing	Medicinal and nuclear services	High	Medium	Medium	High
Thermal image	Very difficult to fool	Expensive	Acceptance of users	Highly secured places	Average	Medium	Low	Very high
Voice print	Inexpensive	Slow and can easily be affected	Reliability	Remote database and banking	Medium	High	High	Low
Signature recognition	Inexpensive	Easily affected	Accuracy and reliability	Industries	Medium	High	Medium	Low
DNA	Extremely accurate and cannot be fooled	Extremely intrusive, expensive	Can be extremely intrusive	Forensics and healthcare	Very high	Low	Low	Extremely high

References

1. Alqahtani A (2016) Evaluation of the reliability of iris recognition biometric authentication systems. In: 2016 international conference on computational science and computational intelligence (CSCI). IEEE
2. Bilge L et al (2009) All your contacts are belong to us: automated identity theft attacks on social networks. In: Proceedings of the 18th international conference on World wide web. ACM
3. Boatwright M, Luo X (2007) What do we know about biometrics authentication? In: Proceedings of the 4th annual conference on Information security curriculum development. ACM
4. Kang BJ, Park KR (2005) A study on iris image restoration. In: Audio-and video-based biometric person authentication. Springer, pp 31–40
5. Chitte PP, Rana JG, Taware S (2012) Iris recognition based security system using RFID. In: Proceedings of the CUBE international information technology conference. ACM
6. Chowhan SS, Shinde GN (2008) Iris biometrics recognition application in security management. Image and signal processing, 2008. Congress on CISP'08, vol 1. IEEE
7. Duarte T et al (2016) Biometric access control systems: a review on technologies to improve their efficiency. In: 2016 IEEE international power electronics and motion control conference (PEMC). IEEE
8. Hämmerle-Uhl J, Raab K, Uhl A (2011) Robust watermarking in iris recognition: application scenarios and impact on recognition performance. ACM SIGAPP Appl Comput Rev 11(3):6–18
9. Fujimasa, Chinzei T, Saito I (2000) Converting far infrared image information to other physiological data. Eng Med Biol Mag 19:71–76
10. Jagadeesh N, Chandrasekhar M, Patil (2017) Conceptual view of the iris recognition systems in the biometric world using image processing techniques. In: 2017 international conference on computing methodologies and communication (ICCMC). IEEE
11. Anil J, Hong L, Sharath P (2000) Biometric identification. Commun ACM 43(2):90–98
12. Daugman J (2004) How iris recognition works. IEEE Trans Circ Syst Video Technol 14(1):21–30
13. Lai K, Steven S, Yanushkevich SN (2014) “Multi-spectral facial biometrics in access control. In: 2014 IEEE symposium on computational intelligence in biometrics and identity management (CIBIM). IEEE
14. Salganicoff M, Zhang GH (1999) Method of measuring the focus of close-up images of eyes. US Patent 5953440
15. Mock K et al (2012) Real-time continuous iris recognition for authentication using an eye tracker. In: Proceedings of the 2012 ACM conference on computer and communications security. ACM
16. Kalka ND, Zuo J, Schmid NA, Cukic B (2006) Image quality assessment for iris biometric
17. Li P, Ma H (2012) Iris recognition in nonideal imaging conditions. Pattern Recognit Lett 33(8):1012–1018
18. Sharma, Shilpi, and J. S. Sodhi. “Implementation of biometric techniques in social networking sites.” *International Journal of Security and Its applications* 8.6 (2014): 51–60.
19. Sheela SV, P. A. Vijaya. "Iris recognition methods-survey." *International Journal of Computer Applications* 3.5 (2010): 19Snelick, Robert, et al (2005) Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. IEEE Trans Pattern Anal Mach Intell 27(3):450–455
20. Solanke, Sarika B., and Ratnadeep R. Deshmukh. ““Biometrics—Iris recognition system” A study of promising approaches for secured authentication.” *Computing for Sustainable Global Development (INDIACom)*, 2016 3rd International Conference on. IEEE, 2016.
21. Verma P et al (2012) Daughman’s algorithm method for iris recognition—a biometric approach. Int J Emerg Technol Adv Eng 2(6):177–185

22. Chen Y, Dass SC, Jain AK (2005) Localized iris image quality using 2-d wavelets. In: *Advances in biometrics*. Springer, Berlin, pp 373–381
23. Yadav V, Monika V, Kaushik VD (2016) A biometric approach to secure big data. In: *2016 international conference on innovation and challenges in cyber security (ICICCS-INBUSH)*. IEEE

Time Series Analysis of Cryptocurrency: Factors and Its Prospective



Sahil Sejwal, Kartik Aggarwal, Soumya Ranjan Nayak,
and Joseph Bamidele Awotunde

Abstract In the recent time as well as the last decade cryptocurrency has been one of the most discussed topic among the researchers all around the world. Different economies across the globe have seen a lot of growth of cryptocurrency over the time and bitcoin especially has seen a growth of 1100% that is why time series analysis of cryptocurrency is of immense significance. Time series analysis can be referred to as the process of taking into consideration a sequence of different points which are observed over a specific time interval. A large number of people start investing into cryptocurrencies without having any knowledge or analyzing the cryptocurrency market because of the hype it has these days and suffer huge losses so designing a model which can predict accurately as to how different cryptocurrencies would behave on basis of previous record can be very helpful and it can help some people in making profit rather than suffering loss. This paper presents a comparative overview of different algorithms like RNN, Linear Regression, GARCH, and ARIMA which can be used for time series analysis and concludes as to which algorithm is best suitable for time series analysis by considering different parameters like RMSE, MAE, etc., Besides this, it also analyzes the different factors which affect the prices of cryptocurrency.

Keywords Bitcoin · Cryptocurrency · Time series · RMSE · ARIMA

1 Introduction

In the collection of data, time always plays a significant role. A process in which time plays an important role in the collection of data is time series analysis. It helps in studying the processes and how different variables move within it. Time series

S. Sejwal · K. Aggarwal · S. R. Nayak (✉)
Amity School of Engineering and Technology, Amity University, Noida, India
e-mail: nayak.soumya17@gmail.com

J. B. Awotunde
Department of Computer Science, University of Ilorin, Ilorin, Nigeria
e-mail: awotunde.jb@unilorin.edu.ng

analysis can be referred to as a particular way of finding out meaning of a pattern of points of data collected along with passage of time. In this process, the different points of data are recorded randomly during a specific period of time [1]. However, it cannot be considered as similar to only collecting data with time. The factor that makes time series analysis as different is that the analysis of data collected in this with time can show variations with time. In other words, we can say that it gives a lot of importance to the time because it also reflects how data adjusts with respect to points of data and the results that are finally obtained. It gives some extra details and a number of dependencies between the data which is crucial in predicting the prices of cryptocurrency [2]. A large number of data points are required for time series analysis, and this helps in ensuring the reliability as well as the consistency. A dataset which is extensive helps in ensuring that the noisy data is taken care of and the fact that the trends are not considered as outliers. Time series has a lot of applications, and primarily, it is used to predict data on basis of historical data. Time series also helps the organizations in understanding the factors behind the trends or patterns with the passage of time. With the different platforms available these days, the visualizations can be done a lot ahead than the line graphs. Mostly, the time series analysis comes into play when there is data which is not stationary which implies the data that constantly fluctuates with time. Finance and retail are the industries where such kind of data can be encountered frequently and that is why time series analysis comes into play a lot in these industries because relative prices of currencies are always changing. Time series also helps the meteorologists in predicting the weather on the basis of how weather has changed previously with time [4]. Time series analysis can be extremely useful also for predicting the prices of cryptocurrency on the basis of how it has behaved previously by taking into account the different factors that affect its prices.

Machine learning has been a trending topic in last 20 years. It has been used in many types of science and engineering for betterment of interpretation of huge datasets. Evolving a new machine learning algorithm specifically means building a model that can give approximately correct output with the inclusion of certain input. For instance, suppose these models are boxes, you give input from one end and get output from the other end but model does not actually work this easily. The mapping from input to output is way more complex than it seems. But the evolution of supercomputers and powerful computers, there has been some ease in the training of the computers through which they can identify hidden complex representation in dataset given by user. A cryptocurrency is a digital currency which is designed to work as medium of exchange of commodities whereas buyer can buy a share, i.e., a percentage of the currency. For example, price of one bitcoin is around 25 lacs INR but if we want, we can buy a percent of bitcoin. So, this flexibility is one of the reasons that why cryptocurrency is popular nowadays. Its predicting is one of the most difficult tasks. There are a lot of factors that affect the price of cryptocurrency and bitcoin such as its high demand among people, controversies created by influential personalities like Elon musk just lifted dogecoin up to 11,800% [5]. Its unpredictable and volatile nature is the main problem which has to be determined for its time series analysis and prediction.

The reminder of the paper is presented as follows: Sect. 2 presented the literature work, and the methodology section is deliberated in Sect. 3. Factors affecting cryptocurrency and algorithm analysis are presented in Sects. 4 and 5, respectively. The concluding remark and future scope are presented in Sect. 6.

2 Related Work

While considering the scenario of cryptocurrencies price prediction, the work in majority of cases takes into consideration the different factors that affect the price of cryptocurrency. A large number of papers have been published in which various factors have been determined. Ladislav initially considered the market of cryptocurrency which was made of no fundamentalists but only speculative leaders [6]. His work inferred the relation between cryptocurrency prices and the search trends. After some time, wavelet coherence analysis was used in order to determine the potential drivers of the prices of Bitcoin and it helped in revealing the specific properties of bitcoin such as speculative ones [7]. Furthermore, 66 coins were taken into consideration in order to get familiar with the reasons responsible for affecting the prices of bitcoin especially in the technical area like cryptologic algorithm which is used for protocol and the computational power as aggregate which is mostly used for mining the cryptocurrencies [8]. The comparative analytical study of literature is presented in Table 1.

The information gathering on getting an idea as to what the price of Cryptocurrency would be using machine learning algorithms has not gained much of importance even though they are pretty efficient. A latent source model was implemented to present cryptocurrency [13], furthermore same has been implemented in practical implication [14] and they observed a return of 89% in around 50 days in which 4.1 was obtained as the Sharpe ratio. Some work has also been done using text from different platforms of social media in order to predict the prices of bitcoin like sentiment analysis has been used for the same in which the Wikipedia reviews were also taken into consideration in which frequency was coupled using SVMs [15]. Liquidity is considered limited in case of exchanges of bitcoin, and this is the reason that market

Table 1 Comparative technique of previous papers

Reference	Technique	Outcome	Accuracy (%)
[9]	LSTM	Bitcoin price prediction	68
[10]	Linear Regression	Bitcoin price prediction	64
[11]	SVM, RNN, KNN	Bitcoin price prediction	74
[12]	ANFIS, SVM	Bitcoin price prediction	60

has higher chances of risk of manipulation. This is one of the reasons that sentiment analysis from social media was neglected by many researchers. A. Greaves alongside Au [16] took into consideration the blockchain in predicting the prices using SVM and ANN which stands for Artificial Neural Network. And the accuracy obtained using ANN was around 55% which led them to conclude that there was limitation in case of prediction using blockchain as well. Zhao with Madan and Saluja [17] also used the same blockchain technique but with a bit of optimization. Besides SVM, they also took into account the binomial GLM which is a generalized linear model and they got accuracy of around 97% but they did not cross-validated which led to generalizability of the results obtained. It is on the basis of this in case of different famous exchanges which are put forward by CoinDesk, hash rate and difficulty have been inculcated while analyzing.

While analyzing the different determinants of price of cryptocurrency, the framework of Barro [18] has also been used in which empirical analysis was conducted [19]. Some of these analysis proved wrong the fact that the macro-level financial developments play an important role in driving the prices of cryptocurrency. Sentiment as a factor responsible for influencing the volatility of these currencies was analyzed [20], and it was inferred that when the bitcoin was in the bubble stage the sentiments affected positively. Dynamic interactions between the values of cryptocurrency and the social media alongside literature based on finance were taken into consideration [21].

Multi-Layer Perceptron has also been implemented for the prediction of the prices of stock [22, 23] but the drawback of it is that only one observation is taken at a time [24]. However, in case of RNN the output of previous layer is stored so that it can be looped back from next layer. It is in this way that memory of all sorts is used in networks gains and that is how it is different and more important than MLP. The term used to describe the network length is window length [25]. Another form of RNN that can be used in time series analysis is LSTM which was initially implemented [26] for the purpose of time series analysis. However, one of the limitation of RNN and LSTM is that it requires a considerable amount of computational time while training.

Besides analyzing what all models are beneficial for time series analysis, the research in this field further increased when NVIDIA developed the CUDA framework in the year 2006 because of which parallel computational efficiency of GPUs increased especially in the area of machine learning. So, research has also been done on whether the model works faster when implemented on a GPU or CPU. This relationship was initially analyzed by Matta et al. [27], and when it was further observed by Catanzaro et al. [28], it was seen that the speed of classification increased by eight times when the SVM was implemented on GPU rather than SVM being implemented on CPU and similarly the training time improved by 9 times on GPU as compared to the training time on CPU. In case of deep neural networks, it was seen in research [29] that the training speed was 40 times better than training deep neural networks in CPU.

3 Methodology

In this paper, we are mainly focusing in comparing all the algorithms used for time series analysis of cryptocurrencies using machine learning. A time series is nothing but a method of analyzing or understanding a sequence of data checkpoints over a chosen interval of time. Time series analysis is predicting the future behavior of that particular data that user has collected. In this paper, we will compare that which algorithm gives the best accuracy for the time series analysis of cryptocurrencies for the chosen interval of time.

The historical data related to bitcoin which acted as our dataset was collected from an open-source Web site called Kaggle. Initially, during the pre-processing phase the unwanted data from the dataset was removed and some of the features were normalized and analysis of significant attributes were performed. The evaluation has been done on RNN, Linear regression, GARCH, and ARIMA. However, in the introduction part, we have given a brief introduction of what is cryptocurrency, how it works, and on what factors it depends upon. After this in the sub sequent sections, we have discussed about what work has been already done on this topic, i.e., what techniques and algorithms have been performed for the time series analysis based on the research offered on the aspect of bitcoin behaviors over time [30] as shown in Fig. 1.

Coming on to Sect. 4, in that we will discuss about that on what factors cryptocur-rencies depend upon and how its prices go upwards and downwards. The flowchart given below shows the basic methodology which was adopted while implementing all the algorithms. In Sect. 5, we follow the comparison of all existing algorithms

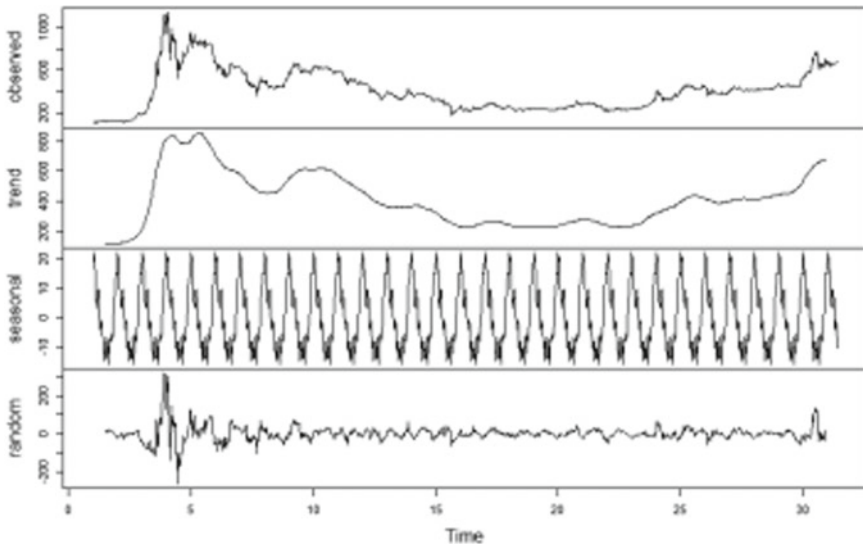


Fig. 1 Behavior of bitcoin over time [30]

which have been used in other papers and find out that which one is best, have most accuracy and least correction factor. Lastly, in Sect. 6 we have given the references of all the research papers we have referred. This paper is just a comparative research paper where our main focus is to find the best suitable algorithm for time series forecasting of cryptocurrency. During our research, we went through different papers and studied about various techniques used for analysis of future behavior of bitcoin and other cryptocurrencies, and presented general flow diagram followed by various researchers are presented in Fig. 2.

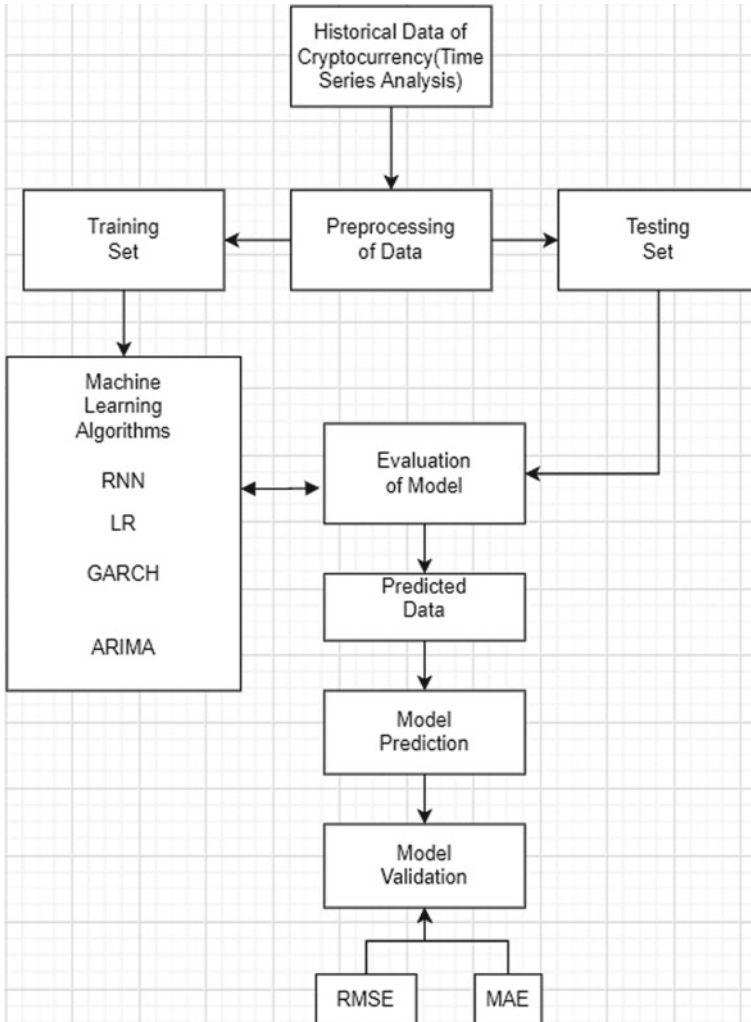


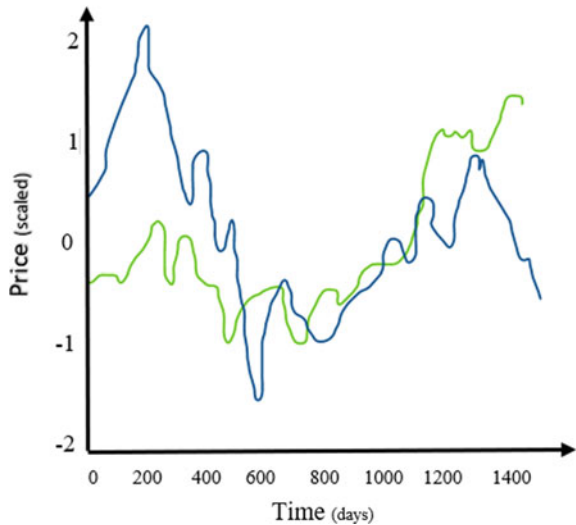
Fig. 2 Flowchart of basic methodology

4 Factors Affecting Cryptocurrency

There a large number of factors that play a crucial role in determining as to how the cryptocurrency prices would behave later and that is why it is immensely important to determine the different factors that determine the prices of crypto currency. A large number of analyses reflect that Bitcoin is the major reason and the trend which is followed by bitcoin is also followed by other cryptocurrencies because it was bitcoin which first implemented the Blockchain technology which is a decentralized database system and this was created by a mysterious group and the blockchain has the potential to hugely affect the business transformational model these days. As soon as the bitcoin starts gaining some friction, alternate coins are introduced in the market but Bitcoin is the main factor because it was the first cryptocurrency introduced in the market. The graph given in Fig. 3 shows how the prices of ETH follow the exact same path of BTC. So, there is a huge chance that when prices of bitcoin increase, the prices of other cryptocurrencies also increase and vice versa, and therefore, it can be one of the major factors which affects the prices of other coins. Here, x-axis represents the days and y-axis represents that how many times has the cryptocurrency gone up or gone down.

Another factor which comes into play in case of cryptocurrencies is the legal issues that revolve around the cryptocurrencies and technology game changers also play a role. One instance can be the statements by influential countries like when China banned the cryptocurrency it badly affected the prices of all the coins and it lead to devastation in crypto market. Another factor which is usually not noticed but affects the prices of crypto coins is the public perception. The more the coins are purchased, the higher is the demand and it leads to increase in the prices which can be taken into account by the help of web scraping through some Web sites. It

Fig. 3 BTC and ETH prices over a time



was recently observed that volatility in the prices of these coins is due to the market sentiment. As per a large number of studies, the “memory” of the time when there were deliberate changes in prices of coins acts as the semi-important factors that determine the prices of cryptocurrency. IT has been seen that long memory is there in individual market system as well which reflects the counteraction with the slow adjustment of shocks.

5 Algorithms Analysis

There are a number of algorithms available for predicting future behavior and analysis of time series of the bitcoin for particular interval of time. So, one of the first algorithm which is used for forecasting of cryptocurrency is recurrent neural network (RNN) with long short-term memory (LSTM).

5.1 RNN with LSTM

RNN is an algorithm which finds a sequence in the data for analysis. This works very well for sequential data such as speech, text, and time series.

In this, the information goes through a cycle where it makes decision and considers the initial input again and again in order to get better results [31]. The models have been evaluated on the basis of RMSE (Root Mean Square Error) and on MAE (Mean Absolute Error) presented in Fig. 4. The performance analysis for RNN is given as given in Table 2.

There are 3 gates in LSTM which are used to preserve and regulate information. The equation of these gates and cell is given below:

$$\text{Input Gate } In_t = \sigma(W_{in}[hs_t - 1], x_t + b_{in}) \tag{1}$$

Fig. 4 Functioning of RNN

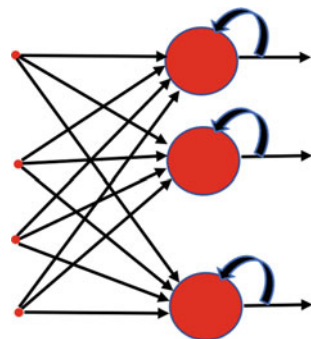


Table 2 Performance analysis of RNN

Model	Parameter	Score
RNN	RMSE	308
RNN	MAE	322

$$\text{Memory Cell } C_t = \tanh(W_c[hs_t - 1], x_t + b_e) \tag{2}$$

$$\text{Forget Gate } f_t = \sigma(W_f[hs_t - 1], x_t + b_f) \tag{3}$$

$$\text{Output Gate } f_o = \sigma(W_o[hs_t - 1], x_t + b_o) \tag{4}$$

Here, W is used for weight, x_t is input vector at time t , b is bias vector, and f , c , t , and o represent input, forget, cell memory, and output gates. LSTM is nothing but is an extension of RNN where it increases its memory for improved results. As far as we have researched, we have found that the accuracy of this model is only 52%, since it was the first technique introduced so it is a decent digit for accuracy. It was introduced in 1990s and was one of the first algorithm used for sequential data.

5.2 Linear Regression

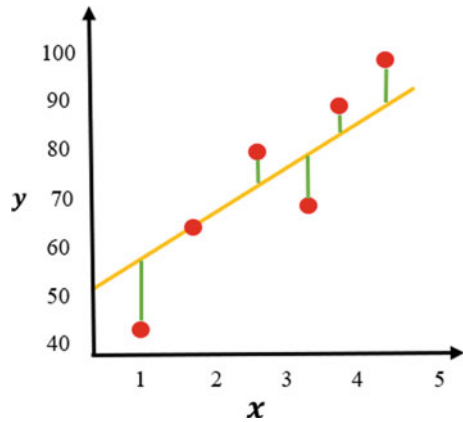
It is another method used for estimation of a certain model. In linear regression, the system identifies a relation between the input variable and the single output value which is linear in nature [32]. For estimation through this model, we require coefficients values such as mean, standard deviation, correlations, and covariance. Now, a relationship is formed between all these factors which represents the output. It has also showed the accuracy of around 77%.

The models have been evaluated on the basis of RMSE (Root Mean Square Error) and on MAE (Mean Absolute Error). The analysis for Linear Regression is given ahead. Here, red dot is the result of random deviations from a relationship which is represented in yellow between a dependent variable (y) and independent variable (x). To find the solution, we draw a line which passes to most of the points or closest to most of the points. For this, we have drawn a yellow line which is called as regression line presented in Table 3 and Fig. 5, respectively. The green line is the distance between the estimated value and the actual value. It represents the root mean square error.

Table 3 Performance analysis of LR

Model	Parameter	Score
LR	RMSE	320
LR	MAE	327

Fig. 5 Representation of linear regression model



5.3 Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

GARCH is a model which is statistical in nature whose function is to analyze time series data where the error of variance is auto correlated [33]. This model is one of the best model if the data is volatile in nature such as bitcoin which varies a lot in short interval of time. One of the main feature of GARCH [34] is that it can identify risks in advance and can calculate the expected returns which contains volatile clusters over a particular period of time. It is mainly used for identifying future behavior of financial data such as stocks, shares, and cryptocurrency. The models have been evaluated on the basis of RMSE (Root Mean Square Error) and on MAE (Mean Absolute Error). The analysis for GARCH is given as shown in Table 4.

The equation of a simple GARCH(1, 1) model is given below:

$$t = \omega + \alpha I X_{2t-1} + \dots + \alpha p X_{2t-p} + \beta 1 \alpha_{2t-1} + \dots + \beta q \sigma_{2t-q} \quad (5)$$

This model works when there is heteroscedasticity, i.e., variation of pattern is regular in the statistical data. The only reason that why this model lacks behind from other models is that it cannot work on the data which is constant. We know that bitcoin is volatile but if it is constant for a period of time then it would not be accurate. Its accuracy is computed around 50.8% but with a very low error rate.

Table 4 Performance analysis of GARCH

Model	Parameter	Score
GARCH	RMSE	298
GARCH	MAE	316

Table 5 Performance analysis of ARIMA

Model	Parameter	Score
ARIMA	RMSE	261
ARIMA	MAE	229.6

Table 6 Performance comparison of all algorithm

Model	Parameter	Score
RNN	RMSE	308
RNN	MAE	322
LR	RMSE	320
LR	MAE	327
GARCH	RMSE	261
GARCH	MAE	229.6
ARIMA	RMSE	298
ARIMA	MAE	316

5.4 ARIMA

It is one of the new model just like GARCH which is used for forecasting the behavior of cryptocurrency. This contains a lot of steps for data pre-processing which makes this model more complicated than any other model we have discussed till now. Other models perform better in beginning and then does not perform well in later steps and some perform better in later steps but not in beginning. But ARIMA is the model which performs better in almost all the steps which gives it upper hand on all the algorithms that we have discussed earlier in this paper. The models have been evaluated on the basis of RMSE (Root Mean Square Error) and on MAE (Mean Absolute Error). The analysis for ARIMA is given as shown in Table 5.

All these models can only predict the future behavior but no one can guarantee that what is exactly going to happen in the future. But after doing the comparative analysis of all the models, it can be said that ARIMA has the most accurate prediction as ARIMA has the least RMSE of 261 while RNN, LR, and GARCH have 308, 320, and 298, respectively, which is far higher than ARIMA, which is presented in Table 6.

6 Conclusion

A large number of people suffer huge losses because of their lack of knowledge and not able to take the correct decision as to which cryptocurrency would behave in the best possible manner. In this paper, we have investigated the different factors that play a major role in determining as to how the coin would behave in the future and

besides that we have analyzed the different algorithms that can be used to design the model that helps in predicting the prices of cryptocurrency and what is the accuracy and precision obtained by different algorithms.

It could be inferred that the GARCH algorithm which has been explained above was least precise and accurate then slightly better than that was RNN-LSTM but the ARIMA algorithm proved out to be the best for performing the time series analysis on cryptocurrency and it has a great benefit in case of linear models and has a precision of 100%. 60% was the usual accuracy which was obtained by the methods which are statistical which came out to be higher than the usual models of machine learning. The algorithm which performed even worse than GARCH model was XGB. We came across the ARIMAX model while researching about this topic. In ARIMAX, the forecasts are produced on basis of autoregressive terms and moving average terms. It takes into consideration the exogenous variables as well. So this can be achieved by making use of ACF and PACF which has various functionalities like it helps in determining the lag order of moving average. So we will try to predict bitcoin price using this model in our future work. This research paper which we have made has a certain number of limitations like data sources which brings us to our future work. As we move ahead, we will be finding how to take into consideration the different factors which affect the prices of cryptocurrency into our model and this paper helped us in realizing that we will be using ARIMAX model to predict the prices of cryptocurrency when we will implement the project.

References

1. Wei WW (2006) Time series analysis. In: The Oxford handbook of quantitative methods in psychology, vol 2
2. Rebane J, Karlsson I, Papapetrou P, Denic S (2018) Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: a comparative study. In: SIGKDD Fintech'18, London, UK, August 19–23, 2018
3. North H, Pairman D, Belliss SE, Cuff J (2012) Classifying agricultural land uses with time series of satellite images. In: 2012 IEEE International geoscience and remote sensing symposium, IEEE, pp 5693–5696
4. Nguyen T, de Bodisco C, Thaver R (2018) Factors affecting bitcoin price in the cryptocurrency market: an empirical study. *Int J Bus Econ Perspect* 13(1)
5. Kristoufek L (2013) BitCoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. *Sci Rep* 3(1):1–7
6. Kristoufek L (2015) What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE* 10(4):e0123923
7. Hayes A (2015) What factors give cryptocurrencies their value: an empirical analysis. Available at SSRN 2579445
8. Mittal A, Dhiman V, Singh A, Prakash C (2019) Short-term bitcoin price fluctuation prediction using social media and web search data. In: 2019 Twelfth International conference on contemporary computing, pp 1–6
9. Chen Z, Li C, Sun W (2020) Bitcoin price prediction using machine learning: an approach to sample dimension engineering. *J Comput Appl Math* 365:112395. <https://doi.org/10.1016/j.cam.2019.112395>

10. Mallqui DCA, Fernandes RAS (2018) Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques. *Appl Soft Comput J*. <https://doi.org/10.1016/j.asoc.2018.11.038>
11. Chen W, Xu H, Jia L, Gao Y (2020) Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *Int J Forecast*. <https://doi.org/10.1016/j.ijforecast.2020.02.008>
12. Shah D, Zhang K (2014) Bayesian regression and Bitcoin. In: 2014 52nd annual Allerton conference on communication, control, and computing. IEEE, Allerton, pp 409–414
13. Chen GH, Nikolov S, Shah D (2013) A latent source model for nonparametric time series classification. arXiv preprint [arXiv:1302.3639](https://arxiv.org/abs/1302.3639)
14. Georgoula I, Pournarakis D, Bilanakos C, Sotiropoulos D, Giaglis GM (2015) Using time-series and sentiment analysis to detect the determinants of bitcoin prices. Available at SSRN 2607167
15. Greaves A, Au B (2015) Using the bitcoin transaction graph to predict the price of bitcoin. *No Data*
16. Madan I, Saluja S, Zhao A (2015) Automated bitcoin trading via machine learning algorithms. <http://cs229.stanford.edu/proj2014/Isaac%20Madan20>
17. Barro RJ (1979) Money and the price level under the gold standard. *Econ J* 89(353):13–33
18. Ciaian P, Rajcaniova M, Kancs DA (2016) The economics of Bitcoin price formation. *Appl Econ* 48(19):1799–1815
19. Bukovina J, Marticek M (2016) Sentiment and bitcoin volatility. University of Brno
20. Mai F, Shan Z, Bai Q, Wang X, Chiang RH (2018) How does social media impact Bitcoin value? A test of the silent majority hypothesis. *J Manag Inf Syst* 35(1):19–52
21. White H (1988) Economic prediction using neural networks: the case of IBM daily stock returns. In: *ICNN*, vol 2, pp 451–458
22. Yoon Y, Swales G (1991) Predicting stock price performance: a neural network approach. In: *Proceedings of the twenty-fourth annual Hawaii international conference on system sciences*, vol 4. IEEE, pp 156–162
23. Koskela T, Lehtokangas M, Saarinen J, Kaski K (1996) Time series prediction with multilayer perceptron, FIR and Elman neural networks. In: *Proceedings of the World congress on neural networks*. INNS Press San Diego, USA, pp 491–496
24. Giles CL, Lawrence S, Tsoi AC (2001) Noisy time series prediction using recurrent neural networks and grammatical inference. *Mach Learn* 44(1):161–183
25. Gers FA, Eck D, Schmidhuber J (2002) Applying LSTM to time series predictable through time-window approaches. In: *Neural nets WIRN Vietri-01*. Springer, London, pp 193–200
26. Matta M, Lunesu I, Marchesi M (2015) Bitcoin spread prediction using social and web search media. In: *UMAP workshops*, pp 1–10
27. Catanzaro B, Sundaram N, Keutzer K (2008) Fast support vector machine training and classification on graphics processors. In: *Proceedings of the 25th international conference on machine learning*, pp 104–111
28. Cireşan DC, Meier U, Gambardella LM, Schmidhuber J (2010) Deep, big, simple neural nets for handwritten digit recognition. *Neural Comput* 22(12):3207–3220
29. McNally, S., Roche, J., & Caton, S. (2018, March). Predicting the price of bitcoin using machine learning. In 2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP) (pp. 339–343). IEEE.
30. Nelson, D. M., Pereira, A. C., & de Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. In 2017 International joint conference on neural networks (IJCNN) (pp. 1419–1426). IEEE.
31. Yan X, Su X (2009) *Linear regression analysis: theory and computing*. World Scientific
32. Troster V, Tiwari AK, Shahbaz M, Macedo DN (2019) Bitcoin returns and risk: a general GARCH and GAS analysis. *Finance Res Lett* 30:187–193
33. Gyamerah SA (2019) Modelling the volatility of Bitcoin returns using GARCH models. *Quant Finance Econ* 3:739–753
34. Malladi RK, Dheeriyaa PL (2021) Time series analysis of cryptocurrency returns and volatilities. *J Econ Finance* 45(1):75–94

A Review on Internet of Things: Communication Protocols, Wireless Technologies, and Applications



Meenu Garg, Gurjinder Kaur, Gurmehr Singh, Gursharan Sandhu, Sheifali Gupta, Soumya Ranjan Nayak, and Muhammad Fazal Ijaz

Abstract Internet of things (IoT) is a cloud-based “extensive global network” that connects numerous devices. Various devices are connected to the Internet for acquiring and storing the data. With the growth of technology, wireless networks, and affordable computer chips, everything can be integrated into the IoT. Wireless IoT can utilize a variety of different wireless communication technologies and protocols to connect various smart devices. Recent advancements in communication protocols and data processing permit the description of a new sort of restricted communication in terms of how objects communicate with one another on the Internet of things. This review paper presents a novel classification for the various traditional IoT network protocols and highlights different wireless IoT technologies along with their applications. In addition, comparative analysis of different features of IoT communication protocols is also presented.

M. Garg · G. Kaur · G. Singh · G. Sandhu · S. Gupta
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura,
Punjab, India
e-mail: meenu.garg@chitkara.edu.in

G. Kaur
e-mail: gurjinder.kaur@chitkara.edu.in

G. Singh
e-mail: gurmehr172118.ece@chitkara.edu.in

G. Sandhu
e-mail: gursharan172119.ece@chitkara.edu.in

S. Gupta
e-mail: sheifali.gupta@chitkara.edu.in

S. R. Nayak (✉)
Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India
e-mail: nayak.soumya17@gmail.com

M. F. Ijaz
Department of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, South
Korea
e-mail: fazal@sejong.ac.kr

Keywords Internet of things · IoT protocols · IoT wireless technologies · IoT applications

1 Introduction

The Internet of things (or IoT) is a combination of traditional fields such as embedded systems, control systems, and automation that permits the automatic transmission of data across a network without human interaction. Sensors and actuators embedded in physical things communicate via wired and wireless networks, which usually use the same Internet protocol (IP) that connects the Internet. After Kevin Ashton presented a completely new technology called “radio frequency identification” (or RFID) at Procter and Gamble, the term “Internet of things” (or IoT) was coined. The Internet of things (IoT) is based on the concept of ubiquitous communication [1]. IoT devices enable everyday things to communicate with humans and provide continuous support. The Internet and other devices can be accessed via wireless technologies such as Wi-Fi, Bluetooth, ZigBee, and RFID. The collected data will be processed, kept, and examined by the user. It has seen the IoT’s potential in the commercial and industrial development of prospering areas. However, data security is a major concern and a tremendous challenge [2]. The Internet has opened several channels for hackers, resulting in data vulnerabilities. However, IoT is committed to delivering the most effective solutions for dealing with this information and its security concerns. In addition, creating a secure method to communicate between social networks and privacy concerns is a big difficulty in IoT. There are numerous wireless technologies available today that operate in the roughly 2.4 GHz ISM band. Wireless devices are exploding in popularity. Thus, at a broad range of applications, the replacement of wires with wireless technology is also underway. For example, Bluetooth is being used to replace cable connections between desktop computers and external devices, while ZigBee is being used to monitor systems that require low energy consumption.

2 Literature Review

The Internet of things (or IoT), being an alarming technology and with the widespread anomalous research interest, presents qualitative work that copes with different factors of standardization in IoT. A lot of work has been done constantly to focus on abundant factors of standardization associated with IoT. Numerous Internet of things applications are already accessible, owing to the fact that a large number of individuals worldwide working in a related industry. Akyildiz et al. [3] made an exclusive IoT prototype, for the development of a bus transportation system in Singapore. With the help of this model, the customer can easily identify and estimate the available bus options for any particular destination, and it also gives an idea about the arrival timing

and crowd on the bus. Sharma et al. [4] proposed a unique IoT-based model, using virtual components for the replacement of the advertising system in large shopping centers and in many more organizations. A humidity controller is also a part of this model, which maintains the humidity inside the shopping malls and organizations without any human effort. Elkhodr et al. [5] highlighted the exceptional characteristics of wireless technologies and the problems with their integration into the IoT. Specifically, it concentrates on Bluetooth low energy, ZigBee, LoRa, and a number of distinct versions of Wi-Fi. Chen et al. [6] examined the efficiency of various IoT protocols in healthcare settings, including CoAP, DDS, MQTT, and UDP. The efficiency of the various protocols is evaluated using a network emulator. This emulator enabled the author to simulate a system with a high latency and a low bandwidth. The scientists compared the performance of the protocols and concluded that DDS's higher bandwidth consumption, high dependability, and data latency make it the optimal choice for healthcare applications. Al-Sarawi et al. [7] examined the issue of selecting the most appropriate technology for a certain application. Thus, a comparison of the general communication protocols used in IoT using diverse criteria has been conducted. Topology, cryptography, power consumption, standards, frequency ranges, data rate, features, security, and range are some of the criteria. According to the author by comparing the protocols, they can determine which one is feasible for a particular function. Alavi et al. [8] stated issues rose due to urbanization in the metropolis. The authors concluded that IoT-enabled technology is essential for smart city expansion to solve certain issues such as waste collection and traffic management.

This paper aims to provide an extensive review to explain all the predominant protocols and also provide newly arising standards protocols as shown in Fig. 1 The presented work exclusively gives an extensive overview of all the various IoT communications protocols, technologies, and their applications. The rest of the paper is structured as follows: The existing IoT protocols are summarized in Sect. 3, followed by IoT wireless technologies in Sects. 4 and 5 present the conclusion of the present work.

3 IoT Protocols

All IoT devices communicate using specific regulating protocols. These protocols are a set of rules that govern how data is transmitted to and from the Internet. They ensure that data from a device or sensor is read and interpreted by one device or sensor and not by another [9]. There are numerous procedures, all of which are briefly addressed in the paper below.

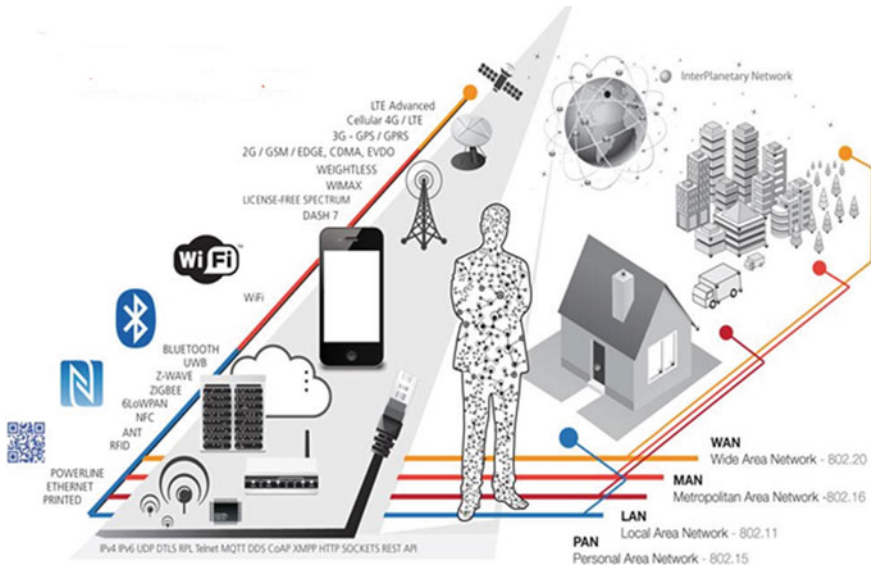


Fig. 1 IoT communication protocols and technologies

3.1 The Internet Protocol Version Four (IPv4)

The Internet protocol, or IP, is a unique numerical identifier assigned to each device connected to a larger computer network. It is used for a wide variety of functions and is further classified into two types: IPv4 and IPv6. IPv4 is the fourth generation of the Internet protocol, which was released in 1981 and is widely used in data communication [10]. It is a connectionless protocol that is implemented in packet switching layer networks. Typically, this 32-bit addressing mechanism is denoted by dot-decimal notations. 32-bit addressing is typically used in IPv4 in five classes, namely A, B, C, D, and E. The classes A to C use distinct bit lengths to define the network host. Pre-allocations of classes D and E have been made for military and future use, respectively. IPv4 can be configured with a wide variety of devices, either manually or automatically, depending on the network type.

3.2 Internet Protocol Version Six (IPv6)

IPv6 is the most recent version of the Internet protocol; it is essentially an upgrade to IPv4. It was created in 1998 by the Internet engineering task force (IETF) to address the long-predicted issue of IPv4 address exhaustion. IPv6 addresses are 128 bits in length and include both source and destination addresses [11]. Addresses are divided into eight groups by colons or hexadecimal characters. Along with increased address

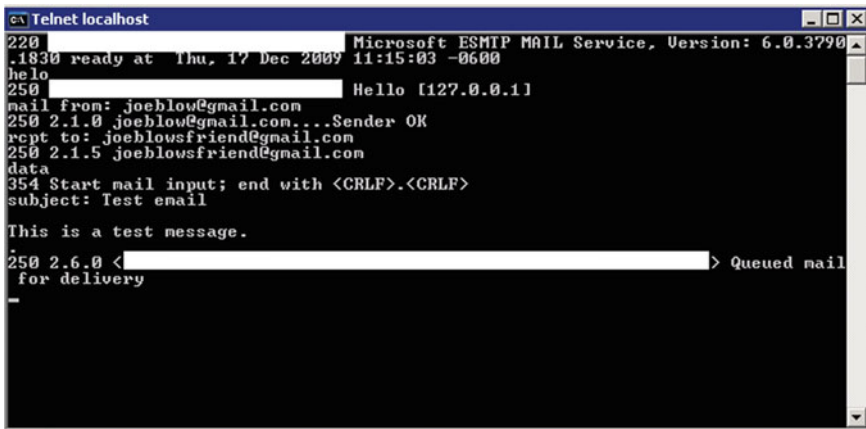
space, IPv6 supports hierarchical address assignment methods, which reduces the expansion of routing tables. IPv4 and IPv6 are both compatible and have no linkages.

3.3 User Datagram Protocol (UDP)

Designed by David P. Reed in 1980, UDP is known as one of the most essential members of the IP suite. It is an unreliable and simple connectionless protocol and offers a minimal transport facility to the applications that do not use the level of service of TCP [12]. The only service it gives is checksums for received data incorruptibility and multiplexing by port numbers to address different functions at the starting place and target of the datagram. UDP is mainly useful for concurrent services like gaming and voice/video communication. As great functioning is required in these services, UDP permits the packets to be dropped rather than handling the ones which are delayed. Also, quite a lot of bandwidth is saved as no error checking is done by UDP, thus making it very efficient in terms of both latency and bandwidth.

3.4 Teletype Network

Telnet is an application protocol that enables users to communicate in a bidirectional, eight-bit format. It was developed in 1969 and included a command-line interface (CLI) for communicating with remote hosts, as illustrated in Fig. 2. Telnet is an acronym for teletype network, and the verb “to telnet” may also be used [13]. The Telnet application allows users to run commands on the server, allowing users to



```

c:\ Telnet localhost
220 [redacted] Microsoft ESMTIP MAIL Service, Version: 6.0.3790
1.1830 ready at Thu, 17 Dec 2009 11:15:03 -0600
helo
250 [redacted] Hello [127.0.0.1]
mail from: joeblow@gmail.com
250 2.1.0 joeblow@gmail.com...Sender OK
rcpt to: joeblowsfriend@gmail.com
250 2.1.5 joeblowsfriend@gmail.com
data
354 Start mail input; end with <CRLF>.<CRLF>
subject: Test email

This is a test message.

250 2.6.0 <[redacted]> Queued mail
for delivery
-
```

Fig. 2 Telnet CLI

operate it and communicate with other servers on the network. Telnet is frequently used to provide remote access to multi-user dungeon (MUD) games and secure internal connections, as well as to test web and mail servers.

3.5 Datagram Transport Layer Security (DTLS)

DTLS is used to secure datagram-based communications. DTLS provides communication security for datagram protocols and is based on TLS with a few modifications to address issues caused by packet rearrangement or loss. DTLS does not guarantee message delivery in any particular order, or even that messages will be delivered at all. As the DTLS protocol uses UDP, it thus deals with packet reordering, datagram loss and data being pretty larger in size than that of a network packet [14]. It is used for web browsing, mail, instant messaging, and voice over IP (VoIP).

3.6 Routing Protocol for Low Power and Lossy Networks (RPL)

RPL is a routing protocol designed for wireless networks with minimal power consumption and a strong probability of packet loss. It is a demanding protocol based on distance vectors that operate on IEEE 802.15.4 and were designed for multi-hop and many-to-one communication as well as one-to-one messaging. This protocol can quickly construct network routes, disseminate routing knowledge, and efficiently adapt the topology.

3.7 Message Queuing Telemetry Transport (MQTT)

MQTT is a featherlight publish/subscribe messaging protocol developed for devices and networks that are strained due to excessive latency, low bandwidth, or instability. It was invented by Dr. Andy Stanford-Clark and Arlen Nipper in 1999. The MQTT design principles were developed in such a way that they minimize network bandwidth and device resource needs while ensuring soundness and some level of delivery assurance. Thus, MQTT is primarily designed for connections with a remote site that require a “minimal code footprint” or that have a limited network capacity.

3.8 *Data Distribution Service (DDS)*

DDS is a data communications model for machine-to-machine (M2M) communication. The DDS integrates system components, resulting in low-latency data transfer, high accuracy and an extendable architecture required for crucial IoT applications that require real-time data interchange. It completely removes the requirement for network programming that manages communications which automates their interaction.

3.9 *Constrained Application Protocol (CoAP)*

CoAP is a protocol for the service layer that is used with managed nodes and networks. These nodes are equipped with 8-bit microcontrollers and a tiny amount of read-only memory (ROM) and Random Access Memory (RAM). It implements a request/response mechanism for application interaction and includes built-in services and resources. The protocol is primarily targeted at hardware that cannot run HTTP, such as 8-bit microcontrollers, or TLS, as well as low-power sensors [15].

3.10 *Extensible Messaging and Presence Protocol (XMPP)*

The extensible messaging and presence protocol (XMPP) is a free and open set of technologies for real-time communication. This enables real-time communication and powers a wide variety of applications, including instant messaging. It collectively means:

- **X (Extensible):** This indicates that it can be customized or expanded to meet specific requirements.
- **M (Messaging):** This protocol is optimized for real-time messaging and features a very competent push mechanism in comparison with other protocols
- **P (Presence):** It indicates if you are online, offline, or occupied in another way.
- **P (Protocol):** A protocol is a collection of standards that enable systems to communicate with one another.
- This protocol is used to communicate with VoIP systems, publish–subscribe systems, file transfer systems, and gaming systems, among others.

3.11 *Sockets*

These are forms of communication that enable the flow of data both locally and across networks. Sockets are classified into four types:

- **Stream Sockets:** If you send three things “A, B, C” through a stream socket, they will arrive in the exact same order as they were sent. These sockets make use of TCP for data delivery. If delivery is not completed, the sender is notified of the failure.
- **Datagram Sockets:** These are connectionless sockets because they do not require an open network to function and instead use UDP.
- **Raw Sockets:** They enable the use of socket abstractions in basic communication protocols. They are useless to the general public and have been supplied solely to persons interested in inventing new protocols or acquiring access to multiple complex protocol facilities. They enable the user to edit a packet’s headers using the sequenced packet protocol (SPP) or the Internet datagram protocol (IDP), as well as to receive the headers on incoming packets.

3.12 Representational State Transfer (REST)

REST, as defined by Roy Fielding in 2000, is a collection of rules and regulations that specify how Web standards (e.g., HTTP and URIs) should be utilized. It is mostly famous for its simplicity and the concept upon which it is built-in order to accomplish its goals. RESTful Web services are services that adhere to the REST architectural style and enable Internet-based interoperability between computer systems. They allow the system to interact with textual representations of Web resources [9].

RESTful systems strive for stability, rapid performance, and extensibility through the reuse of components that may be controlled and updated separately from the rest of the system through the use of a stateless protocol and standard operations. To aid comprehension, Fig. 3 summarizes the REST-based messaging system.

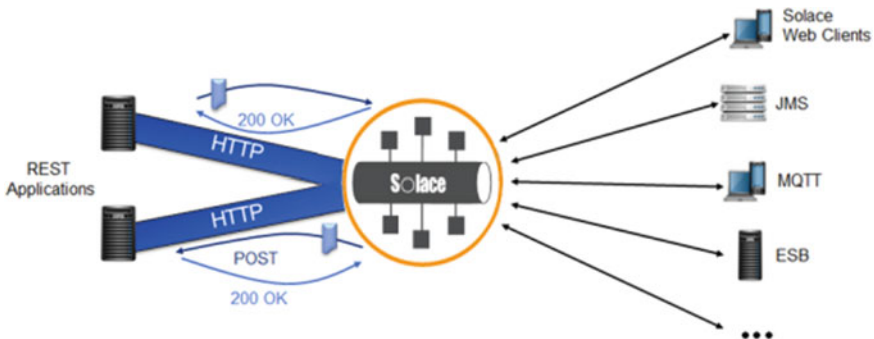


Fig. 3 REST messaging protocol

3.13 Application Programming Interface (API)

An API is a computing interface which influences interactions between numerous software mediators. It is basically a group of routines, protocols, and means for creating software applications. It defines the types of calls or requests that can be made, how we form them, the data formats that can be used, the conventions that must be followed, and so on. It also provides extension techniques so that users can extend current performances in numerous ways and to varying degrees [16].

The names of the OSI model layers, their data units, the TCP/IP model, function, and TCP/IP protocols for the particular layers are mentioned in Table 1

Table 1 Layers of model

OSI model layers	Data units	TCP/IP model	Function	TCP/IP protocols
Application layer	Data	Application	Produced the data, which has to be transferred over the network	CoAP, MQTT, XMPP
Presentation layer	Data	Application	Data extraction and manipulation as needed	FTP
Session layer	Data	Application	Established connections between two processes	API
Transport layer	Segments	Transport	Provide the acknowledgment after the complete data transmission, and re-transmit the data if an error occurs	DTLS, DDS, sockets
Network layer	Packets	Internet	Transfer data from one host to another across multiple networks	IPv4, IPv6, UDP, RPL, 6LOWPAN
Data link layer	Frames	Network access	Ensure that data is transmitted from one node to another without errors	Teletype network
Physical layer	Bits	Network access	Transmit bits from one node to another	Z-Wave, UWB, PLC

4 IoT Wireless Technologies and Applications

4.1 *Wi-Fi*

Wireless fidelity (Wi-Fi) belongs to the family of wireless networks, and it is based upon IEEE 802.11 standard, which are mostly used for LAN networking of devices and Internet access. Wi-Fi is a type of WLAN network. The only drawback of it is that it consumes a lot of power for some of the IoT applications. Its range is up to 50 m. Wi-Fi usually works at 2.4 and 5 GHz of frequencies.

4.2 *Bluetooth*

Bluetooth is commonly used for short-range wireless technology. It is mostly used for pairing and connecting smart watches or bands with our mobile phones. The latest Bluetooth technology used as an IoT protocol is Bluetooth low energy (BLE). BLE provides common Bluetooth range along with lesser power consumption. Bluetooth is based upon IEEE 802.15.1 standard but is now managed by Bluetooth special interest group (BSIG).

4.3 *Ultra-Wide Band*

Ultra-wide band, which is IEEE 802.15.3 standard, is used for wireless transmission of data. It has a very large bandwidth, and therefore, it is used for communications that require a very high bandwidth. UWG spreads information upto a large bandwidth of greater than 500 MHz. It has many disadvantages like higher achievable data rates and high immunity to interference because of its lower spectral power density.

4.4 *ZigBee*

It operates at a frequency of 2.5 GHz; it is mostly used by the industries. It is based upon IEEE 802.15.4 standard and is a low-cost device which consumes very less power. It is mainly focused on battery-powered devices in wireless and IoT applications. Its physical range is up to 10–20 m. The biggest advantage of ZigBee is that it also supports low data rates with the help of mesh networking protocol to avoid hubs in order to create a self-healing architecture.

4.5 *Z-wave*

It is mostly used for home automation, and it has a long range of up to 100 m. It uses low energy radio waves to enable communication between different devices, and its frequency range varies from 800 to 900 MHz; due to this specific frequency range Z-Wave becomes immunized to noise signals and interference which becomes its advantage over other wireless networks, and it also uses encryption techniques to transfer data between different appliances which makes it a secure wireless networking protocol.

4.6 *6LoWPAN*

It is an acronym for IPv6 low-power wireless personal area network. It is based on the IEEE 802.15.4 standard and is primarily focused on enabling IoT devices to communicate via IPv6 packets even when their power supply and processing power are constrained. It operates at a 2.4 GHz frequency. It is supported by a wide variety of networking technologies, including low-power radio frequency, Bluetooth smart, PLCs, and low-power Wi-Fi.

4.7 *Adaptive Network Topology (ANT)*

It is a protocol which is used for ultra-low-power wireless applications to send data between wireless devices in a robust and flexible way. Due to a large number of nodes present, this protocol can operate in a large number of network topologies ranging from peer to peer/star topology to practical mesh topology in personal area networks. ANT is a perfect protocol that can be used for LAN applications in homes and industrial automation applications.

The advantages of ANT protocol in networks are:

- The networks become operational even at very low power.
- The network becomes highly optimized and compact.
- The network becomes more flexible and scalable.

ANT devices can operate at any frequency range from 2400 to 2524 MHz, and operational frequency range of 2457 MHz is reserved for ANT+ devices.

4.8 Radio Frequency Identification (RFID)

In this technology, radio waves are used to capture and store information from an RFID [16] tag even if the tag is not present in the direct line of sight or it is present far away from the sensor. A sensor basically senses all the information from the RFID tag and sends the data collected to the interpreter which further sends the data to a computer. The latest standard of operation of RFID is EPC global class 1 gen—2 protocols. There are two kinds of RFID tags:

- Active tags: These tags are connected to some external power source to turn on their circuitry.
- Passive tags: These tags are very simple and cheap, and these are not connected to any external power source instead it uses reader signals to turn on its circuitry and advanced modulation techniques like QAM and PSK are also not possible in case of passive RFID tags.

4.9 Near-Field Communication (NFC)

NFC consists of protocols used to enable communication between devices placed very close to each other upto the distance of 2 m; these are also used as electronic documents and can be used to facilitate cashless payments using mobile phones or NFC cards.

NFC tags can operate at a frequency range of 13.56 MHz. Three modes of operation of an NFC device are:

- NFC emulation cards: It enables us to use our credit cards and our smart phones for cashless payments.
- NFC readers/writers: It can be used to obtain and store information from an NFC tag or any smart poster.
- NFC peer to peer: This allows two NFC devices to share and transmit data to each other by providing a low-speed connection between both the devices.

4.10 Powerline Ethernet Communication

It is used to provide Internet access in the house using the electrical wiring installed in the house to transfer AC current signals. It does not require any extra wiring to allow the transfer of data signals. Power Ethernet communication basically adds a modulating carrier signal to the existing electrical wiring system to facilitate the flow of data signals over the wiring systems. The limitations of powerline communication are the propagation problem because the power distribution system was developed for transmission of power at frequency range from 50 to 60 Hz. A comparison between the technology, frequency, data range, power usage, and cost is shown in Table 2.

Table 2 Comparative analysis of different technologies

Technology	Frequency	Data rate	Range (feet)	Power usage	Cost
Wi-Fi	Sub GHz, 2.4 GHz, and 5 GHz	0.1–54 Mbps	<300	Medium	Low
Bluetooth/BLE	2.4 GHz	1–3 Mbps	300	Low	Low
Ultra-wide band	1.3 GHz	675 Mbps	>1	Low	Low
ZigBee	2.5 GHz	250 kbps	300	Low	Medium
Z-wave	Sub GHz 800–900 MHz	100 kbps	100	Low	Medium
6LoWPAN	2.4 GHz	250 kbps	380	Low	Low
ANT	2457 MHz	60 kbps	>100	Low	Low
RIFD	2.45–5.8 GHz	640 kbps	6–90	Low	Very low
NFC	13.56 MHz	424 kbps	>1	Medium	Very low
Powerline Ethernet communication	50–60 Hz	400 Mbps	984	Low	Low

5 Conclusion

In this work, an exhaustive review of the IoT Protocols has been presented. Numerous similar protocols have been developed by organizations such as the Internet Engineering Task Force, the IEEE, and the ITU, and many more are being developed for future advancements. The primary objective of this study is to educate developers and service providers about the various IoT communication protocols and IoT wireless technologies. Different protocols along with their history, their basic mechanisms, and their applications have been discussed briefly for a better understanding.

References

1. Sheng Z, Yang S, Yu Y, Vasilakos AV, McCann JA, Leung KK (2013) A survey on the IETF protocol suite for the internet of things: standards, challenges, and opportunities. *IEEE Wirel Commun* 20(6):91–98
2. Granjal J, Monteiro E, Silva JS (2015) Security for the internet of things: a survey of existing protocols and open research issues. *IEEE Commun Surv Tutor* 17(3):1294–1312
3. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. *Comput Netw* 38(4):393–422
4. Sharma V, Tiwari R (2016) A review paper on “IOT” & it’s smart applications. *Int J Sci Eng Technol Res (IJSETR)* 5(2):472–476
5. Elkhodr M, Shahrestani S, Cheung H (2016) Emerging wireless technologies in the internet of things: a comparative study. *arXiv preprint [arXiv:1611.00861](https://arxiv.org/abs/1611.00861)*
6. Chen Y, Kunz T (2016) Performance evaluation of IoT protocols under a constrained wireless access network. In: 2016 international conference on selected topics in mobile & wireless networking (MoWNeT). IEEE, pp 1–7

7. Al-Sarawi S, Anbar M, Alieyan K, Alzubaidi M (2017) Internet of things (IoT) communication protocols. In: 2017 8th international conference on information technology (ICIT). IEEE, pp 685–690
8. Alavi AH, Jiao P, Buttlar WG, Lajnef N (2018) Internet of things-enabled smart cities: state-of-the-art and future trends. *Measurement* 129:589–606
9. Karagiannis V, Chatzimisios P, Vazquez-Gallego F, Alonso-Zarate J (2015) A survey on application layer protocols for the internet of things. *Trans IoT Cloud Comput* 3(1):11–17
10. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. *IEEE Commun Surv Tutor* 17(4):2347–2376
11. Partridge C, Pink S (1993) A faster UDP (user datagram protocol). *IEEE/ACM Trans Network* 1(4):429–440
12. Postel J (1983) Telnet protocol specification. In: RFC 854, ISI
13. Kothmayr T, Schmitt C, Hu W, Brünig M, Carle G (2012) A DTLS based end-to-end security architecture for the Internet of Things with two-way authentication. In: 37th annual IEEE conference on local computer networks-workshops. IEEE, pp 956–963
14. Vasseur J, Agarwal N, Hui J, Shelby Z, Bertrand P, Chauvenet C (2011) RPL: the IP routing protocol designed for low power and lossy networks. *Internet Protocol Smart Objects (IPSO) Alliance* 36
15. Hunkeler U, Truong HL, Stanford-Clark A (2008) MQTT-S—a publish/subscribe protocol for wireless sensor networks. In: 2008 3rd international conference on communication systems software and middleware and workshops (COMSWARE'08). IEEE, pp 791–798
16. Yashiro T, Kobayashi S, Koshizuka N, Sakamura K (2013) An internet of things (IoT) architecture for embedded appliances. In: 2013 IEEE region 10 humanitarian technology conference. IEEE, pp 314–319

Digitization of Inpatient Medical Records Using Electronic Writing Pads in a Teaching Hospital



Deepak Betadur, G. Somu, and P. Naveen Kumar

Abstract *Background* The study was conducted in a large NABH accredited, 2000 bedded tertiary care teaching hospital which had footfall of 2500–3000 patients every day and 300 admissions. There was a robust system for filing of the medical records, and large manpower was employed to store, retrieve, and manage the logistics of transporting records from a central location to the respective clinics or OPDs. The clinics were spread in four building. A lot of paper usage used to occur as all the clinical notes, investigations, and prescription. Most important problem among all was that doctors not comfortable with the regular electronic medical record (EMR) solutions which involves keying in of data. All these issues called for a solution that was handwritten yet digitized. *Objectives* Identifying and designing an EMR solution that would match the requirements vis-a-vis doing a costing study. *Methodology* A standard costing study was done by researchers to identify direct labor costs, equipment costs, material costs between the two options—physical records and electronic records. The end-user challenges in using electronic records were analyzed by questionnaire. A phased implementation plan of digitization in specialty outpatient departments was coordinated. *Results* Cost of each paper file was INR 8.57/-using 156,223 files per year costed INR 1,338,831/-. Storage space rental value equivalent was INR 21,24,000/-. Direct salary costs of 73 employees in medical record department (MRD) is INR 13,406,844/- per year. The digitization was introduced in lesser crowded departments; first, their experiences were shared to customize the product with features in other departments. *Conclusion* It was found to be feasible to transform the hospital from Healthcare Information and Management Systems Society (HIMSS) level-1 digitization to HIMSS level-2 system.

Keywords Electronic medical records · Outpatient services to inpatient services · Digitization of records · Direct labor costs · HIMSS · ISO

D. Betadur

Assistant Professor, Department of Hospital Administration, SDM College, Dharwad, India

G. Somu · P. Naveen Kumar (✉)

Professor, Department of Hospital Administration, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India

e-mail: drnaveenpdr@gmail.com

1 Introduction

Storing the health information of patients in electronic devices belonging to a cluster of systems by medical professionals, from the patients' charts on a computer rather than on papers, is a digital health record or electronic health record [1]. In the 1970s, hospital-wide communications were supported by single computers, the terminals being connected in expanded networks, and functions extrapolated to financial and administrative systems in hospitals like in material management, human resources management, performance appraisal, and then helped in review of laboratory results. In the 1980s, computer networking improved integrated management systems in hospitals and departmental imaging. The 1990s saw the emergence of electronic medical records, and the 2000s saw warehousing and data analytics and clinical decision support systems in electronic medical records. The transformation of electronic health records (EHRs) is an accepted and advocated philosophy and practical issue which can provide swift, safe, high-quality care with reasonable cost to the patient care [1]. IEC 60601-1 (Institute of Electrical and Electronic Engineers) is a set of medical device communication standards, accepted by the International Standard Organization (ISO) as ISO 11073. These standards communicate patient data from medical devices like—ventilators, monitors, and infusion pumps. ISO 18308 defines “Requirements for an Electronic Health Record Reference Architecture,” and ISO/DTR 20514 defines electronic health record definition, scope, and context. In 2021, the digital health market is expected to reach 156 billion US dollars [2]. The prime solution to the many challenges faced by the healthcare sector in maintaining patient information and communicating between hospitals and physicians and departments can be solved by electronic healthcare record (EHR) systems [3]. The American Medical Association has suggested to reduce medical and medication errors, one of the key factors that reduce illegibility by giving major impetus to EHRs, as evidenced by reports published in their journals. Sometimes, patients do not get all the required treatments [4].

E-health projects are essential inclusions of the Digital India initiative to tackle the deficit of more than 1.5 million doctors. Telemedicine and teleconsultation have received a big boost after the COVID-19 pandemic. People can start getting services online like diagnostic imaging and reporting, getting to know about blood or component availability in nearest blood banks, getting online medications, online consultation, appointments, and much more. Smart people and companies are launching digital platforms for creating online access to doctors, pharmacies, information on health problems, knowledge about insurance, etc. [5]. Patient experiences have improved in 64% of their clientele, by giving access to their own health data (including test results, prescriptions, scans etc.), as opined by Indian healthcare professionals in many studies. Digital connectivity can alleviate the shortage of medical professionals by connecting electronically. Telemedicine can replace 30–40% of physical outpatient consultations; India will end up saving \$10 billion in 2025 [6]. Digital Information Security in Healthcare Act (DISHA) [7] has been introduced to

provide for electronic health data privacy, confidentiality, security and standardization, and provide for establishment of National Digital Health Authority and Health information Exchanges, and such other matters related and incidental thereto.

Why perception study is important? Perception study has been included in our study as Perception is process by which individuals organize and interpret their sensory impressions to give meaning to their environment. However, what people perceive can be substantially differ from objective reality. People's behavior is based on their perception of what reality is, not on reality itself. The world as it is perceived is the world that is behaviorally important [8].

Costing of upgrading the hospital medical records to Stage 2 from Stage 1 of HIMSS, i.e., where clinical systems are made internally interoperable, feeding the data generated to a repository which can provide seamless by a single user interface for reviewing all orders, results, and radiology and cardiology images, was done by taking the quotations from different vendors.

2 Methodology

A single center qualitative, observational study was conducted in 2032 bedded tertiary care teaching hospital.

3 Sampling Technique

To understand the perceptions of all 2380 employees to embrace electronic medical records—1600 of them are nurses, 355 are doctors, 305 are paramedical staff, and 120 are clerical staff. According to the Sekaran and Bougie's [9] guidelines, sample of 6% of the target population is appropriate to study the population. 6% of the population to become sample would be 143. According to NABH 5th edition guidelines, 160 could be the sample size for population of 2380. Considering them, appropriate size of our study sample was chosen as 160. In the sample size of this study, the class of employees represented are 47% of nurses, 26% of doctors, 14% of clerks, and 13% of paramedical staff.

4 Study Tool

To understand usage perception of electronic systems, questionnaire was distributed as part of qualitative study: It was close ended, Likert 5-point scale self-administered questionnaire with question items divided into 4 sub-sections, with 20 questions in

total. The respondents had one-month time frame to return the filled forms. About 3 items, 8, 9, 17 were negatively framed; for these questions, the responses were reversed.

5 Study Design

To understand the feasibility of introducing electronic medical records, costs were collected for paper based and electronic systems. For paper-based system, the costs included were under cost heads—direct labor cost, material costs (file, paper), building rental costs, electricity cost, and scrap costs were taken. For electronic-based system, the hardware was already existing in the form of desktops and laptops to faculty. Apart from these the writing pads in OPD, software required, operating system upgradation, security software-related all costs, training manpower and IT manpower support for few years was collected. This data were compared and recommendations suggested.

6 Results

A total of 160 surveys were distributed along with the consent letters. Regular follow-up was made, and total 148 (93%) surveys were returned. Six cases of the completed questionnaire were found to be unusable due to the missing responses. Net responses considered for the analysis were 142(88.75%). The template of questionnaire is given in Table 1.

50% of paramedical medical staff strongly agreed; 58% of the clerical staff agreed that staff should have computer usage skills. But, 24% of the doctors don't believe that computer skills are necessary for using electronic medical records. Highest agreement of digitization impact was among paramedical (69% strongly agreed) and nurses (46%). Less disagreement was seen from all the respondents; among them, 17% of doctors disagreed that digitization will make impact. Regression analysis is used to understand relationship between the independent factors (perception of computer use, perception of legal and ethical issues of digital records, perception of operational change on digitization, perception of impact of EHR adoption) on dependent factor (overall perception of digitizing the hospital). Accordingly, analysis of legal perception and EHR impact perception is insignificant because ($p = 0.79$) and ($p = 0.84$), respectively, hence, they don't have impact on overall perception of digital usage of medical records. Perception of operational change is leading to increase in overall perception of usage by factor 0.9 with the P value of < 0.05 , while perception of computer usage is kept constant (Table 1).

Table 1 Questionnaire distributed to employees to understand their perception and ease of use:

	AGE					
	DEPARTMENT					
	Your role—Doctor/Nurse/Paramedical staff/Clerk					
	Instruction—please write yes or tick between 5 and 1 for your appropriate response					
	5—strongly agree, 4—agree, 3—don’t know, 2—disagree, 1—strongly disagree					
	Questions	5	4	3	2	1
1	Are you comfortable in using desktop/laptop/palmtop/tablet					
2	Entering the data to computer/tablets is easier than on papers					
3	We need more technical skills to handle computer/tablet					
4	Most of my colleagues in hospital are well versed with the computers or tablets					
5	All employees will cooperate to make tertiary care teaching hospital digital					
6	Electronic medical records can reduce medication errors					
7	Digital records are easy for storage and retrieval compared to manual records					
8	Electronic medical records are legally not accepted					
9	Digital consent for the procedures is not accepted					
10	There will be more threat for patient’s digital records being stolen					
11	Digitization can reduce the patient’s waiting time					
12	Total digitization can reduce the manpower requirement in the hospital					
13	Care givers can dedicate more time on patient care by making system digital					
14	Total digitization can improve the patient care					
15	Digital hospital is environment friendly					
16	Total digitization will increase patient inflow					
17	There will not be good acceptance by patients					
18	Daily operations of the hospital can become more complicated on digitizing the hospital					
19	Cost of digitizing hospital outweighs the benefits					
20	All hospital staffs need to be trained before making hospital digital					

7 Hardware and Software Costing

The number of tablets (writing pads) needed to be purchased was decided on the number of beds in the hospital and the number of logins required per writing pad. The values mentioned in Table 1 have been derived by the researchers as below: Total number of beds in study hospital = 2032, having two tablets per eight beds was found to be appropriate (which will be commonly used by nurses, doctors, technicians, other staff); hence, total tablets required are = $2032/4 = 508$, 10% (50) spare

Table 2 The costs for physical records and electronic records

	Cost heads	Paper-based records	Electronic medical records
1	Direct labor (salary) costs	13,406,844	
2	Material (file) costs	5,315,240	
3	Storage space costs	843,000	
4	Electricity consumption costs	699,036	
5	Total cost of digitizing existing medical records		2,847,891
6	Hardware cost		6,708,333
7	Software cost ^a		13,101,144
		23,579,360	22,657,368
8	Environmental impact on trees	2451 trees	

^aSoftware cost quotation included—software licensing + training + hosting + integration cost + data storing + data security

tablets need to be kept, total tablets required = 558. Cost of each writing pad = INR 25,000/-. Total cost of hardware needed is = $25,000 \times 558 = \text{INR } 13,950,000/-$. Company gave three-year warranty to the tablets. Annual hardware cost of inpatient department = $13,950,000/3 = \text{INR } 4,650,000/-$ software cost calculation. For inpatient department, software quote was taken from Chennai-based Company. Quoted INR 100/login/day, with below features included, inpatient EHR module, hosting cost, server cost, data security cost, training cost, 200-man days of customization (8 h/day), integration with existing hospital information system, lab information system, and picture archiving and communication system (PACS), basic informative dash boards, statistics generation, finance and billing modules, regular upgrading cost, number of logins are required for inpatients in the hospital department doctors = 90 (As there is unit-based practice in study hospital, under unit head's name other consultants can login).

Nurses log in calculation: Total nurses in the hospital = 1600, on an average, there will be an in-charge nurse for each 10 nurses including, nursing and assistant nursing superintendents. Logins required = $1600/10 = 160$.

Administrative and IT logins required: This includes superintendent's office and operations team, finance, insurance and billing team, medical records officer, and customer care. Total 50 logins found to be appropriate for administrative team. Total logins required for the hospital = $90 + 160 + 50 = 300$.

Cost of software per day = $300 \times 100 = 30,000$, annual software cost for inpatient department = $30,000 \times 365 = \text{INR } 10,950,000/-$. Total cost of inpatient EHR = Hardware component + software component, Total cost = $4,650,000 + 10,950,000 = \text{INR } 15,600,000/-$. Similarly, electronic medical records cost adoption was projected for outpatient department, the total of which came up to INR 4,209,477/-. When we add both inpatient and outpatient cost, it will INR 20,250,000/-. To this, if we add existing paper records scanning per piece cost, we will finally arrive at an INR 22,657,368/- as the total cost of electronic medical records adoption.

Annual per bed digitization cost = INR 12,776/- per bed (for 75% occupancy rate and outpatient to inpatient conversion rate is 10%), extra cost burden for the hospital = INR 5,533,659/- per year or Rs.2723/- per year/bed.

8 Environmental Impact Study

University to which the hospital belongs has come with new environmental policy to preserve the environment. To align with the university goals environment impact, study was undertaken as part of the study. Results of which is as follows: Number of paper sheets used by the hospital per year = 20,426,496. Paper sheets (of 150gsm) generated from one mid-sized tree = 8333. Number of trees cut down every year to provide papers in our study = 20,426,496 sheets/8333 trees = 2451 mid-sizes trees.

Even tablets consume electricity. The CPU in a tablet consumes the maximum power up to 70–80%, and the rest of the components consume very less electricity. 20–50 W of electricity is used by tablets and laptops that can be reduced down by using the power-saver modes. Expecting that all our tablets are going to be BEE rated with energy star ratings, their power consumption will come down by 30%. 20 W will be the power consumption for each tablet; with an assumption of 5 h per day continuous usage it transforms into 40 kWh per annum. Each kWh being 1 unit, we would end up paying electricity bill of 280 Rupees per tablet per year.

9 Discussion

Of all the 148 respondents in our study, 30% people strongly agreed, and 33% agreed that having computer usage skills is necessary for digitization of the hospital, among them 28% of the doctors strongly agreed to the fact. These results of our study are consistent with study done in Saudi-Arabia which states that 31% of physicians perceived that computer usage skills are necessary before electronic medical records implementation [10].

A previous study in the same hospital inferred that nurses had positive perception toward daily operational changes of adopting the EHR; in our study, we found that 18% of nurses strongly agreed, and 36% of nurses agreed that EHR adoption will positive change in their daily operations [11].

In our study, we found that a cost adopting the EHR from paper-based system is around INR 2.26 crores/year including the transformation cost from paper to paperless system. A study done by Neil Dennis USA stated that both financial and non-financial costs of implementing the electronic medical records in a primary care setup are \$162,000, with \$85,500 in maintenance expenses during the first year [12].

Medical records department policy states that medical records will be kept only up to 3 years from the last visit of the patient. To digitize all the files, it takes 3 years. Extra cost needs to be spent by the hospital for digitizing the outpatient department

as compared to existing paper records for the next 3 years that is Rs.1,730,304/year. Apart from this, the annual cost of electronic medical records for OPD would reduce to Rs.4,209,477/year which is Rs.1,117,587 less than the present hospital expenditure on paper-based medical records.

For inpatient department, MRD policy is 5 years records preservation from the last visit. To completely move from paper to paperless system in IP department, it is estimated to take 5 years. Extra cost to be borne by the hospital for the first 5 years is Rs.3,803,352/year. After 5 years, annual cost would be Rs.15,600,000/year as there is no transformation cost. This cost is only Rs. 500,568 more than the present spending on the paper-based records.

Our results showed around \$5000 per physician per year (only, professors are considered, as there is unit-based practice in this study hospital, with unit being a team of 3–4 faculty under one professor) for the basic electronic health records. In Boston and New York city, a network of records among the community and clinics results in direct costs of software, hardware, manpower support, a \$34,000–\$39,000 bill on implementing the electronic health records per physician per year.

10 Conclusion

Digitization of tertiary care hospital is a long journey as there are many complex processes, procedures taking place every day. It should be planned strategically to adopt the electronic health records in a phase-wise manner with adoption starting from a less complex area to more complex area. Users play key role in adoption of EHR.

References

1. Mogli GD (2015) Chapter 13. Health records: paper to paperless, re-print, vol 1. Jaypee Medical Publications, India, pp 299–300
2. Statista Research Department (2021) <https://www.statista.com/statistics/387867/value-of-worldwide-digital-healthmarket-forecast-by-segment>. Accessed 25 Jul 2021
3. Mihailescu M (2017) Understanding healthcare digitalization: a critical realist approach, <https://aisel.aisnet.org/iciis2017/IT-and-Healthcare/Presentations/7>. Accessed 11 Aug 2021
4. Swartz N (2004) A prescription for electronic health records. *Inf Manag J* 38(4):20–26
5. McKinsey Update June 29, 2021: <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality>. Accessed 8 Aug 2021
6. Press Release, MoS, Science & Technology, Posted on 28th October 2021, <https://pib.gov.in/PressReleaseDetail.aspx?PRID=1767253>. Accessed on 24 Oct 2021
7. F.No Z-18015/2312017-eGov/Government of India/Ministry of Health & Family welfare/eHealth Section
8. Robins S, Judge T, Vohra N (2016) Organizational behaviour. Pearson Publications, 15th edn., 176–183

9. Sekaran U, Bougie R (2016) *Research methods for business: a skill building approach*, 7th edn. Wiley, New York
10. Mohamed BA, El-Naif M (2015) Physicians', nurses' and patients' perception with hospital medical records at a military hospital in riyadh, saudi arabia. *J Fam Community Med* 12(1):49–53
11. Pera NK, Kaur A, Rao R (2014) Perception of electronic medical records (EMRs) by nursing staff in a teaching hospital in India. *Int J Adv Med Health Res* 1:75–80
12. Mostashari F, Tripathi M, Kendall M (2019) A tale of two large community electronic health record extension projects. *Health Aff (Millwood)* 28(2):345–356

Miniaturization of Dual Shaped Monopole Antenna for UWB Application



Ranjeet Kumar, Rashmi Sinha, Arvind Choubey, Santosh Kumar Mahto, Pravesh Pal, and Praveen Kumar

Abstract With a recent development and phenomenal activity in the area of microwave, there is need to maintain quality of service and high data rate. Therefore, requirement to design an antenna which provide these facilities. This paper presents a double regular hexagonal radiating patch, defected ground, and small circular slot etched out from the middle part of radiating patch. It demonstrated the Ultra-Wide band (UWB) operation, which works efficiently in entire band from 2.52 to 12.91 GHz. The overall dimension of proposed antenna is $31 \times 51.5 \text{ mm}^2$ Simulated in HFSS and has almost stable radiation pattern of E- and H-plane, positive gain of 6 dBi, and 134% of bandwidth in the entire band. This microstrip antenna is simulated and fabricated to verify its result, the equivalent circuit model is also constructed to verify simulate a measured result, which works efficiently in S, C, and X band applications.

Keywords Antenna theory · Bandwidth · Microstrip antenna

R. Kumar (✉) · R. Sinha · P. Pal · P. Kumar

Department of Electronics and Communication Engineering, National Institute of Technology, Jamshedpur, India

e-mail: ranjeet2k4kumar@gmail.com

R. Sinha

e-mail: rsinha.ece@nitjsr.ac.in

P. Pal

e-mail: praveshpal9@gmail.com

P. Kumar

e-mail: 2018rsec005@nitjsr.ac.in

A. Choubey

Indian Institute of Information Technology, Bhagalpur, India

e-mail: achoubey.ece@nitjsr.ac.in

S. K. Mahto

Department of Electronics & Communication Engineering, Indian Institute of Information Technology, Ranchi, India

e-mail: skumar@iiitranchi.ac.in

1 Introduction

Recently development in the field of wireless communication system, has led to demand of UWB antenna, which gives wide bandwidth, low transmission rate, less distortion, high gain and most importantly very less probability to intercept. Antenna researcher community face the problem and designed the miniaturized antenna to fulfill the requirement of UWB. The microstrip planar antenna is chosen because of its attractive feature of monitorization, low profile, low manufacturing cost and appropriate for production [1]. There are different shaped of UWB antenna has been demonstrated in [2–15]. A hexagonal antenna shaped with partial ground plane [2], L-shaped slot-based patch antenna for bandwidth enhancement [3], U-shaped aperture dual band UWB antenna [4], low profile band rejected radiating patch antenna [5], to prediction of notch frequency and T-shaped slot in partial ground [6, 7], a miniaturized antenna with band-notch characteristics, multi resonator, triangular-hexagonal based antenna for UWB antenna has been presented [8–11] to highlight the recently published works in the field of antenna.

This works has two regulars hexagonal shaped patch placed at 30° about its center and has a small circular slot at the middle of proposed antenna. It is designed on FR4 substrate of compact size $31 \times 51.5 \text{ mm}^2$. This proposed antenna presented here has compact size, easy to fabricate, wide bandwidth, stable radiation pattern, and positive gain. The bandwidth of antenna may be enhanced by inserting the circular slot at its center.

2 Antenna Design and Implementation

The proposed microstrip feed hexagonal printed antenna and its evolution step are show in Fig. 1. A double regular hexagonal radiating patch, defected ground, and small circular slot etched out from the middle part of radiating patch. In first step, a regular hexagonal shape (ANT1) of radius R is taken as depicted in Fig. 1a. A similar type regular hexagonal shape is placed at 30° to get ANT2 as shown in Fig. 1b. And finally, small circular slot is inserted on the center of radius ' r ' to obtained ANT3 as portrayed in Fig. 1c. This antenna structure is designed on FR4 epoxy dielectric material constant of $(\epsilon) = 4.4$, height $(h) = 1.6 \text{ mm}$, and loss tangent $(\delta) = 0.02$. The width of defected ground plane $(wg) = 13.5 \text{ mm}$, radius of hexagonal patch $(R) = 13 \text{ mm}$, and circular slot $(r) = 2 \text{ mm}$ is the optimized parameter of proposed antenna. Simulated and fabricated proposed antenna captured in Fig. 2a, b.

The resonant frequencies of regular Hexagonal Patch Antenna (HPA) are calculated by Eq. (1) [11].

$$f_0 = \frac{T_{mn}C}{5.714R_e\sqrt{\epsilon_{reff}}} \quad (1)$$

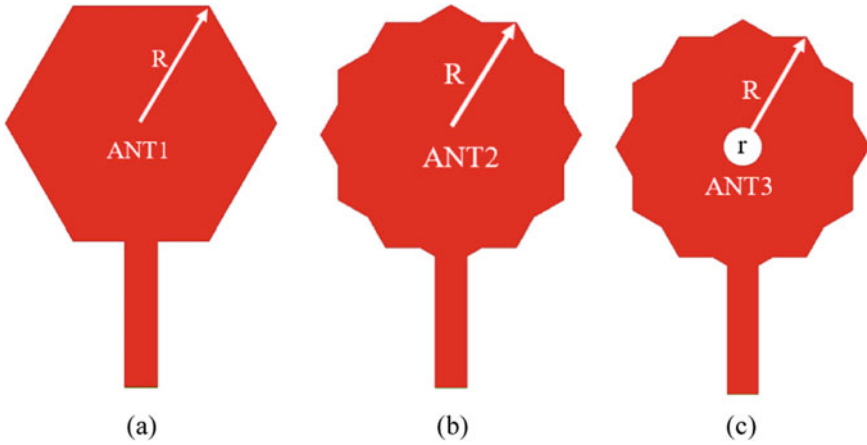


Fig. 1 Deriving stage of dual hexagonal shaped antenna

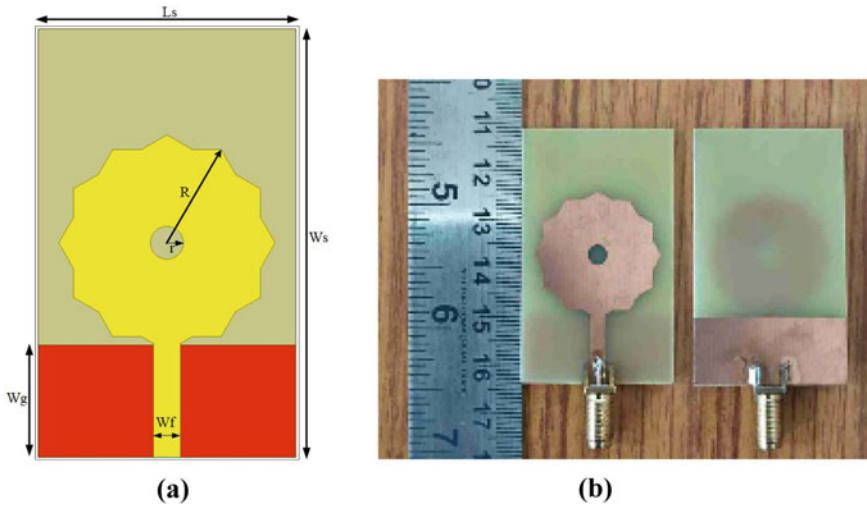


Fig. 2 a Simulated and b prototype antenna

where C = speed of electromagnetic wave in free space, ϵ_{reff} = effective dielectric material constant, the operative radius (R_e) of patch antenna is calculated through Eq. (2) [11].

$$R_e = R_c \cdot \sqrt{\left(1 + \frac{2 \cdot h}{R_c \cdot \pi \cdot \epsilon_r} \left[\ln\left(\frac{\pi \cdot R_c}{2 \cdot h}\right) + 1.7726 \right] \right)} \quad (2)$$

where R_c, h, ϵ_r is the radius of the circular geometry, thickness, and material dielectric constant, respectively. By equating the area of circular and hexagonal shape, the resonance frequency can be determined by Eq. (3) [11].

$$\pi.(R_c)^2 = (1.5).\sqrt{3}.(S_h)^2 \tag{3}$$

Effective dielectric material constant (ϵ_{reff}) is given by Eq. (4) [11]

$$\epsilon_{reff} = \frac{1 + \epsilon_r}{2} \tag{4}$$

where S_h, ϵ_r are the side of hexagonal patch and relative dielectric constant.

3 Finite Element Method (FEM) Analysis

In this section, FEM is used to demonstrate the how and at what extend the potential is varied. Partial differential equation (PDE), in MATLAB toolbox, helps to analyze numerically. There are 223 elements, 55 edges, and 139 nodes available in this antenna. The elliptical, parabolic, hyperbolic form of generic mode and electric, magnetic form of electromagnetic mode are calculated through Eqs. (5–9) [12] and it is graphical representation is depicted in Fig. 3.

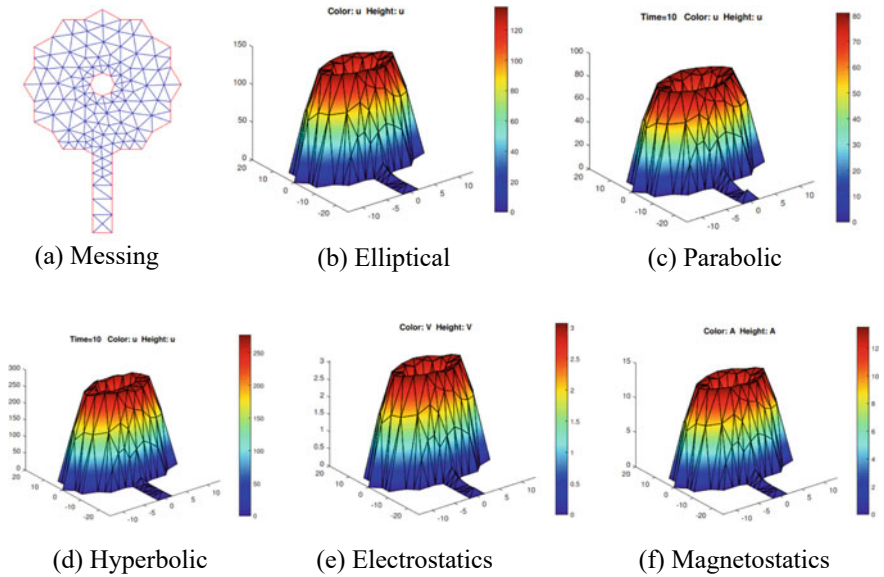


Fig. 3 The variation of potential in different modes

$$-\text{div}(c * \text{grad}(U)) + a * U = f \tag{5}$$

$$d * U' - \text{div}(c * \text{grad}(U)) + a * U = f \tag{6}$$

$$d * U'' - \text{div}(c * \text{grad}(U)) + a * U = f \tag{7}$$

$$-\text{div}(\varepsilon * \text{grad}(U)) = \rho, \mathbf{E} = -\text{grad}(U) \tag{8}$$

$$-\text{div}\left(\frac{1}{\mu} * \text{grad}(\mathbf{A})\right) = \mathbf{J}, \mathbf{B} = -\text{curl}(\mathbf{A}) \tag{9}$$

where U = scalar potential, f = source, \mathbf{E} = electric field, \mathbf{B} = magnetic field density, \mathbf{A} = magnetic vector potential, ρ = charge density, ε = coefficient of dielectric, μ = magnetic permeability, and a, c, d are scalar constant.

4 Lumped Circuit Model Analysis

The return loss characteristic of radiating microstrip patch antenna is developed by help of transmission line model. Its result can be verified by lumped equivalent circuit model. This model can be constructed using basic elements of resistor (R), inductor (L), capacitor (C), and conductance (G). The microstrip feedline, resonating patch, and partial ground can be approximated as a series combination of surface resistance (R_s), and surface inductance (L_s) and is calculated by Eq. (10–11) [13–16]

$$R_s = \frac{1}{W_f \sigma_{\text{cond}} \delta} \tag{10}$$

$$L_s = \frac{1}{W_f \sigma_{\text{cond}} \delta \omega} \tag{11}$$

where $\delta = \sqrt{\frac{2}{\omega \sigma_{\text{cond}} \mu_0}}$ and

The conducting top and bottom layer are separated by FR-4 substrate. Electrically, this suggest a parallel combination of capacitor C and its value is calculated by Eq. (12) [13–16]. The conductance G , due to dielectric loss can be omitted for sake of simple analysis.

$$C = \varepsilon_{r\text{eff}} 2.85 \frac{1}{\ln \left\{ 1 + \frac{1}{2} \left(\frac{8h}{W_{\text{eff}}} \right) \left[\left(\left(\frac{8h}{W_{\text{eff}}} \right) + \sqrt{\left(\frac{8h}{W_{\text{eff}}} \right)^2 + \pi^2} \right) \right]} \right\}} \tag{12}$$

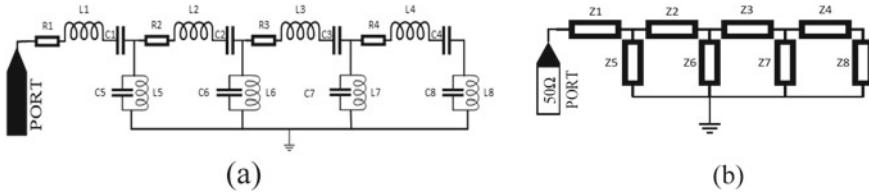


Fig. 4 a Equivalent circuit model antenna, b Reduced form

where

$$W_{\text{eff}} = W_f + \frac{t}{\pi} \ln \left\{ \frac{4e}{\sqrt{\left(\frac{t}{h}\right)^2 + \left[\frac{1}{\pi\left(\frac{W_f}{t} + 1.10\right)}\right]^2}} \right\}$$

where t = conductor height, σ_{cond} = surface conductivity, ω = angular frequency, δ = skin depth, C = capacitance of capacitor, W_{eff} = effective width, and ϵ_{reff} = effective material dielectric constant, μ_0 = permeability in the free space. All assumptions are considered and lumped equivalent circuit is drawn as Fig. 4.

The input impedance ($Z_{\text{in}} = R + jX$) is numerically calculated from Fig. 4b.

$$Z_i = R_i + j\omega L_i + \frac{1}{j\omega C_i} = R_i + \frac{1 - \omega^2 C_i L_i}{j\omega C_i};$$

and

$$Z_n = \frac{j\omega L_n \times \frac{1}{j\omega C_n}}{j\omega L_n + \frac{1}{j\omega C_n}} = \frac{j\omega L_n}{1 - \omega^2 C_n L_n}$$

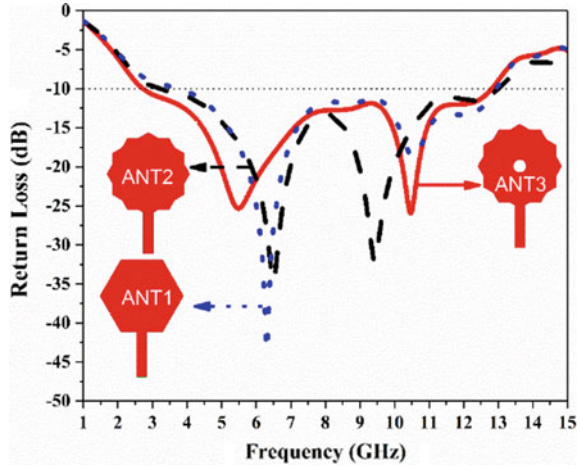
where $i = 1, 2, 3, 4$, and $n = 5, 6, 7, 8$; By putting the value of i and n , Z_1 to Z_8 can be obtained such as $i = 1$ and $n = 5$ will give

$$Z_1 = R_1 + j\omega L_1 + \frac{1}{j\omega C_1} = R_1 + \frac{1 - \omega^2 C_1 L_1}{j\omega C_1};$$

and

$$Z_5 = \frac{j\omega L_5 \times \frac{1}{j\omega C_5}}{j\omega L_5 + \frac{1}{j\omega C_5}} = \frac{j\omega L_5}{1 - \omega^2 C_5 L_5}$$

Fig. 5 Comparison of return loss of evolved antenna



Total input impedance (Z_{in}) is calculated as,

$$Z_{in} = Z1 + \frac{Z5 \times \left[Z2 + \frac{Z6 \times \left[Z3 + \frac{(Z4+Z8) \times Z7}{Z4+Z8+Z7} \right]}{Z6+Z3 + \frac{(Z4+Z8) \times Z7}{Z4+Z8+Z7}} \right]}{Z5 + Z2 + \frac{Z6 \times \left[Z3 + \frac{(Z4+Z8) \times Z7}{Z4+Z8+Z7} \right]}{Z6+Z3 + \frac{(Z4+Z8) \times Z7}{Z4+Z8+Z7}}} \tag{13}$$

The reflection coefficient (Γ) is determined by Eq. (14) with help of Eq. (13). Voltage standing wave ratio (VSWR), and return loss (RL) is given by Eqs. (15–16), respectively.

$$\Gamma = \frac{Z_{in} - Z_0}{Z_{in} + Z_0} \tag{14}$$

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{15}$$

$$\text{Return loss} = 20 \log|\Gamma| \tag{16}$$

where Z_0 is the characteristic impedance (50Ω).

5 Result and Discussion

Performance parameters of the simulated and fabricated antenna are discussed in this section. The proposed antenna (ANT3) in Fig. 1c represents ultra-wide band characteristic as can be seen from Fig. 5. The observations and its values are enumerated in Table 1.

Figure 6, shows the scattering parameters of proposed antenna, which is obtained by simulation, measurement, and lumped circuit model. All outcomes are very close to each other. The simulated bandwidth of this antenna is of 13.09 GHz, ranging from 2.52–12.91 GHz. it can be seen from plot that there is little variation of return loss characteristic at lower and higher resonance frequency. The higher order modes are activated to enhance the wide bandwidth.

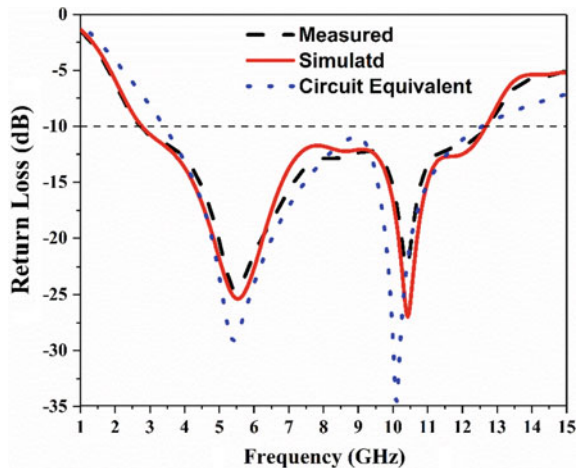
The E and H-plane, at its resonance frequencies of 5.4 and 10.4 GHz are depicted in Fig. 7 and can be seen from it that, simulated and measured radiation patterns are almost stable and has good co-polarization than cross-polarization.

The simulated and measured peak gain of the antenna is 6 dBi at 11.57 GHz depicted Fig. 8a. Average gain and radiation efficiency are 3.5 dBi and 87% in entire operating band as captured in Fig. 8b.

Table 1 Tabulated comparison between resonance frequencies

Antenna	1st resonance	2nd resonance	Bandwidth (GHz)
ANT1	6.5	9.4	3.17–13.00
ANT2	6.2	10.5	3.77–12.94
ANT3	5.4	10.4	2.52–12.91

Fig. 6 Variation of return loss of the antenna



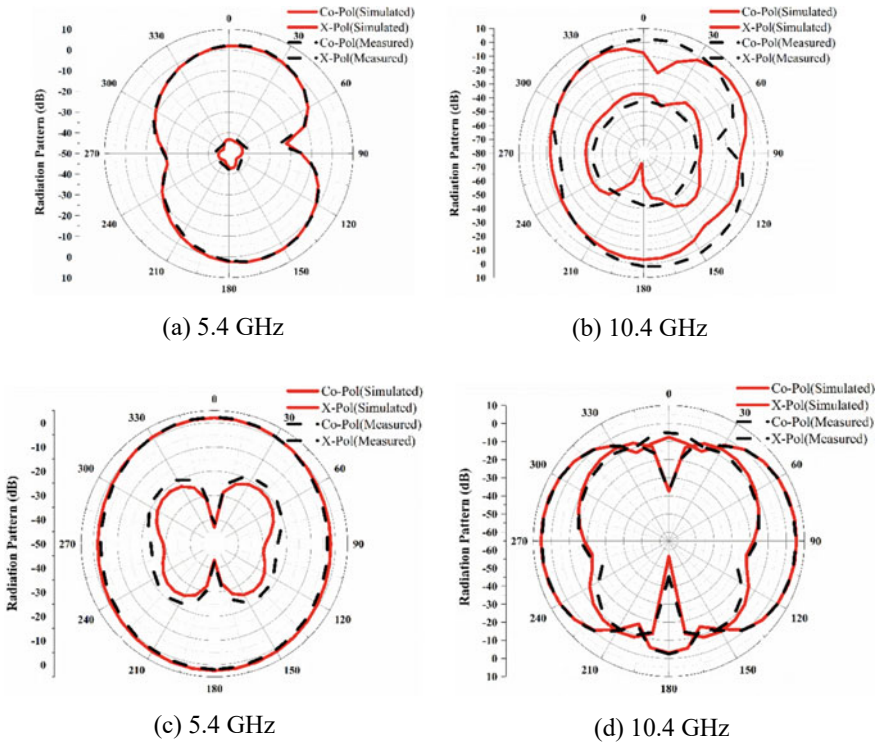


Fig. 7 The simulated and measured a, b E-plane, and c, d H-plane

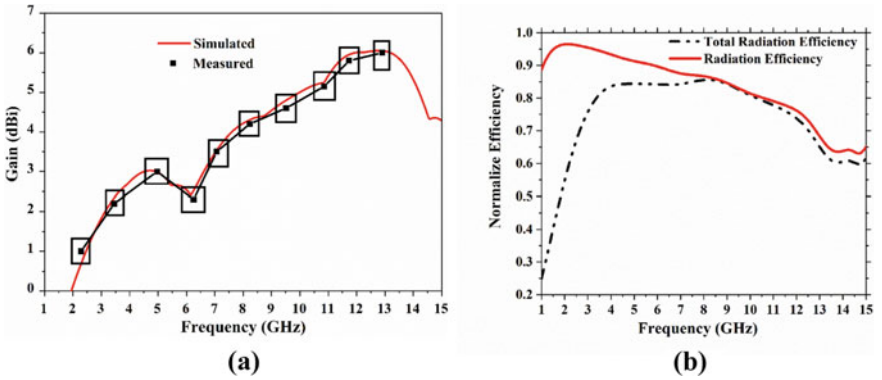


Fig. 8 a Peak gain, and b radiation efficiency versus frequency

6 Conclusion

This paper presents dual hexagonal shaped microstrip antenna with partial ground plane of $31 \times 51.5 \text{ mm}^2$, which exhibits an impedance bandwidth of 10.39 GHz, with VSWR less than 2. The proposed antenna confirms that it maintains the favorable conditions for UWB applications such as WLAN, WiMAX, 5G, S, C, and X band. There is a little discrepancy in simulated and measured results due to permissible fabrication error.

References

1. Yang HD (2005) Miniaturized printed wire antenna for wireless communications. *IEEE Antennas Wirel Propag Lett* 4:358–361
2. Pramudita AA, Kurniawan A, Suksmono AB (2008) Hexagonal monopole strip antenna with rectangular slot for 100–1000 MHz SFCW GPR applications. *Int J Antennas Propag*
3. Zaker R, Ghobadi C, Nourinia J (2009) Bandwidth enhancement of novel compact single and dual band-notched printed monopole antenna with a pair of L-shaped slots. *IEEE Trans Antennas Propag* 57(12):3978–3983
4. Zhou HJ, Sun BH, Liu QZ, Deng JY (2008) Implementation and investigation of U-shaped aperture UWB antenna with dual band-notched characteristics. *Electron Lett* 44(24):1387–1388
5. Jang JW, Hwang HY (2009) An improved band-rejection UWB antenna with resonant patches and a slot. *IEEE Antennas Wirel Propag Lett* 8:299–302
6. Dissanayake T, Esselle KP (2007) Prediction of the notch frequency of slot loaded printed UWB antennas. *IEEE Trans Antennas Propag* 55(11):3320–3325
7. Ojaroudi M, Ghobadi C, Nourinia J (2009) Small square monopole antenna with inverted T-shaped notch in the ground plane for UWB application. *IEEE Antennas Wirel Propag Lett* 8:728–731
8. Liu J, Gong S, Xu Y, Zhang X, Feng C, Qi N (2008) Compact printed ultra-wideband monopole antenna with dual band-notched characteristics. *Electron Lett* 44(12):710–711
9. Chang TN, Wu MC (2008) Band-notched design for UWB antennas. *IEEE Antennas Wirel Propag Lett* 7:636–640
10. Ma TG, Hua RC, Chou CF (2008) Design of a multiresonator loaded band-rejected ultrawideband planar monopole antenna with controllable notched bandwidth. *IEEE Trans Antennas Propag* 56(9):2875–2883
11. Darimireddy NK, Reddy RR, Prasad AM (2018) A miniaturized hexagonal-triangular fractal antenna for wide-band applications [antenna applications corner]. *IEEE Antennas Propag Mag* 60(2):104–110
12. Sadiku MN (2018) *Computational electromagnetics with MATLAB*. CRC Press
13. Bogatin E (1988) Design rules for microstrip capacitance. *IEEE Trans Comp, Hybrids, Manuf Technol* 11(3):253–259
14. Bahl I, Bhartia P, Stuchly S (1982) Design of microstrip antennas covered with a dielectric layer. *IEEE Trans Antennas Propag* 30(2):314–318
15. Garg R, Bhartia P, Bahl II, Ittipiboon A (2001) *Microstrip antenna design handbook*. Artech House
16. Balanis CA (2016) *Antenna theory: analysis and design*. Wiley

Improvement of Isolation Between Co-polarization and Cross-polarization Radiation Using Circular Shorting Columns



Manoj Sarkar, L. Lolit Kumar Singh, Sudipta Chattopadhyay, and Abhijyoti Ghosh

Abstract In this paper, six circular shorting columns combined with rectangular microstrip antenna are proposed to increase the polarization purity in the orthogonal plane with stable radiation pattern. The circular shorting columns have been placed along the length of the patch between the ground plane and patch. From the proposed structure, around 31 dB polarization purity over a wide elevation angle has been obtained as compared to conventional microstrip patch antenna on FR-4 substrate. The proposed structure takes care of the higher-order orthogonal mode at the non-radiating edge without affecting any fundamental (dominant) mode characteristics.

Keywords Rectangular microstrip patch antenna · Cross-polarized radiation · Co-polarized radiation · Circular shorting columns

1 Introduction

In the field of era modem communication, microstrip antenna (MA) is a most suitable radiator which leads toward the compact size. Therefore, antenna designing becoming crucial factor to achieve higher radiation performance in the field of wireless communication. Rectangular microstrip antenna (RMA) is one of the most popular shape among the patch antennas. It has several wide varieties of applications due to its advantages like light in weight, dual-polarization, efficient radiation, and low-cost fabrication process. Apart from these excellent properties, there is some drawback like narrow impedance bandwidth (3–4%), excitation of surface wave, low gain, low power handling capacity, and poor polarization purity [1, 2]. In fundamental TM_{10} mode, RMA is radiating at the broadside direction (co-polarization radiation (CPR)); however, some higher-order orthogonal radiation is also taken place which is called cross-polarization radiation (XPR) [3]. The XPR is the most prominent in the microstrip patch antenna for probe feeding structure which further surges with increasing substrate thickness and dielectric constant [4]. The XPR fields are more

M. Sarkar · L. Lolit Kumar Singh · S. Chattopadhyay · A. Ghosh (✉)
Department of Electronics & Communication Engineering, Mizoram University, Tanhril, Aizawl, Mizoram 796004, India
e-mail: abhijyoti_engineer@yahoo.co.in

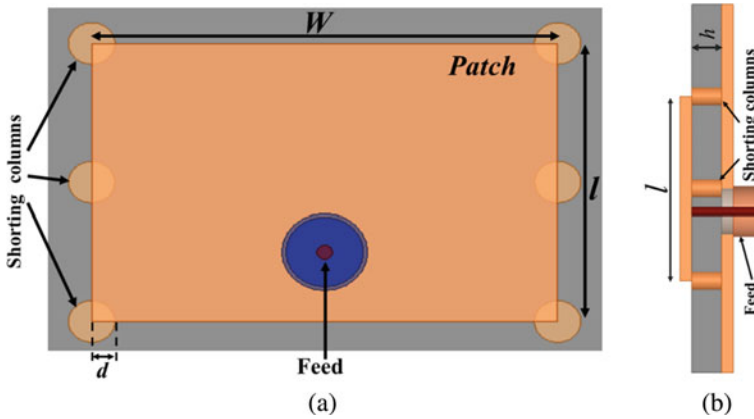


Fig. 1 RMA with circular short **a** view from top, **b** cross-sectional view

significant in the H plane than in the E plane [5, 6]. Improving CPR-XPR isolation is one of the most critical research issues in present day.

Several antenna researchers have exploited various technique to improve XPR performance. Alter the feeding mechanism to reduce the XPR has been reported in [7–9]. Some modified feed structures were investigated in [10] for reduction of XPR around 20 dB. CPR-XPR isolation around 23 dB is obtained by aperture coupling [11]. Modification of ground structure to accomplish CPR-XPR around 10–12 dB is documented in [12].

Different slotted ground structure has been proposed to achieve better CPR-XPR isolation in [13–17]. Around 25 dB CPR-XPR isolation has been documented in those reports.

In the last few years, some new structures of MA have been investigated [18–20], where radiating patch and ground is shorted by connecting different type of shorts for achieving broad impedance bandwidth and higher CPR-XPR isolation.

Therefore, in this investigation, six circular shorting columns have been deployed along the length of the RMA on FR-4 material with dielectric constant (ϵ_r) 4.4 to study the CPR-XPR isolation (Fig. 1).

2 Structural Evolution

There are two major contributors of the XPR in RMA. One is the unwanted radiation from the non-radiating sides of the patch in higher-order orthogonal modes mainly TM_{02} mode, and another one is orthogonal components of electric field vector that are generated mainly from the corner of the RMA when RMA radiates in dominant mode [6]. To minimize these two sources of XPR, three pairs of circular shorting columns with radius (d) have been placed along the length of the RMA to alter the

electric fields at that portion of the RMA. Among these six circular shorting columns, four are placed at the four corners of the RMA to tackle the fields at that position. As soon as the shorting columns with radius $d = 0.2$ mm are placed along the length, the strong electric fields at those location vanish or become very weak. Fig. 2a shows the electric field distribution over substrate at TM_{10} , while Fig. 2b shows the electric field distribution over substrate at TM_{10} after placing the shorting columns.

To find the structure that will provide the best CPR-XPR isolation, the radius of the shorting columns varies starting from 0.2 mm. The variation of CPR-XPR isolation is depicted in Fig. 3 as a function of shorting column radius. It is clearly evident that by increasing shorting column radius, the overall CPR-XPR isolation has improved initially and attains the highest isolation at $d = 0.6$ mm, and then,

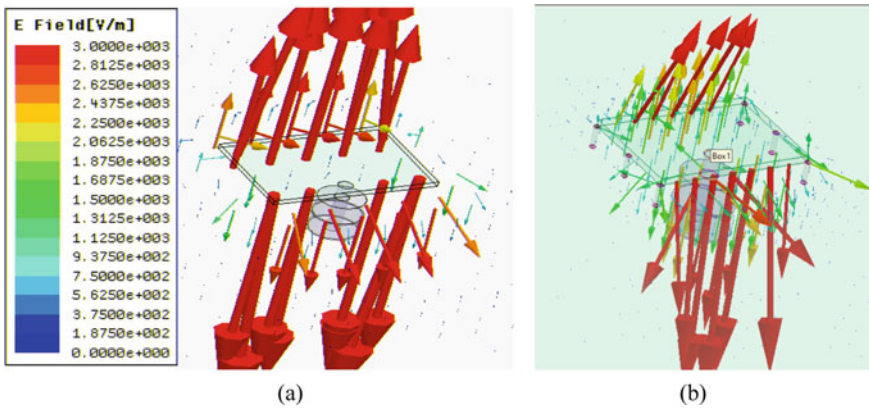


Fig. 2 The distribution of electric field over substrate **a** RMA without shorting columns, **b** RMA with circular shorting columns

Fig. 3 Variation of CPR-XPR isolation of present structure as a function of shorting column radius (d)

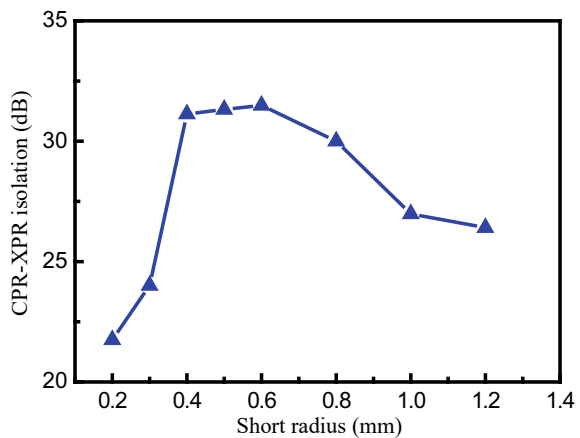
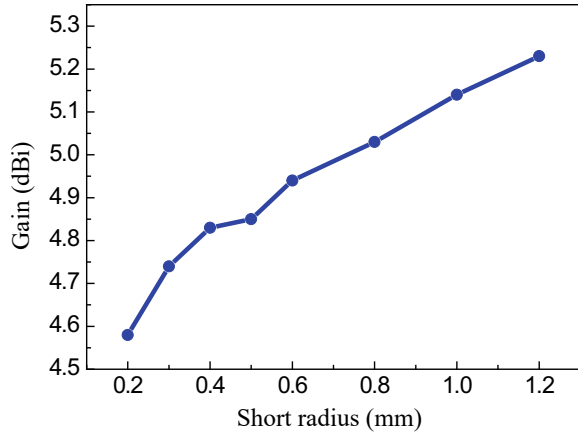


Fig. 4 Variation of gain of present structure as a function of shorting column radius(d)



the CPR-XPR isolation decreases. Optimum isolation of 31.5 dB has been obtained when shorting column radius is 0.6 mm.

The CPR gain with respect to the shorting column radius (d) is shown in Fig. 4. Due to the placing of the shorts along the length of the RMA, the effective patch area where the electric fields can stand shrinks, so the electric field concentration over the patch surface increases. For this, the gain of the present structure should increase. This can be evident from Fig. 4.

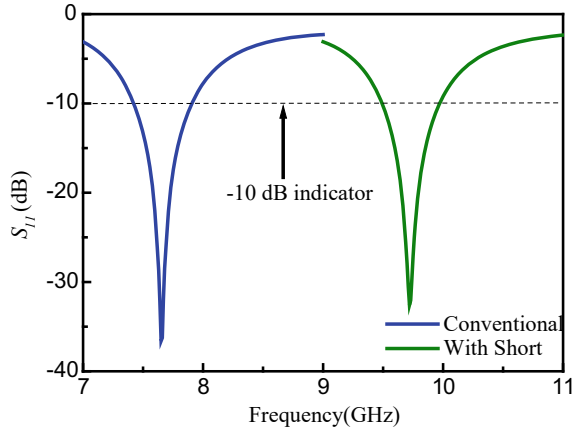
3 Final Structure

At first, a conventional RMA (CRMA) is designed having patch dimension ($l \times w$) of $8 \times 12 \text{ mm}^2$ on FR-4 material with dielectric constant (ϵ_r) 4.4 and height (h) 1.575 mm. The ground plane dimension has been considered as $66 \times 64 \text{ mm}^2$. After those, three pairs of circular shorting columns with radius $d = 0.6 \text{ mm}$ have been deployed along the length of the radiating patch of the RMA where center of the circular shorting columns coincides with patch corner of RMA. This way the final optimized structure has been build which is shown in Fig. 1.

4 Simulated Outcomes from Final Structure

This section discusses the outputs obtained from the present structure with shorting column radius (d) 0.6 mm using [21]. Figure 5 shows the S_{11} profile of conventional RMA and optimum structure of six shorting column inserted RMA. RMA without any circular shorting columns (i.e., conventional RMA) resonant at 7.65 GHz whereas circular shorting columns integrated RMA with radius (d) 0.6 mm resonant

Fig. 5 Simulated S_{11} profile of conventional RMA and present circular shorting column integrated RMA



at 9.72 GHz, but both the structures provide very good impedance matching which is confirmed from the same figure.

The radiation pattern of the E plane has been revealed in Fig. 6 for both CRMA and circular short united RMA. The figure clearly shows that the CPR and XPR of present structure are pretty similar to a conventional RMA. The CPR profile has not been disturbed due to the addition of the shorting columns. The less significant E plane XPR is also below -40 dB in both the cases.

Figure 7 shows the H plane CPR and XPR profile at resonant frequency. The H plane CPR of the shorting column united RMA is fairly similar to that of the CRMA while the XPR reduces significantly around 19 dB as compared to the CRMA. The CPR-XPR isolation in the shorting column united RMA is around 34 dB, while for CRMA, it is only around 14 dB. Present investigation has achieved the said CPR-XPR isolation over the whole elevation angular range.

Fig. 6 Variation of simulated conventional and proposed structure radiation pattern in E plane

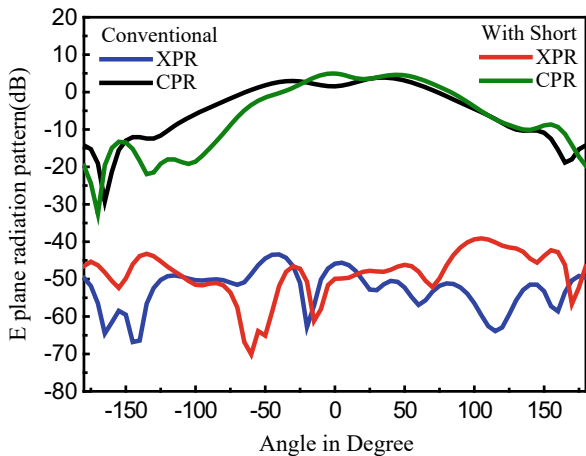


Fig. 7 Variation of simulated conventional and proposed structure radiation pattern in H plane

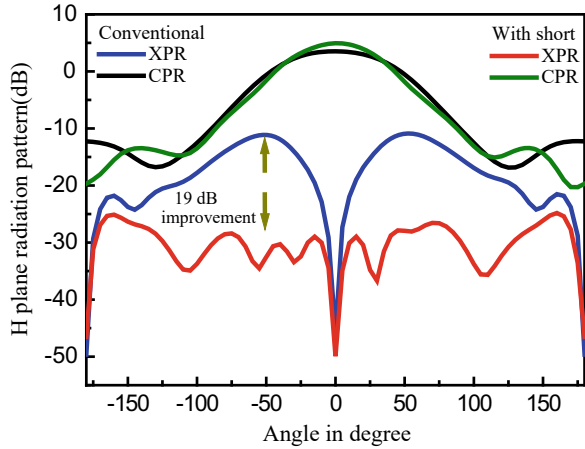


Figure 8 shows the distribution of the electric field vector over substrate and electric field magnitude on the patch surface for the six circular shorting column united RMA with $d = 0.6$ mm. The presence of circular shorting columns reduces the field strength along the length of the patch which are mainly responsible for the XPR. Thus, the XPR for the present model decreases.

5 Conclusion

Three pairs of circular shorting column united RMA has been investigated for enrichment of XPR over CRMA. The CPR-XPR isolation of 34 dB is achieved from the optimum model, while for the same structure without shorting columns, the CPR-XPR isolation is only 14 dB. The improvement of the CPR-XPR isolation is seen over the entire elevation angle region. This investigation may be helpful in era communication where high polarization purity is requisite.

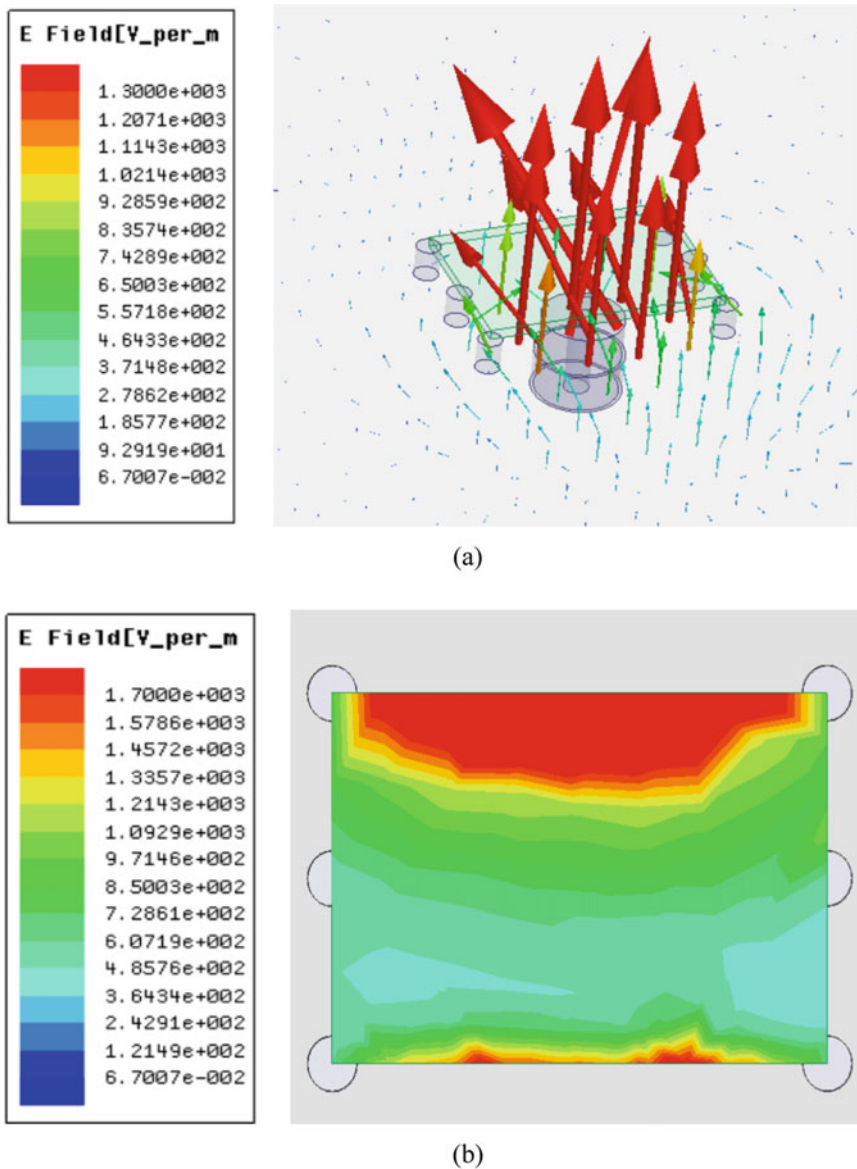


Fig. 8 Proposed structure **a** distribution of electric field vector over substrate, **b** electric field magnitude over patch

References

1. Garg R, Bhartia P, Bahl I, Ittipiboon A (2001) *Microstrip antenna design handbook*. Artech House, Norwood
2. Guha D, Antar YMM (eds) (2011) *Microstrip and printed antennas—new trends. Techniques and applications*. Wiley, Chichester
3. Kishk AA, Shafai L (1986) The effect of various parameters of circular microstrip antennas on their radiation efficiency and the mode excitation. *IEEE Trans Antennas Propag AP-34*, 8:969–976
4. Petosa A, Ittipiboon A, Gagnon N (1999) Suppression of unwanted probe radiation in wide band probe-fed microstrip patches. *Electron Lett 35(5)*:355–357
5. Huynh T, Lee KF, Lee RQ (1988) Cross-polarization characteristics of rectangular patch antennas. *Electron Lett 24*:463–464
6. Chattopadhyay S, Chakraborty S (2018) A physical insight into the influence of dominant mode of rectangular microstrip antenna on its cross-polarization characteristics and its improvement with T-shaped microstrip antenna. *IEEE Access 6*:3594–3602
7. Li P, Lai HW, Luk KM, Lau KL (2004) A wideband patch antenna with cross-polarization suppression. *IEEE Antennas Wirel Propag Lett 3*:211–214
8. Sim C-Y-D, Sim C-Y-D, Chang C-C, Row J-S (2009) Dual-feed dual-polarized patch antenna with low cross polarization and high isolation. *IEEE Trans Antennas Propag 57*:3321–3324
9. Sarin VP, Nishamol MS, Tony D, Aanandan CK, Mohanan P, Vasudevan K (2011) A wideband stacked offset microstrip antenna with improved gain and low cross polarization. *IEEE Trans Antennas Propag 59*:1376–1379
10. Chen ZN, Chia MYW (2003) Broad-band suspended probe-fed plate antenna with low cross-polarization levels. *IEEE Trans Antennas Propag 51(2)*:345–346
11. Gao S, Li LW, Leong MS, Yeo TS (2003) A broad-band dual-polarized microstrip patch antenna with aperture coupling. *IEEE Trans Antennas Propag 51*:898–900
12. Wong KL, Tang CL, Chiou JY (2002) Broad-band probe-fed patch antenna with a W-shaped ground plane. *IEEE Trans Antennas Propag 50*:827–831
13. Ghosh A, Ghosh D, Chattopadhyay S, Singh LLK (2014) Rectangular microstrip antenna on slot type defected ground for reduced cross polarized radiation. *IEEE Antennas Wirel Propag Lett*. <https://doi.org/10.1109/LAWP.2014.2363563>
14. Ghosh A, Chakraborty S, Chattopadhyay S, Nandi A, Basu B (2016) Rectangular microstrip antenna with dumbbell shaped defected ground structure for improved cross polarized radiation in wide elevation angle and its theoretical analysis. *IET Microw, Antenna Propag 10(1)*:68–78
15. Ghosh A, Basu B (2019) Triangular slotted ground plane: a key to realize high gain, cross-polarization free microstrip antenna with improved bandwidth. *Turk J Electr Eng Comput Sci 27(3)*:1559–1570
16. Ghosh A, Chattopadhyay S, Chakraborty S, Basu B (2017) Cross type defected ground structure integrated microstrip antenna for wide band width and improved polarization purity. *J Electromag Waves Appl 31(5)*:461–476
17. Kumar C, Pasha IM, Guha D (2015) Microstrip patch with non-proximal symmetric defected ground structure (DGS) for improved cross-polarization properties over principal radiation planes. *IEEE Antennas Wirel Propag Lett*. <https://doi.org/10.1109/LAWP.2015.2406772>
18. Lee KF, Guo YX, Hawkins JA, Guo YX, Chair R, Luk KM (2000) Theory and experiment on microstrip patch antennas with shorting walls. *IEE Proc Microw Antennas Propag 147*:521–525
19. Ghosh D, Ghosh SK, Chattopadhyay S, Nandi S, Chakraborty D et al (2014) A physical and quantitative analysis of compact rectangular microstrip antenna with shorted nonradiating edges for reduced cross-polarized radiation using modified cavity model. *IEEE Antennas Propag Mag 56*:61–72

20. Poddar R, Chakraborty S, Chattopadhyay S (2016) Improved cross polarization and broad impedance bandwidth from simple single element shorted rectangular microstrip patch: theory and experiment. *Frequenz* 70:1–9
21. HFSS (2008) High frequency structure simulator, Version 13.0. Ansoft Corp, Pittsburgh

Band Rejection in Wideband Partial Ground Plane Antennas Using Defected Ground Structure



Boyapati Bharathidevi, Jayendra Kumar, and Narayana Rao Palepu

Abstract A dual-wideband performance monopole triangular patch antenna is designed, and for wireless communication, applications are presented. The proposed antenna consists of a triangular patch via a microstrip feed line on the substrate and a dumbbell-shaped defected ground structure (DGS) on the partial ground plane. The proposed antenna operates at a dual band of 2.9–3.3 GHz and 5.2–6.4 GHz, a partial ground with dumbbell-shaped DGS implementation. A DGS with a dumbbell shape is operated at 3.2 and 5.8 GHz. The simulated 10 dB bandwidth for return loss is achieved to be 12.90% (2–9–3.3 GHz) and 20.68% (5.2–6.4) in the lower and upper band, respectively, which covers the bandwidth requirements of 5.2/5.8 GHz WLAN and 5.5 GHz WiMAX application bands. The proposed antenna has a very simple planar structure and occupies a small area of 40 mm × 47 mm. The proposed antenna has a desirable VSWR level and radiation pattern which is suitable for wireless communication applications. The introduction of defected ground structure also improved the antenna performance. The parameter of the proposed antenna is varied and discussed in detail to observe its effect on the antenna performance.

Keywords Defected ground structure · Multiband antenna · Microstrip antenna · Triangular patch · Wideband antenna · Partial ground

1 Introduction

The design of a compact light weight microstrip patch antenna has attracted much attention from the antenna community for applications in modern wireless communication systems. The recent advancements in the mobile and wireless communication system increase the demand for higher data rates and large bandwidth. However,

B. Bharathidevi · J. Kumar (✉) · N. R. Palepu
School of Electronics Engineering, VIT-AP University, Amaravathi, Andhra Pradesh 522237,
India
e-mail: jayendra854330@gmail.com

N. R. Palepu
e-mail: narayana.20phd7073@vitap.ac.in

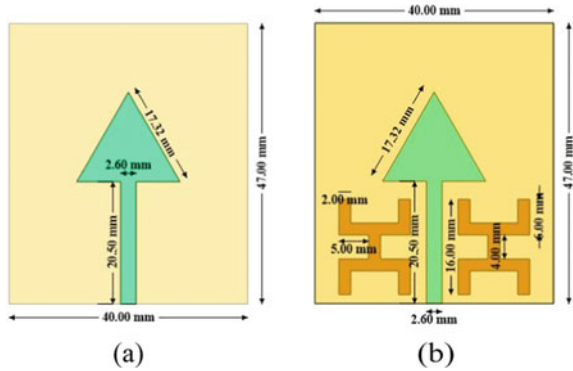
the conventional microstrip antenna suffers from narrow bandwidth problems. The researchers of the antenna community have investigated and proposed various design techniques to enhance the bandwidth of the conventional microstrip patch antenna. Wireless local area network (WLAN) and worldwide interoperability for microwave access (WiMAX) are widely used in modern wireless communication technology [1]. In general, a wireless communication antenna is required to be available to cover a wide frequency bandwidth or several frequency bands [2]. Antennas play a powerful role and are necessary for a good communication setup that does not include a wired cable. In recent years, the demand for small antennas operating at low frequency is increasing; to achieve this, some of the miniaturizing techniques are used [3]. Placing the slots or defects on the ground plane of microwave planner is referred to as defected ground structure (DGS) [4]. DGS is an improving technique for compactness of antenna size, enhancing the bandwidth, and improvement in the gain of the antenna [5]. Different types of defected ground structures are used to attain the wide bandwidth as ohm shaped and fractal structures are implemented [6, 7]. Band-notched filters are also used to operate the antenna to a particular application [8]. Monopole antenna also attained the triple-band for different wireless applications [9, 10]. For cross-polarization suppression, a dumbbell-shaped defective ground structure is implemented [11, 12]. It will also improve the return loss and the antenna performance, the main application is, DGS techniques are used to improve the various parameters of antenna-like narrow bandwidth, low gain and cross-polarization. To obtain the function of unwanted frequency rejection and circuit size reduction [13]. To the low cross-polarization and isolation reduction, two rectangular slots are placed on the ground plane [14].

To attain the wideband, a partial ground plane is implemented [15]. In this paper, a novel dual-band microstrip-fed monopole antenna with the partial ground is presented. With DGS on the partial ground plane, the antenna is able to generate dual impedance bandwidths to cover the frequency bands of the WLAN (5.2/5.8 GHz) and the WiMAX (5.5 GHz), and along with that, the upper band of the S-band is also covering. Wide bandwidth is obtained because of the partial ground plane, and band rejection is accomplished by placing the dumbbell-shaped DGS on the partial ground plane.

2 Antenna Design

The basic structure of the microstrip patch antenna is represented in Fig. 1a and consists of the triangular patch, dielectric substrate and a ground plane. The patch is a conducting material, and it has a triangular shape. The substrate consists of a dielectric material having the dielectric constant of 4.4, FR4 material. This antenna is fed by a line feeding. The other side of the substrate is a ground plane. The dimensions of the antenna are shown clearly in Fig. 1a. An antenna is designed with dielectric substrate 4.4 and a thickness of 1.6 mm. Also, the ground plane has a width of 40 mm and a height of 47 mm. The antenna is excited using a 50 Ω direct microstrip line

Fig. 1 Antenna layout **a** Reference monopole antenna layout, **b** DGS integrated antenna with full ground plan



feeding. One identical slot of width 2.6 mm and height 20.5 mm was combined with the triangular patch and placed symmetrically to the edge of the radiating patch. This type of structure is useful for integration with any microwave component circuit. The resonating frequency of this monopole antenna is at 8 GHz. By placing DGS on the ground plane or implementing the partial ground bandwidth of the antenna can be improved. A normal monopole antenna was designed initially as mentioned in the dimensions in Fig. 1a. It is resonating at a frequency of 8 GHz. The triangular patch is fed through the microstrip feed line.

A dumbbell-shaped defective ground structure is implemented in Fig. 1b. It is resonating at a frequency of 5 GHz. This will apply to the WLAN applications. The dimensions of the DGS are mentioned clearly in Fig. 1b. To get a better wide bandwidth, partial ground is implemented. A dumbbell-shaped defected ground structure is placed on the ground plane. It is resonating at multiple frequencies. The combination of a dumbbell-shaped DGS and a partial ground is used to improve the bandwidth and the return loss for the desired antenna. The partial ground is implemented to the length of 18 mm, and the width is 47 mm. Partial ground implementation is shown in Figs. 2 and 3. This will be useful to wireless communication applications.

Fig. 2 Partial ground plane antenna **a** Top view, **b** bottom view

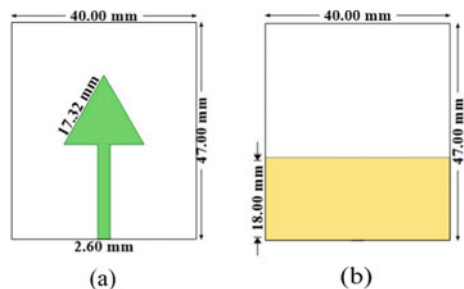
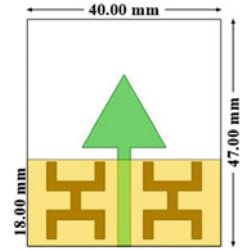


Fig. 3 DGS integrated partial ground plane antenna



2.1 Parametric Analysis

Here, parametric analysis was performed on the ground plane. Figure 4 shows the effect of changing the length of the ground plane. When the length of the ground plane is 15 mm, the antenna is resonating at a single frequency, i.e. 2.9 GHz with a return loss of -11 dB. When the length is increased to 16 mm, the antenna is resonating at two frequencies 3.1 GHz and 5.7 GHz with a return loss of -13.45 dB and -12.03 dB. When the length is increased to 17 mm, in this case, significant result in the return loss as well as bandwidth improvement. In this condition, the antenna is resonating at 3.3 and 5.7 GHz with a return loss of -14.8 dB and -21.49 dB. When the length is 18 mm, in this condition, wide bandwidth is obtained (3.1–6.5 GHz) and resonating at a frequency of 5.5 GHz with the return loss of -28.45 dB. When

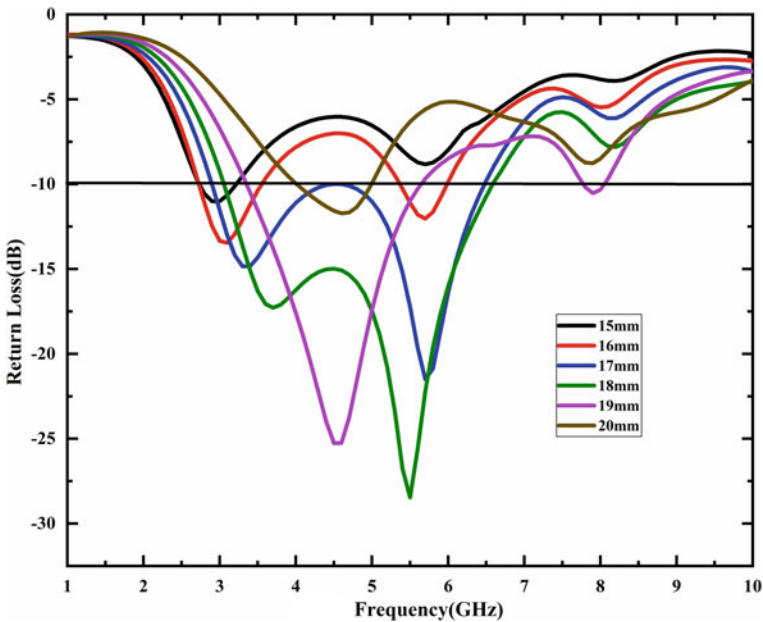


Fig. 4 Scattering parameter of the partial ground plane antenna versus length of the ground plane

the length is 19 mm, antenna is resonating at a single frequency at 4.5 GHz with the return loss of -25.26 dB. When the length is increased to 20 mm, antenna is resonating at 4.6 GHz with the return loss of -11.71 dB. Compared to the remaining four conditions at the length of 18 mm, good bandwidth and return loss are obtained.

3 Results and Discussion

Simulated return loss plots are indicated in Fig. 5. From the figure, it is observed that the designed antenna is resonating at the frequency of 8 and 5 GHz with a return loss of -18.4 dB and -20.44 dB. -10 dB impedance bandwidth of the designed antenna ranges from 7.9–8.1 to the conventional microstrip patch antenna without DGS and 4.9–5.1 with DGS.

When compared with the conventional monopole antenna without DGS, the designed antenna with DGS shows better results. In order to enhance the bandwidth of the triangular patch antenna, a partial ground plane is employed in Figs. 3 and 4. The incorporation of a dumbbell-shaped DGS and partial ground plane reduced the resonant frequency to 3.2 and 5.8 GHz having the improved reflection coefficient value of -20.39 dB and -22.62 dB with IBW of 12.90% (2.9–3.3 GHz) and 20.68% (5.2–6.4 GHz). In Fig. 6, to get the wider bandwidth, partial ground technique is implemented in the design. Due to the reduction in capacitance in turn

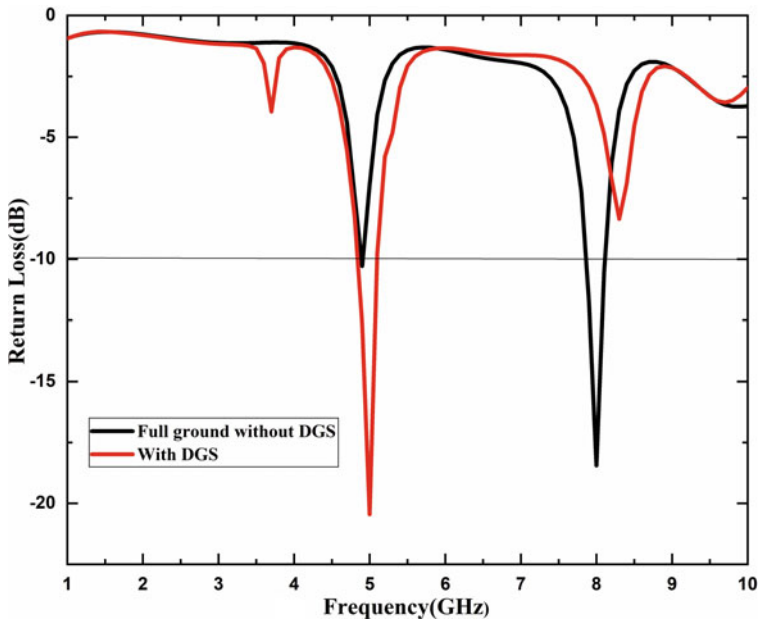


Fig. 5 Scattering parameter of the conventional triangular patch antenna with and without DGS

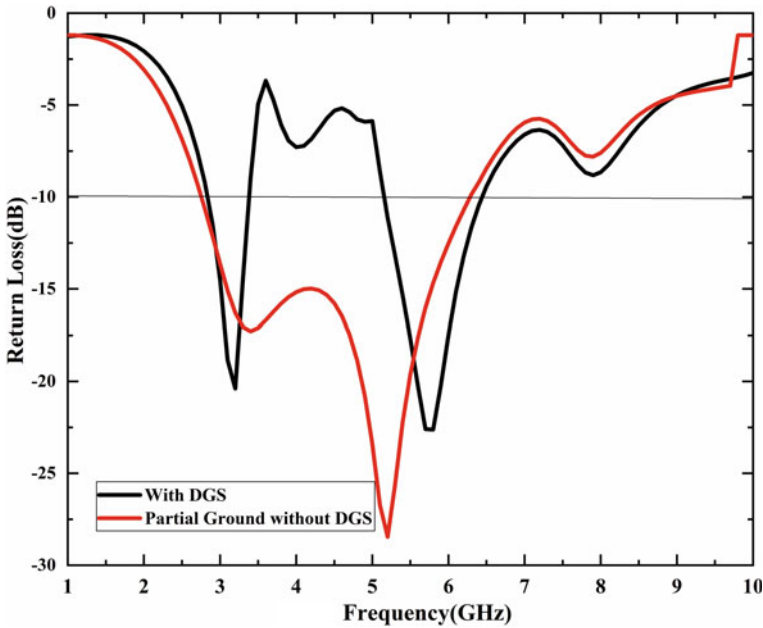


Fig. 6 Scattering parameter of the partial ground plane antenna with and without DGS

reduction in quality factor. The advantages associated with the insertion of the DGS and partial ground implementation is, reflection coefficient and impedance matching of the proposed antenna is much better compared to without DGS and conventional triangular patch antenna. The proposed antenna shows significantly improved bandwidth, and wide bandwidth is clearly observed. Two resonant frequencies are used for WiMAX/WLAN band. Antenna without DGS operates at 3.1–6.6 GHz. It is used for wideband applications. When compared to the conventional monopole antenna with the full ground, by the implementation of the partial ground, significant results are obtained and bandwidth is improved.

Figure 7 shows that the simulated VSWR result for the designed antenna is less than 2 dB. It will fulfil the bandwidth for the requirement to the required frequency range. Better VSWR value is obtained to the partial ground with DGS compared to the remaining conditions. The very low value of VSWR signifies very less mismatch loss thus better impedance matching throughout the operating bandwidth of the proposed antenna. The conventional antenna has a value of VSWR less than 2 that is from 4.5 to 5 GHz, whereas the proposed antenna maintains the proper value of VSWR to achieve significant impedance matching. The simulated radiation pattern of the E-plane and H-plane at 5.5 GHz is represented in Fig. 8. The radiation patterns are omnidirectional to the co-pol and bidirectional pattern to the cross-pol in E-plane. In the E-plane, a bidirectional radiation pattern is attained because of the monopole antenna. The radiation patterns show that in the H-plane the antenna has a nearly omnidirectional radiation pattern, while the E-plane radiation pattern shows a typical

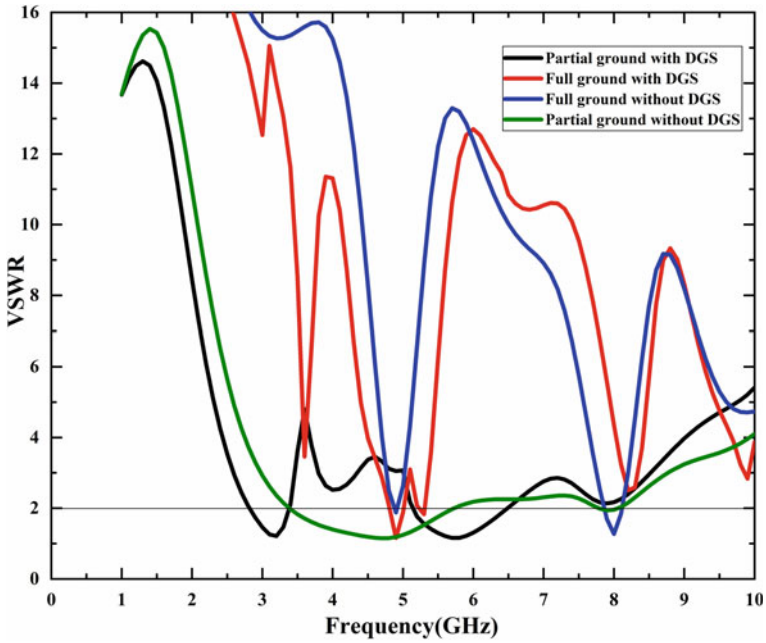


Fig. 7 VSWR of different antenna configurations

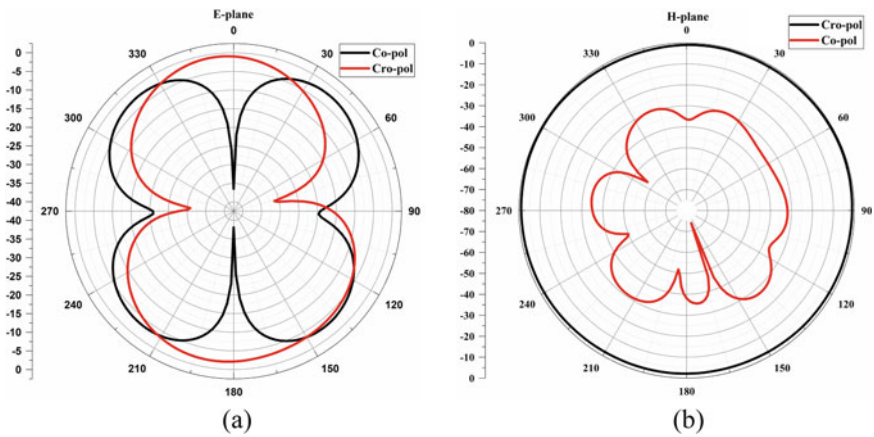


Fig. 8 Radiation pattern of the DGS integrated partial ground plane antenna at 5.5 GHz a E-plane b H-plane

figure of eight at the frequency of 5.5 GHz, which is similar to a conventional dipole antenna. The modifications in the size of the ground plane result in a decrease in inductance value, and hence, there is a change in the resonant frequency, and in turn, the improved radiation pattern is obtained.

4 Conclusion

A dumbbell-shaped DGS is proposed in this paper to the partial ground and full ground of microstrip patch antenna for WLAN, WiMAX and wideband applications. The structure of the proposed antenna is very simple with a small area of 40 mm × 47 mm. The proposed antenna is small in size and provides dual-band impedance bandwidth. The proposed antenna parameters have been simulated, and parametric analysis concerning ground is performed. The proposed antenna resonates at 3.2 and 5.5 GHz with a dual band of % (2.9–3.3 GHz) to the lower band and % (5.2–6.4 GHz) to the upper band. Resonant frequencies and VSWR have also been studied. Return loss and radiation patterns allow the antenna for WLAN and WiMAX applications.

References

1. Wu T, Shi X-W, Li P (2013) Bai H (2013) Tri-band microstrip-fed monopole antenna with dual-polarisation characteristics for WLAN and WiMAX applications. *Electron Lett* 49:1597–1598
2. Kumar J (2016) Compact MIMO antenna. *Micro Opt Technol Lett* 58:1294–1298
3. Chakraborty M, Rana B, Sarkar PP, Das A (2012) Design and analysis of a compact rectangular microstrip antenna with slots using defective ground structure. *Procedia Technol* pp 411–416. <https://doi.org/10.1016/j.protcy.2012.05.064>
4. Richards WF, Hanapi KM, Abdul Rahim SK, Saad BM, Rani MSA, Aziz MZA (2014) An elliptically planar UWB monopole antenna with step slots defective ground structure. *Micro Opt Technol Lett* 56:2084–2088
5. Sheeba R, Jegan Antony Mariciln L, Velan B, Sugadev M (2020) Design and analysis of energy band gap and defective ground structure on array of patch antenna using meta material. *Int Conf Smart Syst Inv Technol*. <https://doi.org/10.1109/ICSSIT48917.2020.9214150>
6. Salamin MA, Das S, Zugari A (2018) Design and realization of low profile dual-wideband monopole antenna incorporating a novel ohm shaped DMS and semi-circular DGS for wireless applications. *Int J Electron Commun* 97:45–53
7. Banerjee U, Karmakar A, Saha A (2019) and pilyali Chakraborty: A CPW-fed compact monopole antenna with defected ground structure and modified parasitic Hilbert strip having wideband circular polarization. *Int J Electron Commun*. <https://doi.org/10.1016/j.aeue.2019.152831>
8. Lakrit S, Das S, El Alami A, Barad D, Mohapatra S (2019) A compact UWB monopole patch antenna with recon- figurable band-notched characteristics for Wi-MAX and WLAN applications. *Int J Electron Commun* 9. <https://doi.org/10.1016/j.aeue.2019.04.001>
9. Li L, Zhang X, Yin X, Zhou L (2016) A compact triple-band printed monopole antenna for WLAN/WiMAX applications. *IEEE Trans Antennas Propag* 15:68–78
10. Yadav A, Verma S (2019) Compact monopole antenna with modified triangular shape for WLAN, WiMAX and Wi-Fi applications. *Int Conf Con Technol*

11. Abhijyoti G, Subhradeep C, Sudipta C, Arnab N, Banani B (2016) Rectangular microstrip antenna with dumbbell shaped defected ground structure for improved cross polarised radiation in wide elevation angle and its theoretical analysis. *IET Micro Antennas Propag* 10:68–78
12. Debatosh G, Manotosh B, Yahia MMA (2005) Microstrip patch antenna with defected ground structure for cross polarization suppression. *IEEE Antennas Wirel Propag Lett* 4:455–458
13. Balanis CA (2005) *Antenna theory analysis and design*, 3rd edn. Wiley, New Jersey, USA
14. Kumar J, Basu B, Talukdar FA, Nandi A (2018) Multimode-inspired low cross-polarization multiband antenna fabricated using graphene-based conductive ink. *IEEE Antennas Wirel Propag Lett* 17:1861–1865
15. Kumar J, Shirgan SS, Patil DB (2014) Miniature wideband 1×2 micro-strip antenna for 4G application. *IEEE Con Comput Netw*

Triple-Band Polarization Independent C-Band Metamaterial Absorber



Kashish Mahindroo, Vani Sadadiwala, Vimlesh Singh, Devender Sharma, and Sarthak Singhal

Abstract A triple-band metamaterial absorber is proposed and discussed in paper for C-band application. The proposed absorber surface is simulated for FR4 substrate by CST microwave studio electromagnetic simulator. The unit cell of metamaterial structure resonance appears at frequency of 4 GHz (−14 dB), 4.7 GHz (−13.5 dB), and 4.9 GHz (−26 dB). Absorptivity for simulated design is 96% for 4 GHz, 95% for 4.7 GHz, and 99.8% for 4.9 GHz. Electromagnetic microwave radar/absorber plays a vital role in military and defense techniques as well. This type of absorber also works for cloaking and shielding of drones which are used in defense sectors.

Keywords Metamaterial · Absorptivity · Polarization · Independent · Triple band · RCS · Drone

1 Introduction

Absorbers are the metamaterial structures that are designed to absorb different frequencies and wavelengths of different electromagnetic waves. These metamaterial structures are the structured unique material designed to absorb the highly exotic narrow bandwidth of tunable frequency range for very high accuracy, and these metamaterials are designed to form optical geometry which has less loss and more absorptivity [1–4]. The paper gives a glimpse of the general theory of the

K. Mahindroo · V. Sadadiwala (✉) · D. Sharma
Aeronautical Engineering, FET, MRIIRS, Faridabad, India
e-mail: vanisadadiwala82@gmail.com

D. Sharma
e-mail: devendersharma.fet@mriu.edu.in

V. Singh
Electrical & Electronics & Communication Engineering, FET, MRIIRS, Faridabad, India
e-mail: vimlesh.fet@mriu.edu.in

S. Singhal
Electronics & Communication Engineering, MNIT, Jaipur, India
e-mail: sarthak.ece@mnit.ac.in

electromagnetic wave absorber and the survey of metamaterials that are suitable to design the absorber which is affordable, available, and feasible. It also talks about RAM, i.e., the radar absorbing materials which are more favorable in designing the absorbers of good frequencies. The main aim of the paper is to focus more upon the designs of the absorber with wide bandwidth, generally, the absorber which is capable enough to overcome and clamp down the different frequencies of reflection, transmission, and signals of unwanted EM waves to the lowest grade of detection and fulfill the purpose of cloaking, absorbing, and shielding.

The drones, majorly an autopilot autonomous system, are developed for military and defense sectors. On contrary, misuse of this autonomous system is seen on a large scale including entry in restricted areas of terror. For military safety and concerns, this absorber can work well for the shielding and protection of drones from the enemy radar systems. This method is collectively known as antiterrorism technique. To minimize the detection of drones from enemy system, radar cross section (RCS) is needed to be minimized. This reduction in RCS is done using metamaterial absorbers [5, 6].

Absorbers are the major application of these metamaterials as the work on the theory of interference. However, the magnetic and the electric resonance of metamaterial limits the behavior of electromagnetic wave. The requirement of wideband comes into frame and this depends on the thickness and frequency range; as for small frequencies, large band gaps are reached easily, while for large frequencies, the width should be minimized; and the most efficient result is obtained when all the resonant are combined together in multiple layer fashion [7–10]. Lossy material is usually used in enhancing the bandwidth of the absorber. Different sized patches are used for better results. Ring and plate patches are the more effective patterns for patches as they can absorb a very high range of EM as well as radio waves. Basically, the motive is to create a perfect ranging metamaterial [11].

2 Theoretical Approach of Absorber

The nature of metamaterial depends upon the structure, size, and geometry it is embedded with rather than its constituent materials. It is prominently seen in many designs that the presence of metamaterial affects the permittivity as well the permeability of the structure and the technical design in which they are used. The metamaterial becomes more interesting because they exhibit the double negative behavior, negative refractive index, left-handed behavior, and much more. Metamaterial absorbers are the highly satisfying absorbers as they are capable of converting the EM waves into heat [12–16].

Metamaterial absorbers get activated with electric field, and magnetic response is generated via resonant absorption; this results in the generation of current and the anti-parallel currents. These metamaterials have evolved so far that the absorbers designed nowadays generate microwave optical bands. In most of the designs, the

range of EMI wave is unknown resulting in a broadband of frequencies indicating the polarization occurs from everywhere.

Achieving a perfectly absorbing EM band is impossible, so if one obtains maximum 4 wide frequency absorber bands using resonators, then such absorbers are considered desirable for working on stealth technology as they indicate the minimum transmission of EM waves and their reflection [17–22].

Sometimes, it is possible to achieve 5–6 peaks of absorption in same graph but they have very narrow width which becomes negligible while designing the actual patch of model. Even in multilayer absorbers, it is tough to maximize the bandwidth when multiple peaks are generated. Recent updates in nanotechnology have somehow solved these problems of absorption in the absorbers as it has introduced the concepts of micro-designing resonating structures with simple patterns. Many composite structures are introduced to achieve a wideband resonance. For desirable band, selection of structures has been improvised by selecting the appropriate frequency selective surface with suitable conductive patterns on selected dielectric medium. However, such absorbers are quite bulky and breakable and have high mass density but still, they possess numerous advantages as they have thick bandwidth and they provide quality for reducing the reflection.

3 Role of Bandwidth and Thickness in Metamaterial Absorber

The absorption of metamaterial absorbers depends upon the electric and the magnetic resonance behavior; it is a major reason for its limiting bandwidth. In most of their working fields, we require higher bandwidth to work with. Achieving the wider bandwidth is the challenging task as it requires working with the thickness of metamaterial. EM wave absorbers are working with the absorption of the EM waves present, and the rest of the waves are reflected. One of the most effective ways to obtain the effective wider range is to design multiple resonance units in a single unit in a multiple layer fashion, and as a result, the multiple phenomena give rise to better absorption. However, the fabrication techniques become quite complicated while designing such designs, and maintenance of placements of each cell is very critical for achieving good frequency bandwidth [23–25].

However, multilayering fashion provides a good compromise between the frequency range and absorption of electromagnetic absorbers by merging the layers of resonance and this improves the resulting quality and provides a broader bandwidth.

EM absorber could work much better if there are no restrictions on the thickness of material as they are capable to absorb maximum EM waves. The wide use of EM absorber is seen in radio wave absorption as it works better in the RF spectrum and provides better results there. A part of multilayering absorbers FSS absorbers are also used in improving the results. They are the simple periodically arranged structures and dielectrics with more reasonable properties and frequency range. They have

fractal, ring designed patches on the metamaterial surface which help in increasing the rate of absorption.

Circuit analog absorber (CAA) is the another approach of obtaining broader bandwidth. These absorbers use chip resistors but the only challenges are that one needs to select the appropriate place to fix the chip so that a wide range can be obtained. These structures have loops on high impedance surface, and both the layers are separated by the air gap. The air gap is not responsible for increasing the broadness of the absorber. The performance factor of CAA depends upon the angle of incidence of the EM wave [10, 26–28].

4 Structural Design of Unit Cell of Proposed Meta Surface

A triple-band metasurface absorber for C-band application has been designed in this paper. Design of simulated unit structure is shown in Fig. 1 and has metallic top of thickness 0.038 with conductivity of copper material in upper and lower layer of 5.8×10^7 S/m. Proposed metasurface is simulated for FR4 dielectric substrate with relative permittivity $\epsilon_r = 4.4$, $Z_{top} = 1.6$ mm dielectric thickness and tangent loss $\tan = 0.02$.

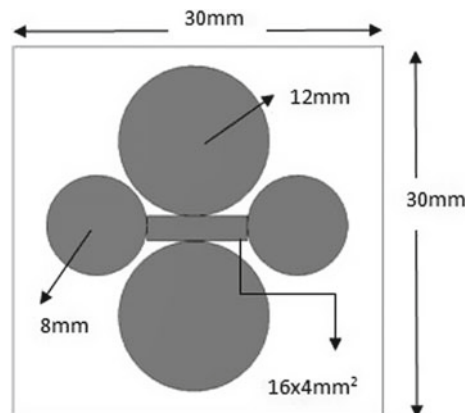
The ground of proposed metasurface unit cell is covered with copper so power transmitted $|S_{21}|^2$ from the ground plane is zero. For angular frequency w absorptivity $A(w)$, reflected power $S_{11}(w)$ and transmitted power $S_{21}(w)$ is formulated as [19, 29]:

$$A(w) = 1 - |S_{11}(w)|^2 - |S_{21}(w)|^2$$

For no power transmitted from the ground surface, $|S_{21}| = 0$.

So,

Fig. 1 Proposed metasurface unit cell design



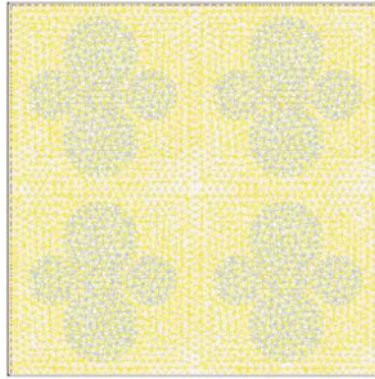


Fig. 2 Mesh view of 2×2 metasurface

$$A(w) = 1 - |S_{11}(w)|^2$$

For minimum reflection, absorption rate is high. Lossy nature of dielectric layer and resistivity of metallic surface absorb incident electromagnetic waves in metasurface. Figure 2 shows mesh view of 2×2 metasurface.

5 Result of Proposed Metasurface

The unit cell of metastructure is simulated by CST microwave studio 2014 version [30].

Reflection coefficient S_{11} and absorption $A(w)$ characteristic for unit cell metasurface are shown in Figs. 3 and 4. Reflection coefficient at 4 GHz is -14 dB, 4.7 GHz

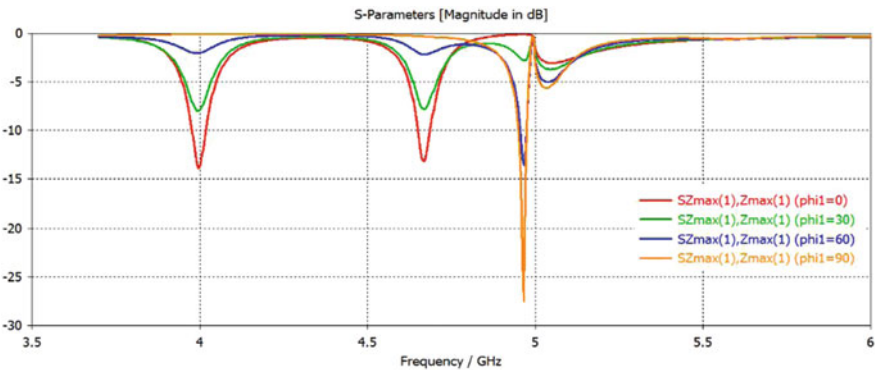


Fig. 3 Reflection coefficient for simulated metasurface for 0° , 30° , 60° , and 90°

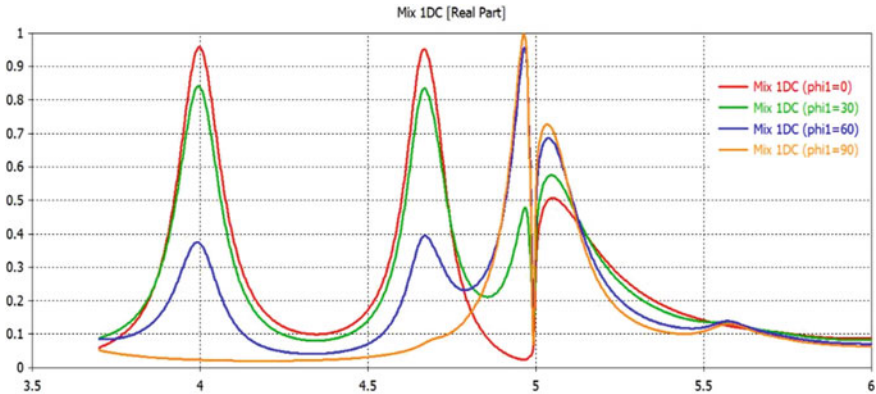


Fig. 4 Simulated absorptivity at 0°, 30°, 60°, and 90° metasurface

is -13.5 dB, and 4.9 GHz is -26 dB. Absorption for simulated design is 96% for 4 GHz, 95% for 4.7 GHz, and 99.8% for 4.9 GHz.

6 Conclusion

A triple-band metamaterial absorber has been presented in paper for C-band application. Designed absorber shows more than 95% absorptivity for three frequencies of C-band. The proposed structure is simulated for 0°, 30°, 60°, and 90° phase angle. The paper has a clear vision of presenting considerable absorbers which can be further used in the military for the purpose of defense and protection. The proposed metamaterial surface can be used in reduction in size of radar cross section, electromagnetic interference, shielding, defense, security, and other applications. The purpose is to design the absorber which can withstand the thin technology and can give better results using the accessible amount of metamaterial which is present in abundance and can be extracted easily and economically.

References

1. Caloz C, Itoh T (2005) Electromagnetic metamaterials: transmission line theory and microwave applications. John Wiley & Sons, Hoboken, NJ
2. Munk BA (2005) Frequency selective surfaces: theory and design. John Wiley & Sons, Hoboken, NJ
3. Munk BA (2009) Metamaterials: critique and alternatives. John Wiley & Sons, Hoboken, NJ
4. Capolino F (2009) Applications of metamaterials. CRC Press, Boca Raton, FL
5. Pisa S, Piuzzi E, Pittella E, Lombardo P, Genovese A, Bloisi D, Nardi D, d'Atanasio P, Zambotti A Numerical and experimental evaluation of the radar cross section of a drone

6. Farlik J, Kratky M, Casar J, Stary V (2016) Multispectral detection of commercial unmanned aerial vehicles. MDPI in 2019
7. Bhattacharyya S, Ghosh S, Vaibhav Srivastava K (2013) Triple band polarization-independent metamaterial absorber with bandwidth enhancement at X-band. *J Appl Phys* 114:094514
8. Ayop OB, Abd Rahim MK, Murad NA, Samsuri NA, Dewan R (2014) Triple band circular ring-shaped metamaterial absorber for X-band applications. *Prog Electromag Res M* 39:65–75
9. Wang BX, Wang LL, Wang GZ, Huang WQ, Li XF, Zhai X (2014) Theoretical investigation of broadband and wide-angle terahertz meta-material absorber. *IEEE Photonics Technol Lett* 26:111–114
10. Jang T, Youn H, Shin YJ, Guo LJ (2014) Transparent and flexible polarization-independent microwave broadband absorber. *ACS Photonics* 1:279–284
11. Holloway CL, Kuester EF, Gordon JA, O'Hara J, Booth J, Smith DR (2012) An overview of the theory and applications of metasurfaces: the two-dimensional equivalents of metamaterials. *IEEE Antennas Propag Mag* 54:10–35
12. Li M, Yang HL, Hou XW, Tian Y, Hou DY (2010) Perfect meta-material absorber with dual bands. *Prog Electromag Res* 108:37–49
13. Cheng Y, Yang H, Cheng Z, Wu N (2011) Perfect metamaterial absorber based on a split-ring-cross resonator. *Appl Phys A: Mater* 102:99–103
14. Liu X, Tyler T, Starr T, Starr AF, Jokerst NM, Padilla WJ (2011) Taming the blackbody with infrared metamaterials as selective thermal emitters. *Phys Rev Lett* 107:045901
15. Hajizadegan M, Ahmadi V, Sakhdari M (2013) Design and analysis of ultrafast and tunable all optical metamaterial switch enhanced by metal nanocomposite. *J Lightwave Technol* 31:1877–1883
16. Kildishev AV, Boltasseva A, Shalaev VM (2013) Planar photonics with metasurfaces. *Science* 339:1232009
17. Panwar R, Puthucheri S, Agarwala V, Singh D (2015) Fractal frequency-selective surface embedded thin broadband microwave absorber coatings using heterogeneous composites. *IEEE Trans Microw Theory Technol* 63:2438–2448
18. Soheilifar M, Sadeghzadeh R (2015) Design, fabrication and characterization of stacked layers planar broadband metamaterial absorber at microwave frequency. *AEU-Int J Electron Commun* 69:126–132
19. Ghosh S, Bhattacharyya S, Chaurasiya D, Srivastava KV (2015) Anultrawideband ultrathin metamaterial absorber based on circular split rings. *IEEE Antennas Wirel Propag Lett* 14:1172–1175
20. Montaser AM (2016) Design of metamaterial absorber for all bands from microwave to terahertz ranges. *Int. J. Adv. Res. Electron. Commun. Eng.* 5:1475–1481
21. Zuo W, Yang Y, He X, Mao C, Liu T (2017) An ultra wideband miniaturized metamaterial absorber in the ultrahigh-frequency range. *IEEE Antennas Wirel Propag Lett* 16:928–931
22. Ahmadi F, Idab N (2017) A broadband ultrathin metamaterial absorber using tilted parallel strips. In *Proc SPIE* 10103:101031V-V101041
23. Sood D, Tripathi CC (2016) Broadband ultrathin low-profile meta-aterial microwave absorber. *Appl Phys A* 122:332
24. Ozden K, Yucedag OM, Kocer H (2016) Metamaterial based broadband RF absorber at X-band. *AEU-Int J Electron Commun* 70:1062–1070
25. Lee D, Kim HK, Lim S (2017) Textile metamaterial absorber using screen printed channel logo. *Microw Opt Technol Lett* 59:1424–1427
26. Munaga P, Ghosh S, Bhattacharyya S, Srivastava KV (2016) A fractal-based compact broadband polarization insensitive metamaterial absorber using lumped resistors. *Microw Opt Technol Lett* 58:343–347
27. Sen G, Islam SN, Banerjee A, Das S (2017) Broadband perfect metamaterial absorber on thin substrate for X-band and Ku-band applications. *Prog Electromagn Res C* 73:9–16
28. Vu DQ, Le DH, Dinh HT, Trinh TG, Yue L, Le DT, Vu DL (2018) Broadening the absorption bandwidth of metamaterial absorber by coupling three dipole resonances. *Physica B* 534:90–94

29. Bhattacharyya S, Ghosh S, Chaurasiya D, Srivastava KV (2015) Wide-angle broadband microwave metamaterial absorber with octave bandwidth. *IET Microw Antennas Propag* 9:1160–1166
30. CST Microwave Studio Suite 2014, CST Inc., 2007

A Study on the Ionospheric Propagation Using GNSS Receiver Over Hill Region



Sankha Subhra Debnath and Swastika Chakraborty

Abstract The Global Navigation Satellite System (GNSS) has been a very integral part of our modern-day communications and navigation. Initially, it was limited for military systems only but now with involvement of both commercial and civilian access to the data, the possibility of development for new applications is endless. So far only there has been a very limited number of Global Systems and a few regional systems which are likely to change in the coming days as more countries are looking to make their presence in earth's orbit. The applications of GNSS are helping enormously for mapping, surveying, precise positioning, ionosphere research, earth data monitoring and more. But with the newly emerging trends and innovations in the commercial and research and development by so many private enterprises, the need for GNSS will be greater than ever before. More devices will need the use of GNSS to integrate for various applications. The fast-paced innovation has led to the democratization of the system to a large extent. With more development and adoptions of next generation autonomous technology such as driverless cars, automatic traffic management system and more access to high-speed global network, GNSS will play and will have more impact on everyday life in the coming years. Thus, it is important to start finding use cases that can have a large-scale impact. This paper focuses specifically on the use case of GNSS for ground-based measurements that can be further developed for more niche applications. The GNSS receiver and the data used in the study are located at Sikkim Manipal Institute of Technology, Sikkim, India. The study presented here is to show that the system is viable and can be developed into products with very important use cases and applications that can be of interests for both commercial and military uses.

Keywords GNSS · GPS · RINEX · RTKLIB

S. S. Debnath
Agartala, West Tripura 799007, India

S. Chakraborty (✉)
ECE Department, Narula Institute of Technology, Kolkata 700109, India
e-mail: swastika1971@gmail.com

1 Introduction

Global Navigation Satellite System (GNSS) refers to any satellite constellations that can perform and sending positioning, navigation and timing, metrics and services both on a regional and global basis. The GNSS systems are now one of the most important systems for global communications, earth observations, data collecting and monitoring; currently, we have a handful of major countries who have their own GNSS operating that is fulfilling the global requirement of the applications in operation currently, but it is changing rapidly as more countries are looking forward to making their presence known in earth orbits and space in general. Currently, we have IRNSS/NavIC (Indian Regional Navigation Satellite System) of India, QZSS (Quasi-Zenith Satellite System) of Japan, GLONASS (Global Navigation Satellite System) of Russia, Galileo by European Union, DS (Bei Dou Navigation Satellite System) of People's Republic of China and GPS (Global Positioning System) by USA. GNSS is extensively used for many ground-based measurements as well as Precise Point Positioning (PPP) solutions. This paper focuses by approaching the ground-based measurements of GNSS systems and the further developments of complex PPP solutions following various future trends and applications.

Most of the applications with GNSS was for military and government uses but now with more and more new commercial products that are being developed with GNSS related services, many data and information have been made available for public and commercial usage. It has become much easier now for research and developments at academic level or for commercial use. Thus, there is a very big scope of many new applications that can be developed. Currently, the continuous need of PNT (positioning, navigation and timing) services has played a major role in bringing down the costs commercially and is helping for continuous improvement of the user experiences. Thus, more 'integration of GNSS with related services can help to create' instead of 'more integration of GNSS and related services can help' create and develop applications that can be very significant for government, military and other commercial applications.

2 Literature Surveys

To undergo the development of the study, the fundamentals of GNSS receiver systems and its operation have been studied thoroughly. The study is region focussed; thus, the observation for the system is also analysed thoroughly by understanding existing IGS data and post-processing.

In this study, we are focussed on the ground-based measurements of our system. Literature is there [1] to understand how ground-based systems can be developed with GNSS systems. Real-Time Kinematics Library (RTKLIB), [2] an open-source

programming package for GNSS positioning, has been extensively used for post-processing and analysing procedures for the development of the low-cost RTK-GPS receiver.

The study and the development of the project have been carried out by analysing and understanding the current GNSS systems and operations and how new applications and the trends are growing in the specific sector.

In this study, we are focussed on the ground-based measurements of our system. This paper on [1] ‘Development of approach procedure design criteria for systems based on Global Navigation Satellite System (GNSS) with Ground Based Augmentation System (GBAS)’ has been studied to understand how ground-based systems can be developed with GNSS systems.

‘Development of the low-cost RTK-GPS receiver with an open-source program package RTKLIB’ [2]: this paper has been an important study for our work since we have extensively used RTKLIB for our post-processing and analysing procedures.

Another paper that has been studied for understanding the principles behind the ground-based measuring applications is ‘On the potential of GPS-InSAR combination to improve the accuracy of ground deformation monitoring: simulation-based validation’ [3].

For understanding more principles of operations and its working, the paper ‘Assessment of Dual Frequency GNSS Observations from a Xiaomi Mi 8 Android Smartphone and Positioning Performance Analysis’ [4] has provided great insight.

One of the major problems to be solved is to reduce the multipath errors associated with any GNSS systems; for instance, any ground-based measurements suffer from multipath errors. This paper ‘Multipath errors and mitigation techniques’ [5] has been studied thoroughly to understand more in-depth problems of the multipath that affects GNSS systems.

To understand and analyse more about the current trends and perspectives on why it is important for the development of new applications based on satellite navigations, this paper has been very helpful ‘Status, perspectives and trends of satellite navigation’ [6]

3 Data

In this study, a high-precision GNSS module (model number ZED-F9P) developed by a u-blox have been used for the receiver system. It is a multiband receiver that can deliver accuracy at centimetre precision within seconds. The geographical location of where the receiver is set up is situated at Sikkim Manipal Institute of Technology from the department of electronics and communication engineering. Many datasets were observed and collected for a span of 5–6 months. In this study, we sampled out the data of 31st March, 2021 for specific reasons such as very clear weather and consistent sample data observations throughout the day.

The study also requires using data from the International GNSS service (IGS), Crustal Dynamics and Data Information System (CDDIS) and NASA’s Archive of

Space Geodesy Data. The data for the study focussed on this paper ranged for a specific period and date. We at present have used the IGS data of IISC, Bangalore, India (13.02 N-Lat, 77.57 E-Lon) specifically in our study for more accurate results. IGS data of IISC has been used specifically as it is one of the nearest to the original receiver used in the study and thus will give us more accurate observations.

For this study, the data used from both the GNSS receiver and the CDDIS archive (https://cddis.nasa.gov/Data_and_Derived_Products/GNSS/daily_30second_data.html) is 31st March, 2021 and the time frame along which the post-processing has been done for further plotting is also kept the same.

4 Methodology

The study conducted is region based specifically for hill region. More observation and interpretation of the data from GNSS and GPS systems can be explored for a successful implementation of different types of real-time applications. The first section of the study focussed on receiving raw data from the GNSS receiver of the satellite consisting of high accuracy geolocation data and signal. The received raw data is thus converted into Receiver Independent Exchange format (RINEX) where we have the observation files and navigation files. The raw receiver data has the '.ubx' format which have the log files of the data and can be visualized with u-blox for much detailed information.

The raw data must be in RINEX format for post-processing. The raw data is converted to the RINEX format where RTKLIB, an open-source program package for GNSS positioning, is used. We can generate the observation (.obs) and navigation files (.nav) which will enable us for post-processing and analysis of our raw GNSS received data. RTKLIB is used further for post-processing of our observation and navigation files. Another positioning file (.pos) will be generated. The plotting of the post-processed RINEX files is possible now which enables us to get the desired results for our raw GNSS data.

In the next section of the study, the obtained RINEX files are used which is provided by the CDDIS of NASA. This data can be obtained freely from their websites. With the CDDIS data, the observation data and positioning data can be obtained for the required date, time and region. To get more accurate and real estimation, the data is taken from the nearest geographical stations to the study region. In this case, the IGS data from IISC, India, has been used extensively for all the different factors. The post-processing of the obtained IGS data can be performed with the help of RTKLIB for further analysis. The plotting can be done in the same steps as mentioned before to obtain the required parameters for the study that is essential to compare the IGS data to the obtained RAW data from the region.

After the Post-processing plotting and analysis of the original raw GNSS data received and the freely obtained IGS data from CDDIS, the final comparative study is being conducted with all the post-processing plotting for the various parameters we have obtained in both the cases. The comparison of the post-processing plots is

done separately for L1 band and L2 band, which are the frequency bands focussed for this study specifically. The plotting is done separately for both the IGS data and the GNSS original received data. The different parameters which are the most important factors for understanding the propagation in the region such as satellite visibility, multipath, SNR and elevation mask are analysed and studied carefully for all the processed data. Below is brief description of the most important factors which are observed in the study.

4.1 Multipath

Multipath in GNSS is where a signal reflection is superimposed on our required signal thus creating distortion. Any surfaces on the path of the signal can cause even the minute of reflections. It can be buildings, trees, walls and even water. These multipaths can create errors in our direct signals. Multipath error can also be minimized by upgrading the antenna quality that can mitigate signals coming from low elevation directions.

Thus, it is important to reduce the multipath effect to reduce the errors. Ground-based measurements can help to minimize the multipath in this case provided the most optimal environments and improvements in carrier filters are taken into consideration. The more the multipath error is reduced, the more accurate positioning can be achieved.

4.2 SNR

Any GNSS/GPS receiver measures the signal strength also called the signal to noise ratio (SNR). It is one of the most significant parameters that is much needed in any GNSS systems. The SNR is very helpful in evaluating the performance of the acquisition in a receiver. For the ground-based measurements in this study, the SNR of the GNSS receiver should be compared to the IGS/CDDIS data. SNR signals >40 dB are generally considered as excellent signal. Thus, the more the SNR, the better is our main signal.

4.3 Elevation Mask

When the satellites are low on the horizon, the elevation mask halts the receiver from using the satellites. An optimal elevation mask angle is set for the receiver. Signal multipath and atmospheric errors are highest for low elevation satellites; thus, it is very important to filter those observations. The default elevation mask generally is 10°. If we lower the elevation mask less than 10°, our system performance will not be

that great. For ground-based observations, it is to make the optimal elevation mask for the minimal error.

4.4 Satellite Visibility

The satellite geometry is also an important parameter that impacts the position estimation by the receiver. The GPS needs at least 5 satellites above the local horizon at any given time. The number of satellites visible changes throughout the day. A minimum of 4 satellites is must for estimating user location and time. Thus, the more the satellites used for the estimation, the greater is the quality of positioning and accuracy. It is must to have 5 common satellites for RTK initialization and 4 minimum satellites to be continuously tracked to have a viable RTK solution.

5 Results

The post-processing of the GNSS and IGS data has been performed separately. In this study, the plotting for the GNSS and IGS data has been done for L1 and L2 bands, respectively. Figure 1a, b shows the sky plot for the receiver and IGS data for L1 band. From both cases, it is seen that satellites G09, E04, R07, G04 and G11 are common and are continuously present for the time interval which implies that time and location estimation can be measured more accurately as well as multipath error can be reduced dramatically over the hill region. Similarly, Fig. 2a, b shows the sky

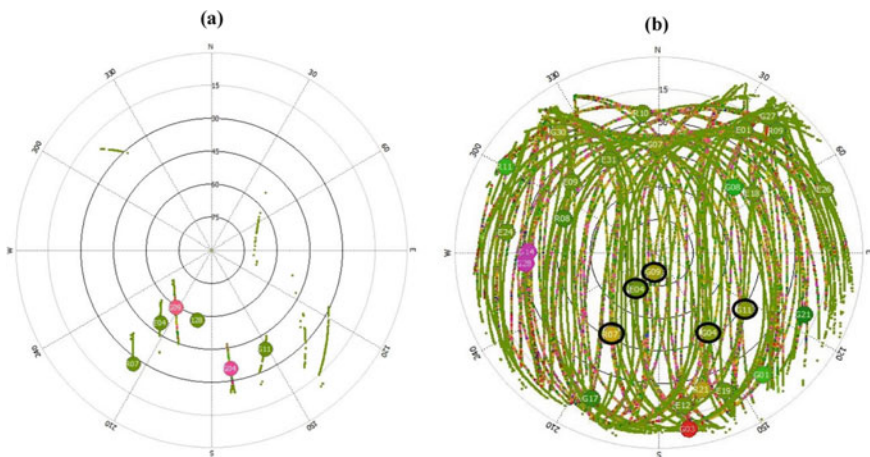


Fig. 1 Comparison of **a** original data and **b** CDDIS data of IISC IND for MP-sky plot L1 for 31st March, 2021, time period—4:30 to 5:35

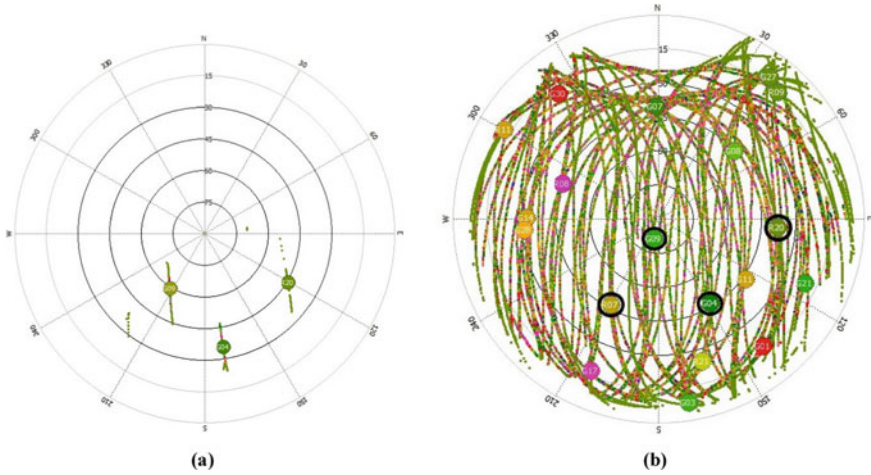


Fig. 2 Comparison of **a** original data and **b** CDDIS data of IISC IND for MP-sky plot L2 for 31st March, 2021, time period—4:30 to 5:35

plot for the L2 band, and the satellites G09, G04, R20 and traces of R07 are common and continuously available. Only 4 satellites are needed for forming a 3-D position fix and RTK initialization and a minimum of 4 to be continuously tracked to generate a RTK solution. Figures 3 and 4 give the total chart for the satellite visibility for the time duration. In Fig. 3, we have the satellite visibility plot for L1 band. We can infer the line plotting for G09, E04, R07, G04 and G11 which is also present in the IGS

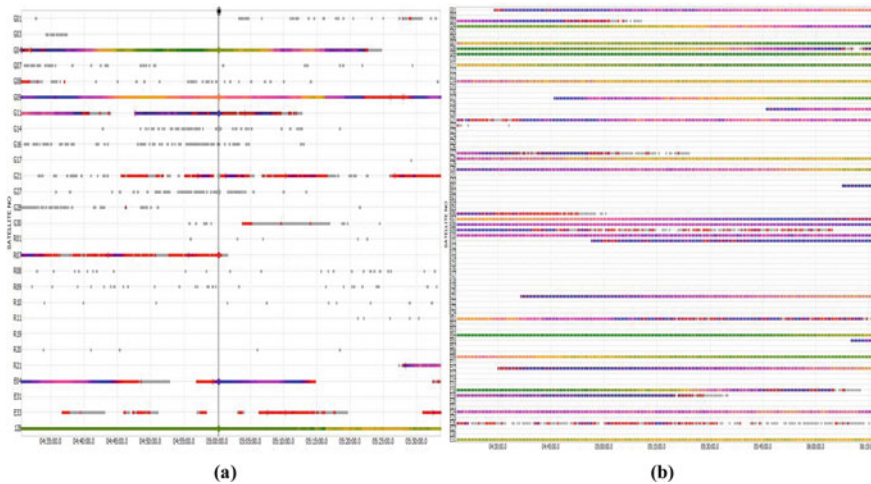


Fig. 3 Comparison of **a** original data and **b** CDDIS data of IISC IND for sat-visibility for L1 for 31st March, 2021, time period—4:30 to 5:35

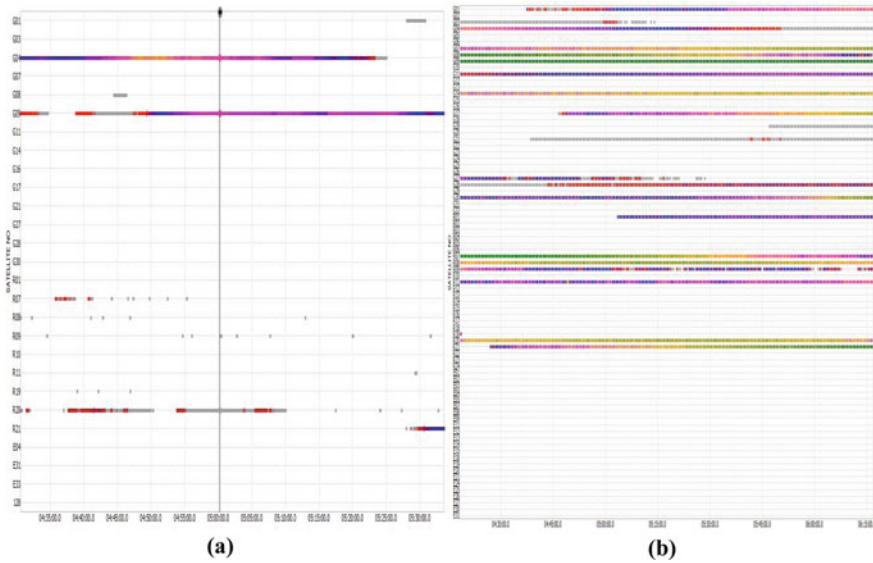


Fig. 4 Comparison of **a** original data and **b** CDDIS data of IISC IND for sat-visibility for L2 for 31st March, 2021, time period—4:30 to 5:35

data. In Fig. 4, G04, G09, R07 and R20 are the satellites visible for L2 band. Due to atmospheric interference, the R07 and R20 have a discontinuous visibility for L2 band. Comparing the Fig. 4a, b which is the IGS data, it is clear that all the common visible satellites observed are correct. With more time interval observed for different days, we can ensure continuous visibility.

Figs. 5 and 6 give the comparison for SNR, multipath and elevation angle for the processed receiver observations and the IGS data. The plotting has been done separately for both L1 and L2 bands each. For each visible satellite, we get the SNR and elevation plots within the time interval. Each satellite can be inferred on its own from the data with the help of RTKLIB. For the SNR, it can be deduced that for dataset the SNR >40–50 dB Hz, which is excellent for the direct signals. The SNR is constant throughout the whole-time interval; thus, it can be defined as a very good receiver system. From the sky plot, the satellites which are continuously visible will have the best SNR. So, it is clear from the analysis that the signals are in excellent condition with comparisons to the main IGS data. For the multipath plots, it can be seen in Figs. 5a and 6a that the distribution is geometry-free and the largest noise at edges is due to the low elevation rays of the satellite from the local horizon. For Figs. 5b and 6b, the multipath is geometric. For the best signal related to any purpose, the condition of the sky is the most important thing. The elevation mask angle is a must to filter out the satellites from our receiving end that are too low at the local horizon. The height also comes into factor with respect to the height of the GNSS receiver. From Figs. 5 and 6, it can be stated that the visible satellites within the time interval have a very good elevation angle, i.e. the elevation angle >10°–20°.

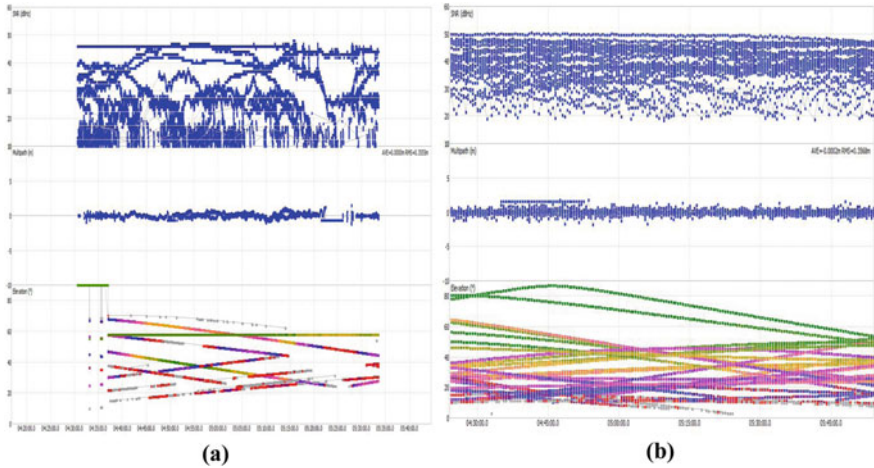


Fig. 5 Comparison of **a** original data and **b** CDDIS data of IISC IND for SNR/MP/EL, L1 for 31st March, 2021, time period—4:30 to 5:35

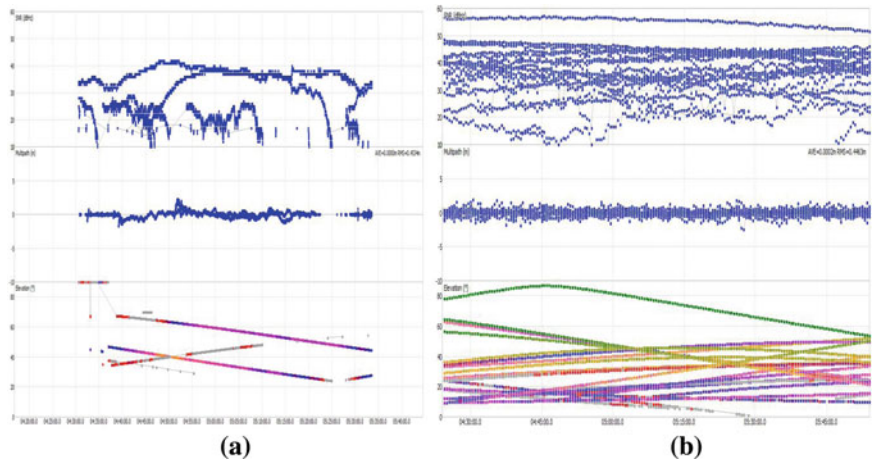


Fig. 6 Comparison of **a** original data and **b** CDDIS data of IISC IND for SNR/MP/EL, L2 for 31st March, 2021, time period—4:30 to 5:35

6 Discussions and Conclusion

The study and its observations have been conducted over a hill region which is specific for the study. From the post-processing and analysing of the observations, it can be stated from the plots that the elevation angle has a correlation to the visibility of the satellites. Thus, the post-processing analysis from the GNSS data is performing in a very good condition and supports all the factors that requires for a viable RTK

solution operating from a hill region. With all the factors corresponding well with the IGS data, the study shows that such GNSS system is very capable of operating in such a terrain and in turn can generate viable RTK solutions. Multipath is also one of the major factors to reduce errors of precise positioning. From the study, we can infer that the visibility of satellites, SNR and elevation angle all are suitable and can be operated in the region; thus, error reduction for multipath can also be tackled. Ground-based measurements such as the system used in this study can tackle the problem, and more ground-based studies will be conducted to find more efficient error margin mitigation techniques. The study has outlined the main factors of why ground-based measurements should be considered. The growing need of Internet of Things (IoT), smart applications, space-based applications and products and high precision GNSS systems is a key to solve the precise positioning problem. There will also be applications of ground-based measurement systems growing exponentially as more nations and private sectors are getting involved in this space-based economy.

Thus, from this study, it can be concluded that such GNSS system can be further developed for more RTK solutions and PPP solutions for all such type of hilly region where the applications of such GNSS-based systems are still very limited. We can further test out current algorithms for multipath and error mitigation and even develop new techniques for hilly terrains. The applications have a great range of scope for a vast number of industries. Advanced machine learning algorithms and neural network implementation with this GNSS system can yield very robust real-world applications, especially for military and defense purposes. Advanced machine learning algorithms and neural network implementation with this GNSS system can yield very robust real-world applications, especially for military and defense purposes. The study gives a comprehensive overview of how applications are based on GNSS, and GPS systems can be developed by understanding how it operates in such a region. It can also help to develop systems specific to such hilly regions. More innovative RTK solutions can be derived for more research and development. The most promising applications that could bear fruit are IoT-based devices, military products on such terrains, precise mapping and navigation-based applications.

References

1. René D, Fricke M, Schulz H (2002) Development of approach procedure design criteria for systems based on global navigation satellite system (GNSS) with ground based augmentation system (GBAS)
2. Takasu T, Yasuda A (2009) Development of the low-cost RTK-GPS receiver with an open-source program package RTKLIB
3. Simonetto E, Durand S, Burdack J, Polidori L, Morel L, Nicolas-Duroy J (2014) Combination of INSAR and GNSS measurements for ground displacement monitoring. *Procedia Technol* 16:192–198
4. Robustelli U, Baiocchi V, Pugliano G (2019) Assessment of dual frequency GNSS observations from a Xiaomi Mi 8 Android Smartphone and positioning performance analysis. *Electronics* 8(1):91

5. Lau L (2021) Chapter 4—GNSS multipath errors and mitigation techniques. In: Petropoulos GP, Srivastava PK (eds) *GPS and GNSS technology in geosciences*. Elsevier, pp 77–98
6. Hein GW (2020) Status, perspectives and trends of satellite navigation. *Satell Navig* 1:22

Ka Band Tropospheric Scintillation Estimation Over North East Indian Region



Nirmal Rai, Swastika Chakravarty, Kapila Sharma, Gopal Thapa,
and Rinkila Bhutia

Abstract Estimation of K_u/K_a band radio signal scintillation is done over eight North East Indian locations using meteorological database of NASA Giovanni. A signal fade of at least one dB is found even for ten percent of time during the observation period of one year for rainy summer season as well as non-rainy winter season. The effect of elevation angle is quite significant while estimating K_a band signal fade. As frequency has a very strong correlation with scintillation intensity, the design of SATCOM link with the use of higher frequency band needs accurate estimation of signal fade depth due to scintillation.

Keywords K_u/K_a band · Scintillation · North East India · Signal fade · Signal intensity

1 Introduction

Future generation satellite communication promises to deliver higher bandwidth and increased data rate to satisfy the ever-increasing demand of technology in society. As C band and K_u band are already congested, K_a band receives more attention nowadays for the design of earth space communication link. In the long-range communications, link reliability suffers from several factors including weather conditions like rain, fog, and clouds, which causes attenuation as well as scintillation. Tropospheric

N. Rai (✉) · S. Chakravarty · K. Sharma · G. Thapa
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: nirmal.r@smit.smu.edu.in

K. Sharma
e-mail: kapila.s@smit.smu.edu.in

G. Thapa
e-mail: gopal.t@smit.smu.edu.in

R. Bhutia
Sikkim University, Gangtok, Sikkim, India
e-mail: rbhutia@cus.ac.in

scintillation is countable above 3 GHz frequencies, and as ionosphere acts as a transparent layer above 10 GHz, there is a need to study the channel fading due to the turbulence caused in the atmosphere. Limited surveys have been done to quantify tropospheric scintillation over the tropical countries including India. ITU scintillation model is exploited to estimate tropospheric scintillation of Bangladesh [1], a subtropical country vulnerable to extreme climatic condition based on measured climatic parameters, like relative humidity and temperature. In this prediction, four major cities, namely Dhaka, Chittagong, Rajshahi, and Sylhet of Bangladesh, are selected for prediction of scintillation. As scintillation is caused by the change of refractive index of the propagating medium due to atmospheric turbulence (like temperature, humidity, etc.), the channel fade due to scintillation can happen in rainy as well as non-rainy condition. There are several prediction models like ITU-R [2], Karaswa [3], Otung [4], and Ortgies [5], already developed for estimating scintillation in subtropical countries, and very few models are based on the data of tropical countries [6]. Further, in many literatures, there exists substantial deviation of the model predicted scintillation fade with respect to actual fade due to scintillation observed in the experimental measurements. Conventional model like the ITU-R model is developed on the basis of the data collected at temperate region. Therefore, there is a need to validate the models for tropical region before using it by radio propagation engineers for satellite link design purpose.

Tropical Indian region, especially Northeast (NE) Indian hilly region as shown in Fig. 1, experiences the orographic rain as here moist air reaches the hill topography with a force to rise over mountains and hills [7]. This phenomenon forces the air to cool and condense, forming clouds. The air continues to be forced over the mountains and so it drops its moisture as relief rain or orographic rain. However, so far very few studies have been done on (NE) hill region to quantify scintillation fade due to very frequent orographic rain. This paper proposes a preliminary estimation of scintillation intensity and fades over NE Indian region based on ITU-R model.

2 Data and Methodology

The dataset used for this study is from the NASA Giovanni [8] data analysis system which has been recognized as a useful tool to access different types of remote sensing data like surface temperature, humidity, etc. all over the globe as a gridded dataset. Elevation angle considered here typically is 8° as low elevation angle can produce worst case in signal fade due to scintillation.

The telecommunication industry and academic organizations throughout the world recognize the work of ITU-R Study Groups for efficient use and management of spectrum/orbit resources and future radio communication systems characteristics. This work also takes the help of ITU-R model to estimate signal fade due to scintillation at K_a band frequency.

The performance parameters required for the method include average surface ambient temperature ($^\circ\text{C}$) (t), average surface relative humidity (%) (H), operating



Fig. 1 North eastern states of India

frequency (f): earth station antenna elevation angle (θ), earth station antenna diameter (D), and antenna efficiency (η).

The steps for estimating scintillation fade are given below:

1. Calculate the water vapor pressure, e_s [hPa],

$$e_s = 6.1121e^{\frac{17.502t}{t+240.97}}$$

2. Calculate the water vapor pressure e

$$e = \frac{He_s}{100}$$

3. Calculate the wet term of the radio refractivity (N_{wet})

$$N_{wet} = 3.732 \times 10^5 \frac{e}{T^2}$$

T = temperature (kelvin)

4. Calculate the standard deviation of the signal amplitude (σ_{ref})

$$\sigma_{ref} = 3.6 \times 10^{-3} + 10^{-4} * N_{wet}dB$$

5. Calculate the effective path length (L)

$$L = \frac{2h_L}{\sqrt{(\sin \theta)^2 + 2.35 * 10^{-4} + \sin \theta}} m$$

where h_L = height of the turbulent layer;

6. Calculate the effective antenna diameter (D_{eff})

$$D_{\text{eff}} = \sqrt{\eta} D m$$

7. Calculate the averaging factor of antenna:

$$g(x) = \sqrt{3.86(x^2 + 1)^{11/12} \sin \left[\frac{11}{6} \tan^{-1} \frac{1}{x} \right] - 7.08x^{5/6}}$$

With $x = 1.22D_{\text{eff}}^2(f/L)$, (where, f = carrier frequency).

8. Calculate the standard deviation of the signal for the considered period and propagation path

$$\sigma = \sigma_{\text{ref}} f^{7/12} \frac{g(x)}{(\sin \theta)^{1.2}}$$

9. Compute the time percentage factor, $a(p)$, for the time percentage p , of concern in the range $0.01 < p \leq 50$:

$$a(p) = -0.061(\log_{10} p)^3 + 0.072(\log_{10} p)^2 - 1.71 \log_{10} p + 3.0$$

10. Compute the scintillation fade depth for the time percentage p :

$$A_s(p) = a(p) \cdot \sigma \text{ dB}$$

3 Result and Discussion

Scintillation intensity and fade are estimated over eight NE Indian locations depending on the meteorological database of Giovanni [7] during the whole year of 2016. As satellite link design engineers require mitigating channel fading in a very efficient manner, dynamic information on rain fade is very important in this regard. Therefore, dynamic analysis of rain fade is carried out during the summer months predominated by orographic relief rain over the NE hills and during the winter months also. From Fig. 2, it is evident that for even ten percent of time, there is a fade of 1 dB at Gangtok to 2 dB at Shillong due to scintillation at summer months. At winter months, fade reduces to 0.8 dB at Gangtok to 1 dB at Shillong depending on the amount and intensity of rain as evident from Fig. 3 for the same ten percent of time.

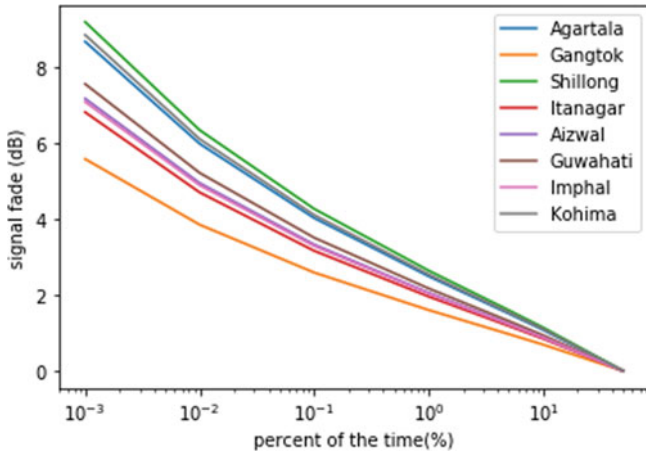


Fig. 2 Signal fade distribution for summer months

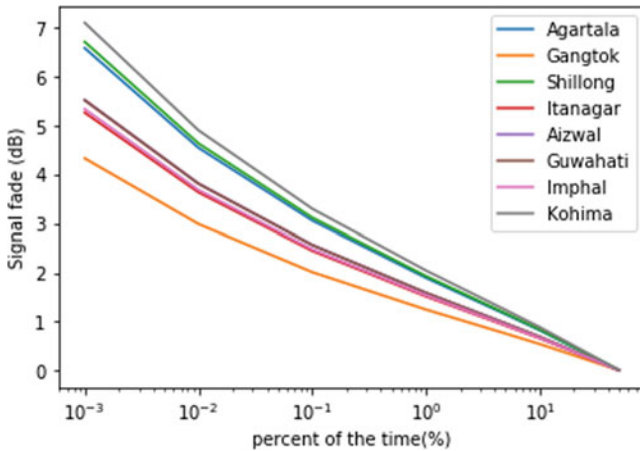


Fig. 3 Signal fade distribution for winter months

Scintillation intensity when plotted for the whole year of 2016, Fig. 4 shows that in the months from May to September, over the NE Indian hills, there is a variation of 0.5–0.9 dB depending on the location and amount of rain as here we have considered wet scintillation only.

Elevation angle plays an important role while quantifying scintillation as low elevation angle forces the signal to pass through longer path through atmosphere which increases the possibility of scintillation. Therefore, as shown in Fig. 5, for the elevation angle 8° scintillation fade depth rises approximately to 2 dB even for one percent of time. As shift to higher frequency is the need of the hour, investigation is done in this paper to find scintillation intensity by increasing frequency up to 20 GHz.

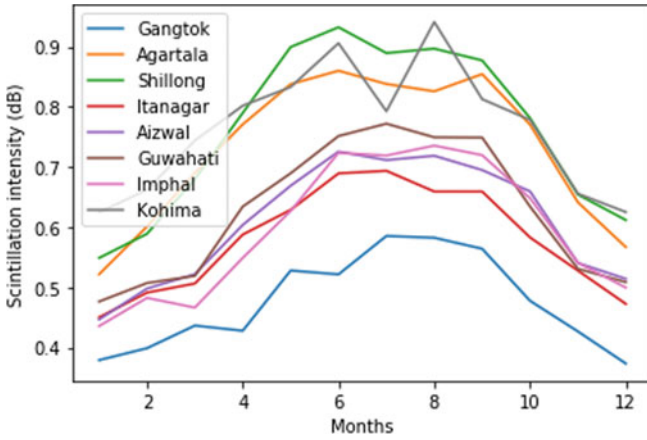


Fig. 4 Annual scintillation intensity variations

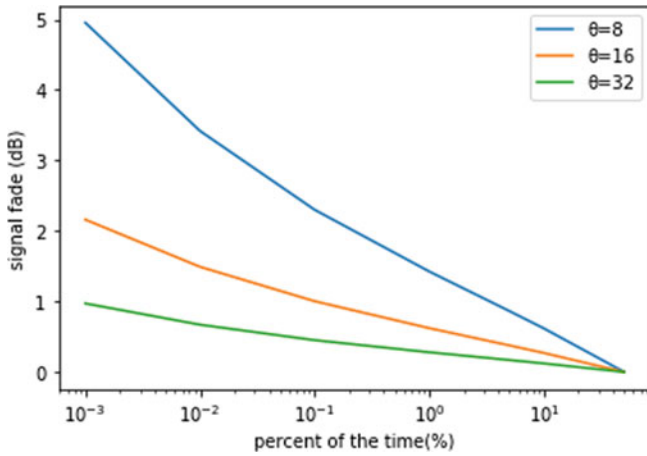


Fig. 5 Signal fade variations with elevation angle (degree)

A positive correlation is found between scintillation intensity and the frequency over the frequency band of 8–20 GHz as shown in Fig. 6. Depending on rain amount, there is a significant variation of scintillation intensity for a particular frequency between the two location Shillong and Gangtok.

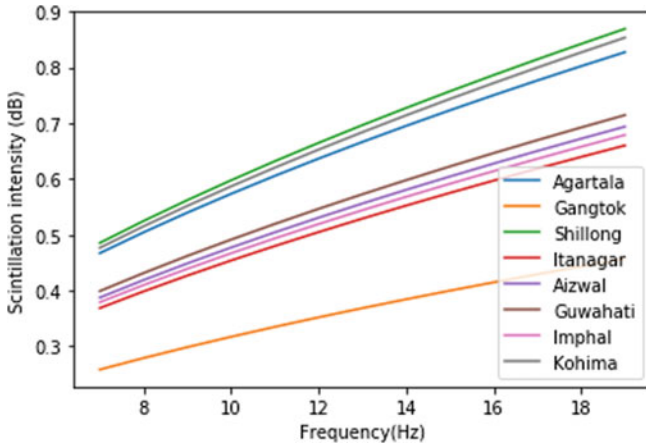


Fig. 6 Variations of scintillation intensity with frequency

4 Conclusion

Effect of scintillation on signal fade depth and signal intensity is thoroughly analyzed and estimated over NE Indian hills predominated by orographic rain using ITU-R model and available meteorological database. A substantial amount of signal fade due to scintillation is found by analyzing one year database. Further investigation is needed, considering long-term available dataset to actuate the amount of signal fade. Validation is also very essential with the help of simultaneous beacon signal experiment if experimental arrangement is available over this region.

Acknowledgements The authors would like to acknowledge NASA, GIOVANNI web-based application for providing relative humidity and surface temperature satellite information. The author thankfully acknowledges Sikkim Manipal Institute of Technology, Sikkim, for providing logistic support for conducting the research.

References

1. Hossain MS, Samad MA (2015) The tropospheric scintillation prediction of Earth-to-Satellite Link for Bangladeshi climatic condition. *Serb J Electr Eng* 12(3)
2. International Telecommunication Union (2009) Propagation data and prediction methods required for the design of Earth-space telecommunication systems, recommendation ITU-R P.618-10
3. Karasawa Y, Yamada M, Allnutt JF (1988) A new prediction method for tropospheric scintillation on Earth-space paths. *IEEE Trans Antenna Propag* 36(11):1608–1614
4. Otung IE (1996) Prediction of tropospheric amplitude scintillation on a satellite link. *IEEE Trans Antenna* 44(12):1600–1608

5. Ortgies G (1993) Prediction of slant-path amplitude scintillation from meteorological parameters. In: International symposium on radio propagation, Beijing, China, 18–21 Aug 1993, pp 218–221
6. Yee ACC, Mandeep JS et al (2011) Development of new tropospheric scintillation prediction model for country in tropical climate. In: Proceeding of the 2011 IEEE international conference on space science and communication (IconSpace) 12–13 July 2011, Penang, Malaysia
7. Karmakar S et al (2019) Tropospheric scintillation at Ku/Ka band in Orographic Region. In: 2019 3rd International conference on electronics, materials engineering & nano-technology (IEMENTech)
8. Nasa, Giovanni, <https://giovanni.gsfc.nasa.gov/giovanni>

Compact Insert Fed Monopole Antenna with Four L-Slot Insertion Design for Wireless Applications



Y. Venkata Lakshmaiah, Bappaditya Roy, and R. Dewan

Abstract In this paper, we present a modified printed monopole antenna for wireless applications. The proposed antenna consists of an inserted fed to radiating patch with a four L slots. The designed antenna was found to operate at three wireless frequency bands of 2.3–5.86, 5.8–6.37 and 9.3–10.5 GHz. This design achieves better impedance matching without any additional circuits. The parametric studies and radiation characteristics show the antenna has good characteristics and is applicable for mobile wireless applications.

Keywords Monopole · Inserted fed · L slot · Wireless frequency

1 Introduction

The demand of compact size antennas is increasing in a modern communication system because of their wideband and multiband characteristics [1]. Due to their very low weight, less cost, low profile and easy to fabricate, researchers are attracted to design a monopole antenna [2, 3]. Modern technology supports various wireless applications and operated on different frequency bands. Hence, there is a huge need for multiband, wideband, super wideband antennas [4, 5]. Nowadays, due to advancements in wireless technologies, the demand for wideband antennas is increased.

A lot of new research has been proposed by several antenna researchers in the field of microstrip antennas. Many techniques are adopting to achieve wideband characteristics such as partial ground plane, defected ground plane (DGS), fractal patch and monopole structure [6–10]. In this work, using CST MW Studio simulation tool design one compact size antenna. Using CST tool, generate the radiation patterns, electric field and magnetic field for different frequencies as well as plotting return

Y. V. Lakshmaiah · B. Roy (✉)
VIT-AP University, Amravati, Andhra Pradesh, India
e-mail: bappaditya.roy@vitap.ac.in

R. Dewan
Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

loss values for the S_{11} port parameters in the calculating operate frequency range. The design that exhibits the optimum return loss and radiation pattern will be simulated. This project is to design, simulate and verify the simulation results with different parameters value of the proposed microstrip monopole antenna.

2 Antenna Design

The design and schematic configuration of the reference antenna is shown in Fig. 1. In the first stage of design, a monopole antenna of width (W_p) = 17 mm and length (L_p) = 19.5 mm is designed to work as a reference antenna. The dimension of the substrate plane is $50 \times 50 \text{ mm}^2$ and the ground plane is $50 \times 16.3 \text{ mm}$. The substrate of the proposed antenna is Rogers RO 3003 having a relative permittivity 3 and thickness of 1.5 mm. The location of the rectangular feed point is in the middle of the patch and the ground plane. The microstrip rectangular feed is given to the patch. Second stage, four L shapes were used as two symmetrical L shapes on the upper patch ($l_1 = 13 \text{ mm}$, $w_1 = 1 \text{ mm}$ and $l_2 = 6 \text{ mm}$) another two symmetrical L shapes on the lower patch ($l_3 = 7 \text{ mm}$, $w_2 = 1 \text{ mm}$, $l_4 = 4 \text{ mm}$ and $l_5 = 2 \text{ mm}$) is shown in Fig. 2. The various dimensions of the proposed structure are indicated in Table 1.

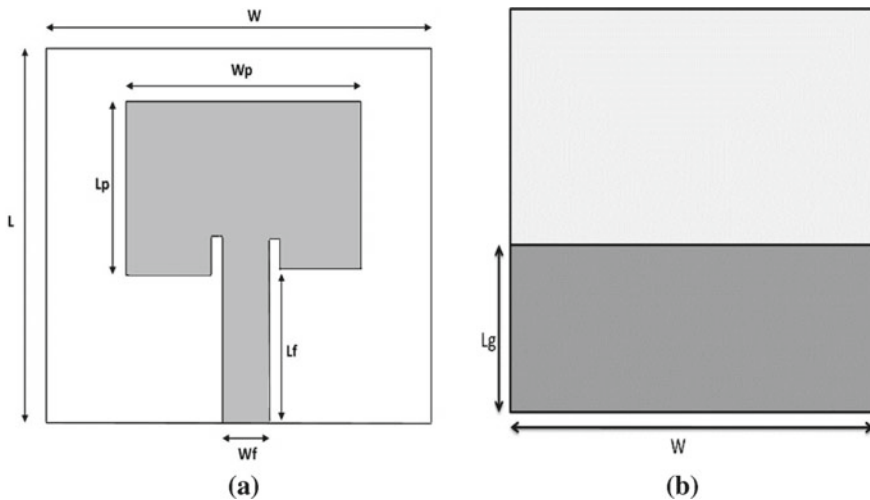


Fig. 1 **a** Front view of the reference antenna, **b** back view of the antenna

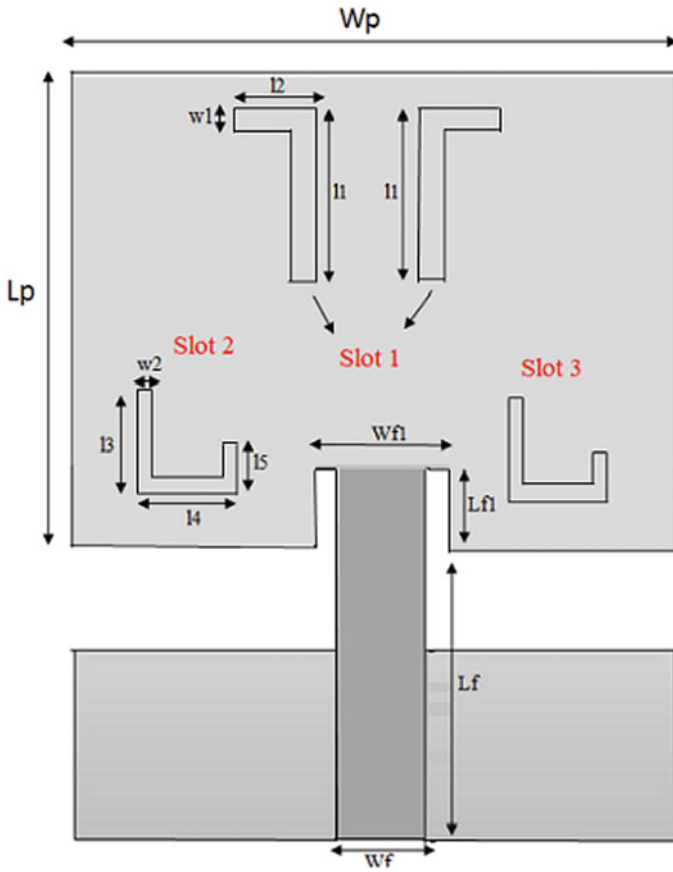


Fig. 2 Final antenna structure

Table 1 Table summary of antenna dimensions

Parameter	W	L	Sh	Lg	mt	Wp	Lp	Lf	Lf1
Values (mm)	50	50	1.5	16.3	0.035	17	19.5	17	3
Parameter	Wf	Wf1	l1	l2	w1	l3	l4	l5	w2
Values (mm)	2.5	4.5	13	6	1	7	4	2	1

3 Results and Analysis

The proposed structure was simulated on CST Microwave studio. Figure 3 shows the simulated reflection characteristics of the proposed structure. In the proposed antenna, we inserted four “L” slots which makes the structure more compact and it is also important for three resonant bands of frequencies. In Fig. 4, the return

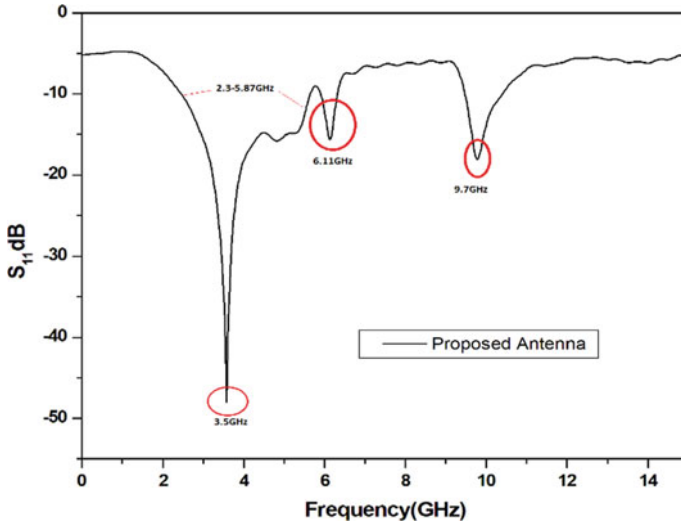


Fig. 3 Proposed antenna reflection coefficient (dB) versus frequency

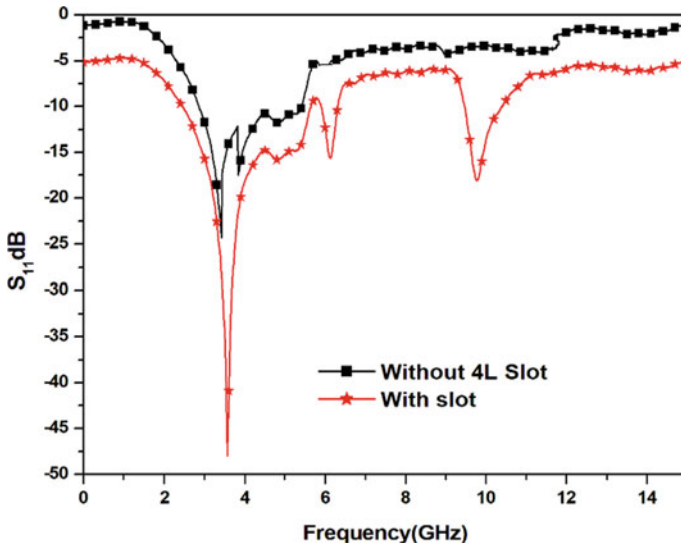


Fig. 4 Reflection coefficient versus frequency plot for without and with slot

loss variation using slots is shown. It will observe that the maximum bandwidth (2.3–5.86 GHz) is obtained where the other two frequencies are (5.8–6.37 GHz) and (9.3–10.5 GHz). Each “L” slot is responsible for each resonant frequency.

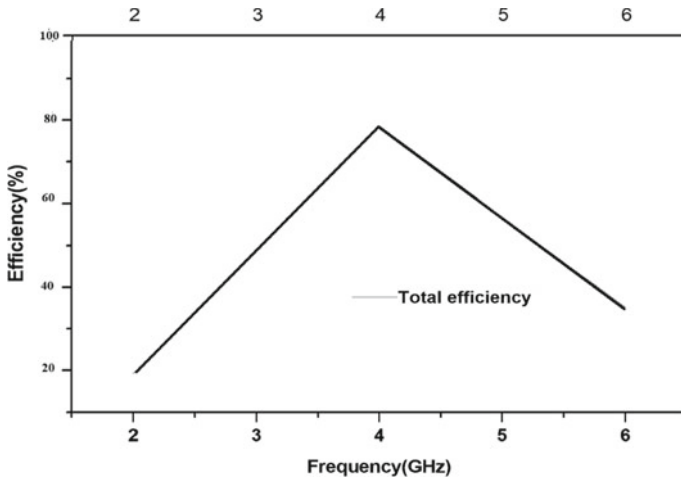


Fig. 5 Efficiency of the proposed antenna

Accurate or perfect antenna efficiency is important for any wireless antenna. Figure 5 will show the efficiency of the proposed structure where the maximum efficiency is 70% at the 4.1 GHz frequency range.

Impedance is one of the prime important parts of a microstrip antenna. Placing of rectangular fed in the proper position is affected in resonant frequency. Figure 6 shows the impedance versus frequency plot.

To know the dielectric constant and loss tangent, we must characterize a substrate, the dielectric constant is prime important to determine the impedance of the circuitry. The proposed structure design on Rogers RO 3003 substrate has a thickness of 1.5 mm. Figure 7 shows the effect of the reflection coefficient on different substrates.

The simulated far-field gain patterns were plotted at different frequencies as shown in Fig. 8. It can be seen from Fig. 8 at the lower frequency, the far-field patterns are like conventional monopole and slight distortion when the frequency increased.

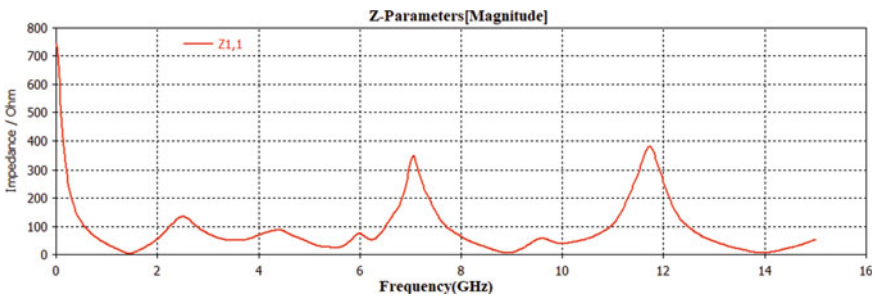


Fig. 6 Impedance versus frequency plot

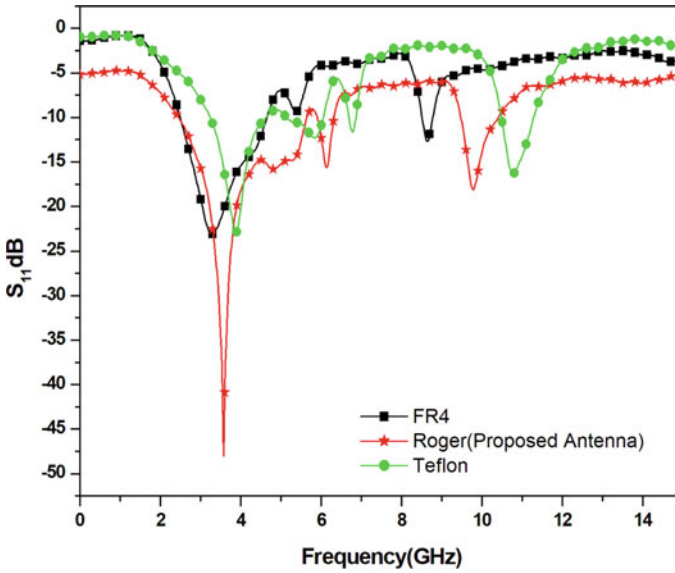


Fig. 7 S_{11} versus frequency plot at different substrate

The surface current distribution pattern of the proposed structure is shown in Fig. 9. The two L-type slots allow finding current distribution in the patch which allows maximum electric current throughout the transmission system.

4 Conclusion

In this article, we design a monopole antenna with the help of four “L” slots. Proposed structure is simulated on CST MW studio. As per our proposed structure and all results, we can conclude that the proposed compact monopole antenna perfectly matches the various wireless frequency bands.

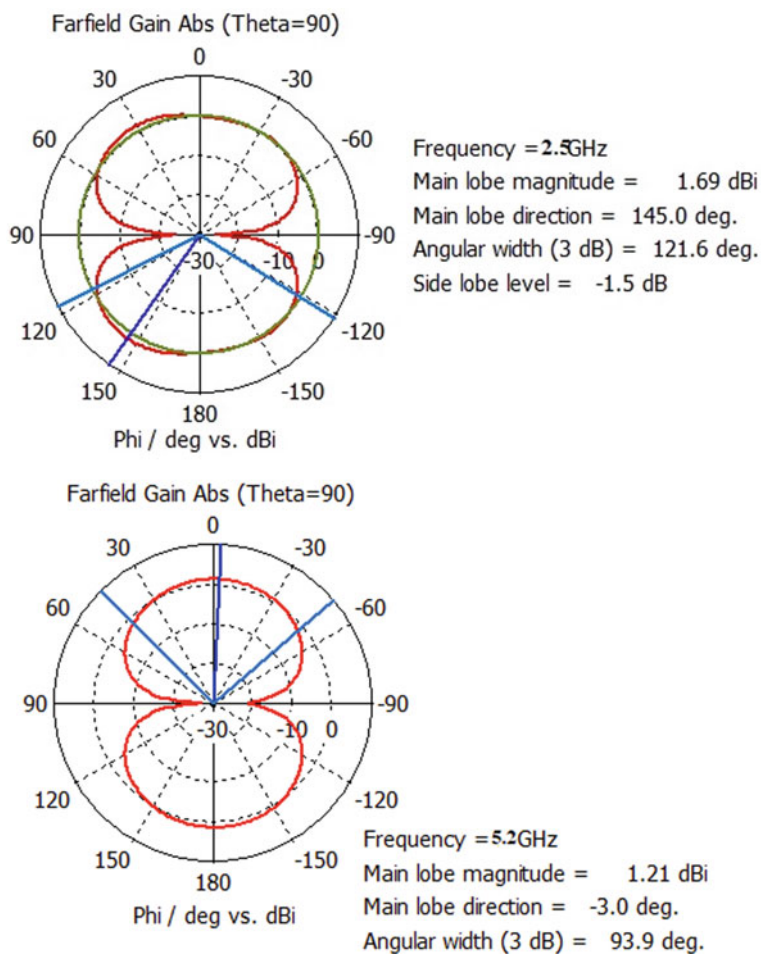
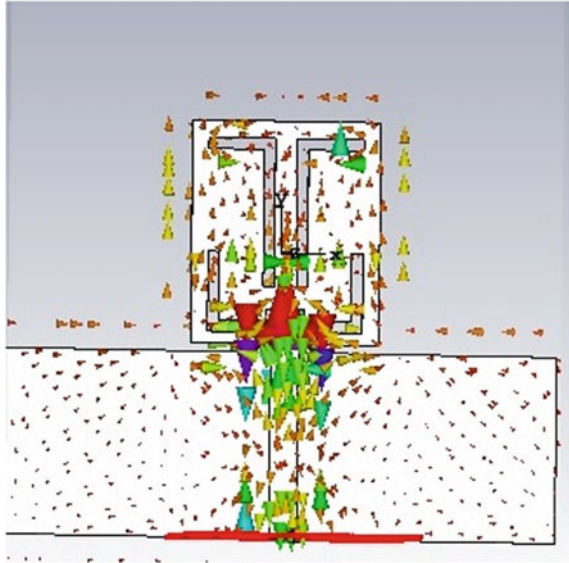


Fig. 8 Far-field gain plot at different frequency

Fig. 9 Current distribution of proposed structure at 5.2 GHz



References

1. Rabike SV, Sahu SD, Khobragade SV (2014) Fractal antenna for multi-frequency applications using PIN diode. *J Comput Electron*. <https://doi.org/10.1007/s10825-014-0640-6>
2. Mishrs, A., J. A. Ansari, K. Kamakshi, A. Singh, M. Aneesh, and B. R. Vishvakarma, "Compact dual band rectangular microstrip patch antenna for 2.4/5.12 GHz wireless applications," *Journal of Mobile Communication, Computation and Information*, 2014, doi: <https://doi.org/10.1007/s11276-014-0783-1>.
3. Bargavi, K., K. Sankar, and S. A. Samson, "Compact triple band H-shaped slotted circular patch antenna," *IEEE Conference on Communication and Signal Processing*, 1159–1162, 2014.
4. Reddy VV, Sarma NVSN (2014) Tri-band circularly polarized Koch fractal boundary microstrip antenna. *IEEE Antennas Wirel Propag Lett* 13:1057–1060
5. Thakare, Y. B. and Rajkumar, "Design of fractal antenna for size and radar cross-section reduction," *IET Microwave, Antennas and Propagation*, Vol. 4, No. 1, 175–181, 2010, doi: <https://doi.org/10.1049/iet-map.2008.0325>.
6. Bappaditya Roy, Ankan Bhattacharya, A.K. Bhattacharjee, S.K.Chowdhury-"An UWB Monopole Antenna for WLAN and WiMAX Applications", *ICIIS 2014*, pp.1–3, 15–17, DOI:<https://doi.org/10.1109/ICIINFS.2014.7036506>
7. Bakariya PS, Dawari S, Sarkar M (2015) Triple band notch UWB Printed monopole antenna with enhanced bandwidth. *International Journal of Electronics and Communications (AEU)* 69:26–30
8. Chen KR, Sim CYD, Row JS (2011) A compact monopole antenna for super wideband applications. *IEEE Antennas Wirel Propag Lett* 10:488–491. <https://doi.org/10.1149/LAWP.2011.2157071>
9. Srivastava DK, Khanna A, Saini JP (2015) Design of a wideband gap coupled modified square fractal antenna. *J Comput Electron*. <https://doi.org/10.1007/s10825-015-0740-y>
10. Choukiker, Y. K. and S. K. Behra, "Design of wideband fractal antenna with combination of fractal geometries," *IEEE Conference on Information, Communication and Signal Processing* (2011). *ICICS*. <https://doi.org/10.1100/ICICS.2011.6174226>

Deep Learning Approach for Wind Power Forecasting



Nishant Saxena, Rahul Kumar, Rachit Saxena, Sri Krishna Mishra, Ritu Jain, and Sujit Kumar Verma

Abstract One of the most important tasks for any electricity system is to balance supply and demand. This is a dynamic equilibrium. An amount is projected for power demand based on historical patterns. This demand estimation will assist the system operator in determining how much electricity to deliver to meet demand or decrease waste. Demand appears to fluctuate over time, requiring supply to be adjusted correspondingly. Traditional fossil-fired power plants have a far simpler time adjusting their output than wind power plants. The output of thermal power plants can be programmed to respond to variations in demand. Wind power, like most other forms of renewable energy, is subject to unpredictable swings. Wind electricity, unlike natural gas power plants, cannot be sent. The maximum wind power output is mostly determined by the current wind conditions. We have proposed a novel approach of deep learning for forecasting wind at the location of Dewas, Madhya Pradesh, India. The wind power has been forecasted using the machine learning regression and long short-term method for the comparative study on MATLAB and concluding the more accurate predictions than the conventional statistical and other conventional techniques.

Keywords Wind power · Wind power forecasting · Weibull's probability density function · Deep learning · Machine learning and regression method

1 Introduction

Wind is intermittent, which is one of the features. There are times when there is practically little evidence of the wind, and other times when people can readily tell it is windy simply by hearing it. The topography has a significant impact on wind. Because of the difference in air temperature between land and ocean, wind

N. Saxena (✉) · R. Kumar · R. Saxena · S. K. Mishra · R. Jain
Suresh Gyan Vihar University, Jaipur, India
e-mail: nishant.gitee@gmail.com

S. K. Verma
GLA University, Mathura, India

is generally greater near the beach. A comparable phenomenon, higher wind speed around some valleys, is the outcome of a similar scenario: a temperature differential, in this case between mountain and valley. Wind profiles differ from season to season, with stronger winds being seen more commonly India during the spring. While the rate of wind penetration continues to rise, the function of wind power prediction has become increasingly crucial. Because demand estimate is never perfect, every power system has a fair capacity to respond to variations in demand [1]. When the penetration rate is low, power systems do not need to be concerned about the variability of wind power supply. This is because, for example, a decrease in wind speed resulting in decreased wind power supply is very comparable to an increase in demand, which the power system can manage [2]. Wind power projections are divided into four groups based on their time scales: extremely short-term (few seconds to 30 min), short-term (30 min to 6 h), medium-term (6 h to 1 day) and long-term (more than 1 day).

1.1 Very Short-Term Wind Power Predictions

Very short-term prediction has a time scale that typically spans from 1 min to half an hour, but it may potentially be as short as a few seconds. The findings of prediction can be utilized for wind turbine control [3], real-time grid operation [4] and other applications. This approach also took a small amount of time to compute.

1.2 Short-Term Predictions of Wind Power

Wind power forecasts with time frames of 30 min to 6 h are classified as short-term forecasts. This group of methods is mostly used to plan load dispatch in order to make wind power more economically competitive [5].

1.3 Medium- and Long-Term Wind Power Predictions

Predictions between 6 h and 1 day and longer than 1 day are classified as medium term and long term, respectively. These predictions can be helpful in the decisions of unit commitment, reserve requirement, maintenance, etc. [6–10]. It is worth noting that research done for long-term prediction is less than that for shorter time horizons.

2 Proposed SVM Based Wind Power Prediction Model

The support vector machines (SVMs) are controlled machine learning techniques that may be used for both regression and classification. However, they are most commonly employed in categorization issues. SVMs were initially developed in the 1960s, but they were improved around 1990. In comparison to other machine learning algorithms, SVMs feature a unique implementation method. They've recently gained a lot of traction due to their capacity to handle numerous continuous and categorical variables. In multidimensional space, an SVM model is essentially a representation of various classes in a hyperplane. SVM will generate the hyperplane in an iterative manner in order to reduce the error. SVM's objective is to split datasets into classes such that a maximum marginal hyperplane may be found (MMH). Seeking support vectors to determine a linear hyperplane, an SVM appears to be only capable of representing a linear relationship between input and output. Original input data, on the other hand, can be transformed into a more sophisticated form with many dimensions. The linear relationship that may be expressed can be discovered in this new feature space.

$$\hat{P}_q = w_0 + \sum_{j=1}^m w_j \vartheta_j(e_q) \quad (1)$$

where $\{\vartheta_j\}$ a set of nonlinear mapping to the new feature space. When transferred back to the original feature space, the nonlinear relationship between input equation and prediction \hat{P}_q is obtained accordingly.

3 Proposed Deep Learning Method-Based Wind Power Prediction Model

The ANN is one of the machine models that requires work to train; in the case of the DNN, which has several hidden layers, the learning process is more difficult. Nonetheless, the capacity to understand more complicated and abstract relationships is a benefit. The recurrent neural network (RNN) is a deep neural network that uses recurrent neural networks (RNNs). Artificial neural networks (ANNs) have been built and enhanced [11–14]. RNN is a time series-based sequence model that can train and extract features directly from inputs in the time series domain. It may be utilized in a wide range of applications with time series inputs. It is utilized here for wind power forecasting using time series sequences. However, because of the exponentially rapid declining gradient norm toward 0 and blast issues, RNN suffers from gradient evanescence and has a limited capacity to learn long-term temporal correlations. The LSTM neural network was designed to solve the vanishing gradient problem [15–19].

The LSTM network is made up of an input layer, an output layer and multiple recursive concealing layers in between, as illustrated in the diagram. A recursive hiding layer is made up of several memory modules, each of which includes one or more self-contained memory units controlled by three gates: the input gate, forgetting gate and output gate. The basic structure of LSTM network is shown in Figs. 1 and 2. Furthermore, the schematic diagram of neurons is shown defining the various gates in the network.

At time interval t output gates values are computed as follows:

$$i_t = \text{Sigmoid}(IW_i \times X_t + RW_i \times h_{t-1}) + b_i \tag{2}$$

$$f_t = \text{Sigmoid}(IW_f \times X_t + RW_f \times h_{t-1}) + b_f \tag{3}$$

$$o_t = \text{Sigmoid}(IW_o \times X_t + RW_o \times h_{t-1}) + b_o \tag{4}$$

$$c_t^* = \text{TanH}(IW_c \times X_t + RW_c \times h_{t-1}) + b_c \tag{5}$$

Fig. 1 Basic structure of LSTM network

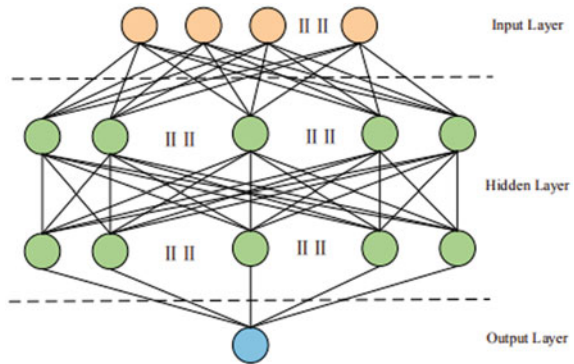


Fig. 2 Schematic diagram of neuron

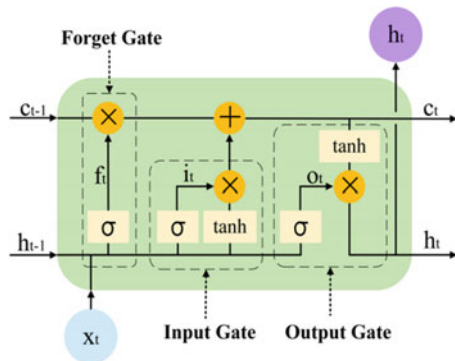


Table 1 Activation function of LSTM method

Activation function	Formula
Sigmoid	$\text{Sigmoid}(x) = \frac{1}{1+e^{-x}}$
TanH	$\text{TanH}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

The activation function is a smooth nonlinear function that specifies the neuron’s output in response to a single or many inputs. The activation functions for the LSTM structure are the hyperbolic tangent (TanH) and sigmoid functions, which are specified as follows in Table 1.

4 Results and Conclusion

Estimators and Evaluators: In the study and analysis purpose, several indicators are identified and used as an estimator for the evaluation of methods and accuracy of predictions made. Some have been used here like, mean absolute error, mean square error and root mean square error discussed in brief below.

Mean Absolute Error: The normalized mean absolute error (NMAE) is an error quantity that is defined in Eq. 6 and looks at the average of the absolute error of the forecast.

$$\text{NMAE}_j = \frac{1}{N} \sum_{k=1}^N |\varepsilon_{k+j|k}| \tag{6}$$

Mean absolute deviation (MAD) is another popular term for mean absolute error (MAE). This figure reflects the size of the total error that has occurred as a result of predicting and should be as little as feasible. This inaccuracy is scale dependent and is influenced by data processing and measurement scale.

Mean Squared Error: The normalized mean squared error (NMSE) is an error quantity that is calculated using the normalized sum of squared error (NSSE) described in Eq. 7 and looks at the average of the squared errors:

$$\text{NMSE}_j = \frac{1}{N} \sum_{k=1}^N \varepsilon_{k+j|k}^2 \tag{7}$$

Positive and negative mistakes do not cancel one other out in this error, and big individual errors are penalized more severely.

Root Mean Square Error: The normalized root mean squared error (NRMSE) is a type of error that squares the NMSE. It’s defined in Eq. 8

$$\text{NRMSE}_j = \text{NMSE}_j^{1/2} = \left(\frac{1}{N} \sum_{k=1}^N |\varepsilon_{k+j|k}|^2 \right)^{1/2} \tag{8}$$

Support vector machine (SVM) results: Using the SVM method, the dataset is applied to developed MATLAB program. Two different methods of SVM have been applied on the proposed dataset, linear SVM and medium Gaussian SVM method (GMSVM). Based on the above discussed evaluation indices, the proposed methods have been evaluated. Figure 3 shows the prediction plot of linear SVM method, and Fig. 4 shows the Gaussian median SVM method.

It can be clearly seen that GMSVM method gives the better prediction with less error which are also been verified by the error comparison table for both the models. The GMSVM method provides the RMSE value of 0.63013 which is much lesser than comparative to linear SVM method as 2.0135. Furthermore, the training time required in the GMSVM is also better then linear SVM method.

Furthermore, response plots for the linear SVM and Gaussian median SVM method are shown in Figs. 5 and 6. It can be clearly seen that the GMSVM method response with wind speed is better and more nearer to the true values. The predictions of the power is more accurate toward the wind speed. Figures 7 and 8 show that

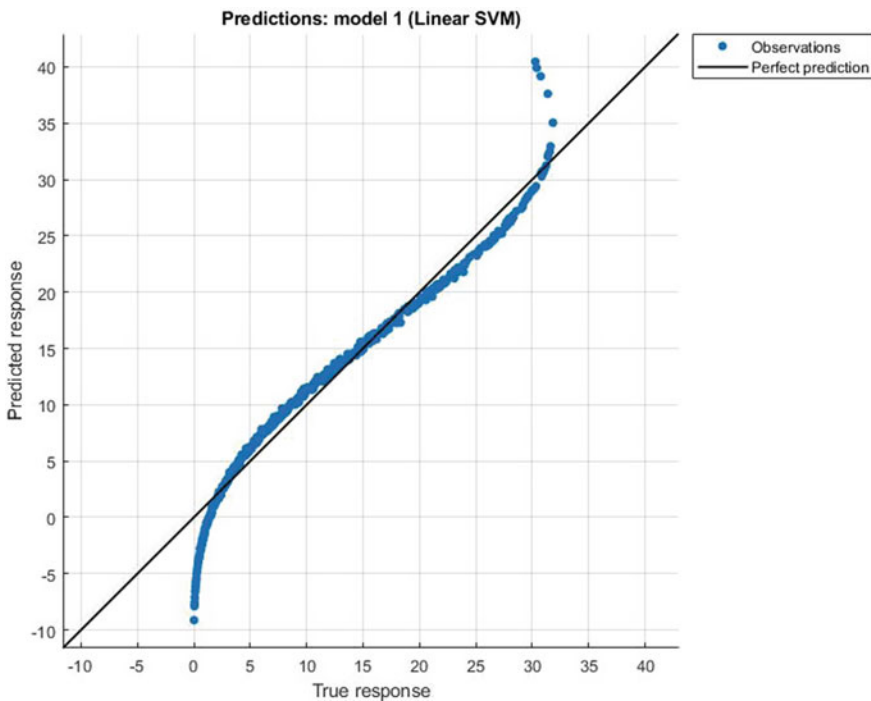


Fig. 3 Response plot of linear SVM

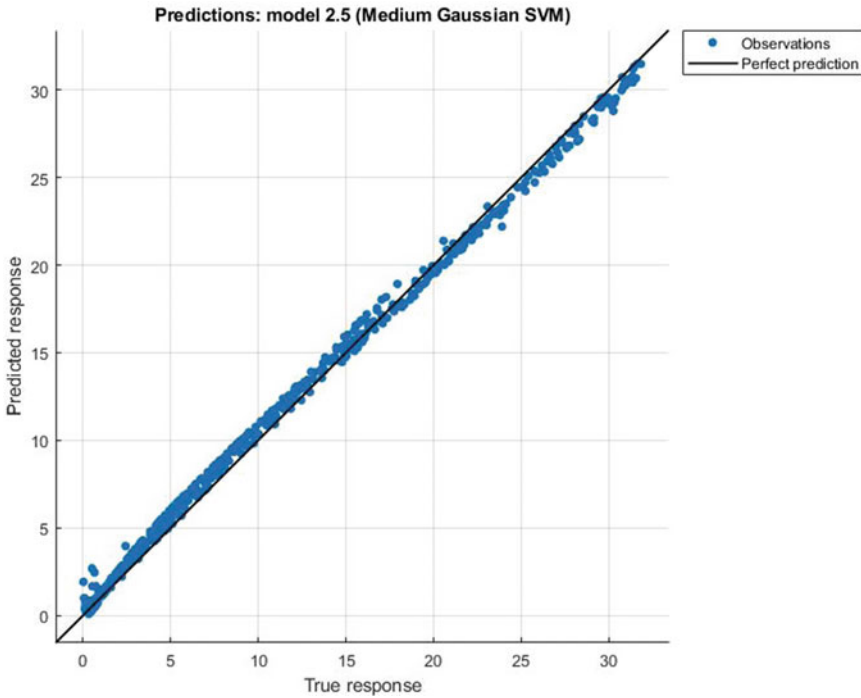


Fig. 4 Response plot of Gaussian median SVM method

residual plots of both linear SVM and Gaussian median SVM method. The figures clearly depict the no. of residuals generated for the predicted and true values.

Long short-term memory (LSTM) method results: RNN—LSTM methods have been implemented on the MATLAB for the forecasting of wind power for next interval. The same dataset of wind power, temperature and wind speed has been applied, and the method machine learning model has been developed using MATLAB deep learning toolbox. The training has been done to the 70 percent of the data, and remaining has been kept for the testing.

After the implementation and setting of above parameters, the results were obtained and shown from Figs. 9, 10, 11 and 12. It can be clearly seen from Fig. 9 that the root mean square error reached to below 0.2 with and loss of the training of less than 0.05 defining the better performance than previously discussed two regression methods, viz. linear support vector machine and Gaussian median support vector machine methods. Figure 4.8 shows the forecasted values of the wind power for the testing data which can also be verified by Fig. 10. Figure 11 shows that the forecasted value of wind power is obtained as similar to the wind power actual during the same period of the days.

After finding the predictive errors for both the proposed approaches, it has been found that deep learning-based LSTM method is more profound and relatively

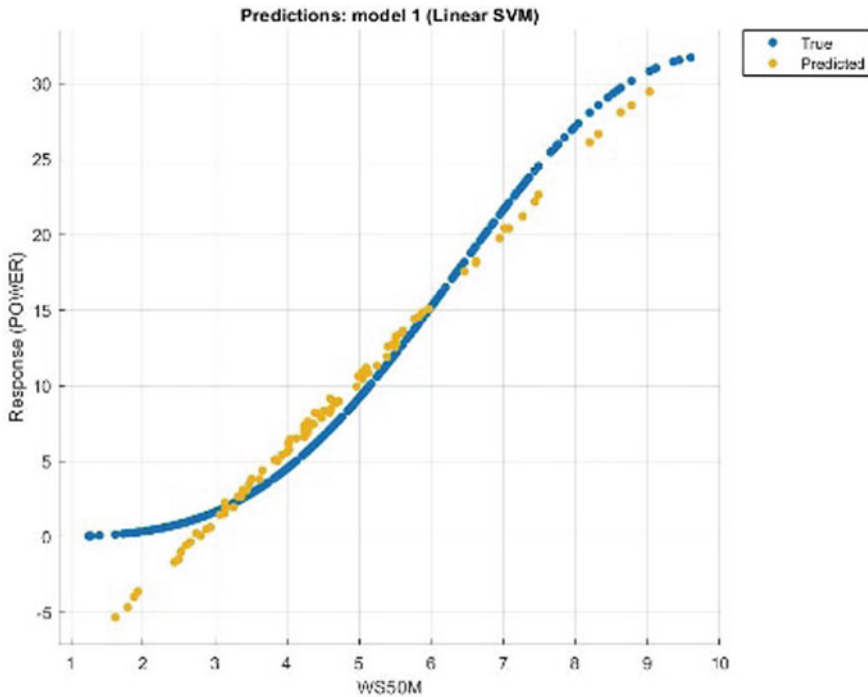


Fig. 5 Response plot of linear SVM for Power

predicts more accurate results comparative to regression-based SVM methods. The comparative results are shown in Table 2. Figure 12 shows the graphical comparison between the various techniques used in the proposed methods.

5 Conclusion

The study provides the more accurate forecasting with less errors in predictions using the novel approaches of support vector machine and deep learning techniques. The deep learning technique is found to be more reliable in comparison to SVM. The comparative Table 2 clearly describes that the RMSE error for the deep learning methods is less than linear SVM and Gaussian median SVM method. With the similar methodology, we can also develop the machine learning models for predicting the solar and wind power for the different sites in India giving more reliable operation for the scheduling and power forecasting in the vicinity. With the increase in the demand of the renewable power and initiatives taken by the government of India, the demand is consistently increasing and more accurate prediction models will be in requirement for forecasting with lesser time and higher accuracy.

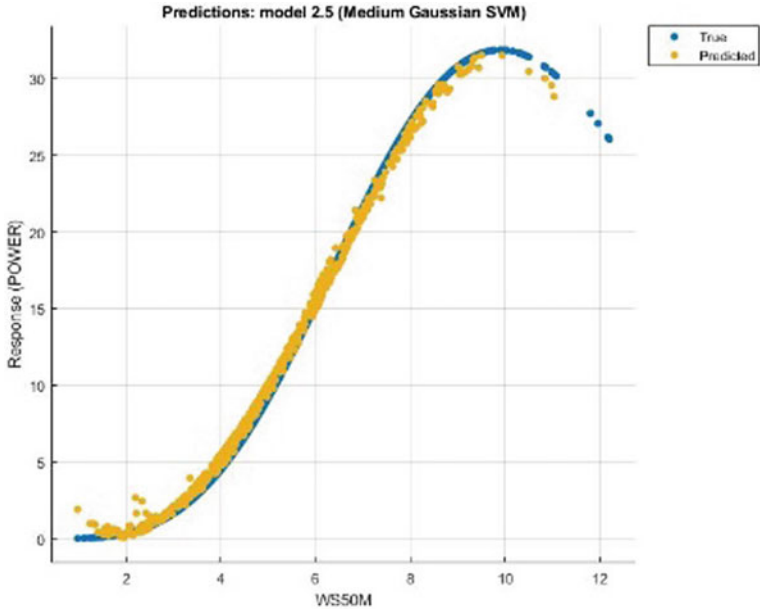


Fig. 6 Response plot for Gaussian median SVM method

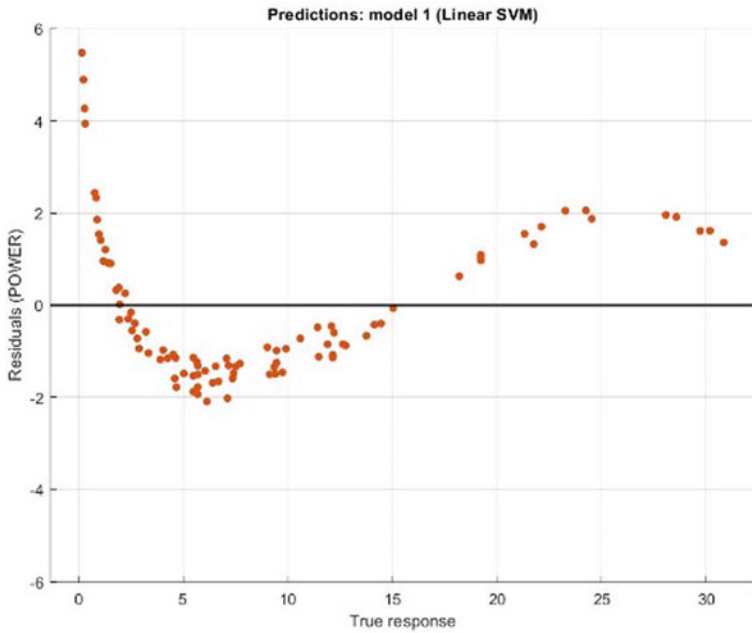


Fig. 7 Residual plot of linear SVM method

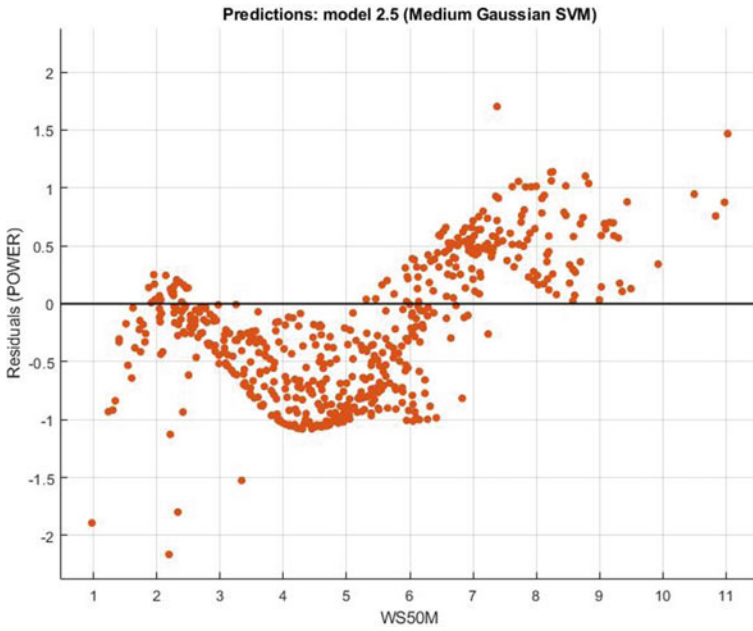


Fig. 8 Residual plot of Gaussian median SVM method

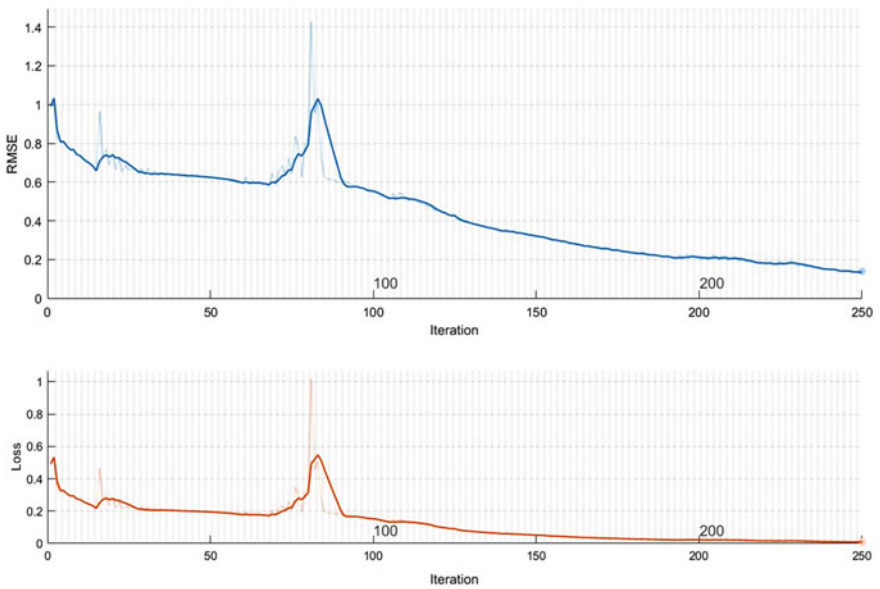


Fig. 9 Training progress of machine learning model with 250 iterations

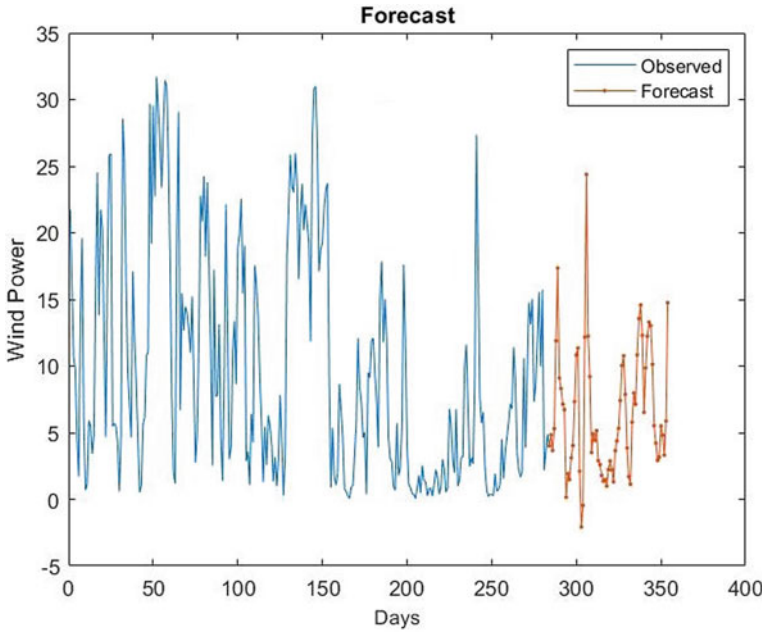


Fig. 10 Forecast plot of the wind power for 30% testing data

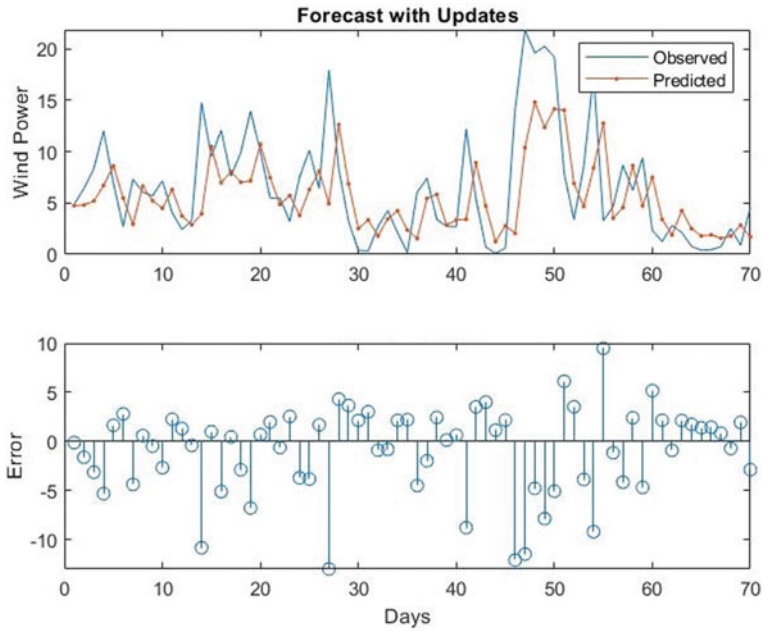


Fig. 11 Forecast with update plot for wind power using LSTM method

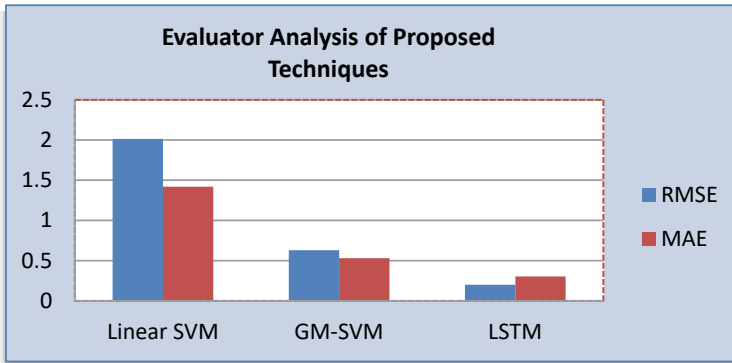


Fig. 12 Comparison graph for the proposed methods and results

Table 2 Comparative analysis of proposed methods

Error	Linear SVM	GMSVM	LSTM
RMSE	2.0135	0.63013	0.2005
MAE	1.4195	0.53114	0.3021

References

1. Saroha S, Aggrawal SK (2015) A review and evaluation of current wind power prediction technologies. *WSEAS Trans Power Syst*
2. Colak I, Sagirolu S, Yesilbudak M (2012) Data mining and wind power prediction: a literature review. *Renew Energy* 46:241–247
3. Chang W-Y (2014) A literature review of wind forecasting methods. *J Power Energy Eng*, 2:161–168. <https://doi.org/10.4236/jpee.2014.24023>
4. Chen K, Yu J (2014) Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. *Appl Energy* 113:690–705
5. Catalao JPS, Pousinho HMI, Mendes VMF (2015) Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal. *IEEE Trans Sustain Energy* 2(1):50–59
6. Potter CW, Negnevitsky M (2006) Very short-term wind forecasting for Tasmanian power generation. *IEEE Trans Power Syst* 21(2):965–972
7. Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos PS (2004) A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Trans Energy Convers* 19(2):352–361
8. Monfared M, Rastegar H, Kojabadi HM (2009) A new strategy for wind speed forecasting using artificial intelligent methods. *Renew Energy* 34:845–848
9. Chang W-Y (2013) Application of back propagation neural network for wind power generation forecasting. *Int J Digital Content Technol Its Appl* 7:502–509
10. Liu D, Niu D, Wang H, Fan L (2014) Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew Energy* 62:592–597
11. Maldonado-Correa J, Solano J, Rojas-Moncayo M (2021) Wind power forecasting: a systematic literature review. *Wind Eng* 45(2):413–426. <https://doi.org/10.1177/0309524X19891672>
12. Saxena N, Ganguli S (2015) Solar and wind power estimation and economic load dispatch using firefly algorithm. *Procedia Comput Sci* 70:688–700
13. Saroha S, Aggarwal S (2015) A review and evaluation of current wind power prediction technologies

14. Saroha S, Aggarwal SK (2018) Wind power forecasting using wavelet transforms and neural networks with tapped delay. *CSEE J Power Energy Syst* 4(2):197–209. <https://doi.org/10.17775/CSEEJPES.2016.00970>
15. Gupta A, Gupta K, Saroha S (2020) Solar irradiation forecasting technologies: a review. *Strateg Plan Energy Environ*, pp. 319–354–319–354. <https://doi.org/10.13052/SPEE1048-4236.391413>
16. Saroha S, Aggarwal SK (2019) Wind power forecasting using wavelet transform and general regression neural network for ontario electricity market. *Recent Adv Electr Electron Eng (Formerly Recent Patents Electr Electron Eng*, 13(1):16–26. <https://doi.org/10.2174/2352096512666190118160604>
17. Singla P, Duhan M, Saroha S (2021) A comprehensive review and analysis of solar forecasting techniques. *Front Energy* 2021:1–37. <https://doi.org/10.1007/S11708-021-0722-7>
18. Syed Ali M, Narayanan G, Saroha S, Priya B, Thakur GK (2021) Finite-time stability analysis of fractional-order memristive fuzzy cellular neural networks with time delay and leakage term. *Math Comput Simul* 185:468–485. <https://doi.org/10.1016/J.MATCOM.2020.12.035>
19. Demolli H, Dokuz AS, Ecemis A, Gokcek M (2019) Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers Manag* 198:111823. <https://doi.org/10.1016/j.enconman.2019.111823>

Driver Drowsiness Detection System



Pratik Mahapatra, Shivam Raj, and Amrita Biswas

Abstract Road accidents are common often leading to serious injuries and often death. There are several causes of road accidents. Fatigue and drowsiness in drivers being one of the prime causes. Being a potential cause for danger on the road, one of the best ways to prevent this is to implement a drowsiness detection system. The proposed work implements a driver drowsiness detection system based on computer vision. A webcam is used to capture the face of the driver, and eye aspect ratio of the driver is used to detect if the driver is sleepy. An alcohol alert module (based on Arduino ethanol gas sensor MQ3) is also included. The system will alert the driver with alarms if the driver is found in drowsy state, and a message will be sent to the owner of the vehicle in case driver is drowsy or drunk. The owner will also be able to remotely monitor the activities of the driver.

Keywords Driver alert · Drowsy · HOG · Facial landmark · Eye aspect ratio · Computer vision

1 Introduction

Aim of this work is to implement a system to detect drowsiness in a driver and set off an alarm and also send message automatically to the owner of the vehicle or concerned authorities.

The current statistics reveal that in 2019 in just India alone, 1,54,732 people died due to road accidents [1]. 21% car accidents were caused due to drowsiness causing drivers to make mistakes. The National Highway Traffic Safety Administration data states each year, 100,000 accidents are due to fatigue in the drivers. Car accidents have led to 1550 deaths, 71,000 injuries and 12.5 billion dollars monetary loss [2]. On the other hand, NHTSA says everyday about 28 people in the USA die in drunk-driving crashes which means one person in every 52 min [3]. In developing countries, fatigue along with poor infrastructure is recipe for disaster. Unlike drugs or alcohol fatigue

P. Mahapatra · S. Raj · A. Biswas (✉)
Sikkim Manipal Institute of Technology, Majitar, India
e-mail: amrita.a@smit.smu.edu.in

is difficult to measure or sense. There are no clear and predefined tests available for fatigue detection.

So probably, the best that can be done is create awareness among the drivers and driver handlers about fatigue-related accidents and encouraging drivers to take a break when they are tired or drowsy. Money motivates drivers to make poor decisions like driving all night even with fatigue. Most drivers are not aware of the enormous risk associated with driving when tired. Some countries have imposed restrictions on the number of hours a driver can drive at a stretch, but it is still not enough to solve this problem as its implementation is very difficult and costly. So one solution to this problem is a driver drowsiness detection system which can be included in the vehicles.

The proposed system detects fatigue based on the eye blinks. The system checks if the eye blinks are more than a certain pre-specified amount or the facial posture is according to some specific positions pre-fed into the system. A driver drowsiness detection system has the potential to save many lives by continuously monitoring and analyzing the driver's eyes and facial expression. If the system detects frequent closing of eyes or yawning the driver could be alerted through loud alarms.

Researchers have tried to develop systems that can be divided into four categories:

- (1) Driving state of the vehicle- Drowsiness detection is based on the fact that when drivers are drowsy and sleepy the vehicle may move away from the center of the lane; thus, these methods take action only after the driver is almost asleep and hence usually very late to prevent accidents [4].
- (2) Physiological indicators like electroencephalogram (EEG) are analyzed for fatigue detection. But physiological sensors are expensive and are usually wearable devices and may not be comfortable for the driver [5].
- (3) Information fusion combines both physiological indicators as well as the driving state of the vehicle. But the driving environment and the drivers manipulation will have a role to play as far as the effectiveness of the system is concerned [5].
- (4) Computer vision-This approach is advantageous because the system works based on the facial features and expressions of the driver. So it is noninvasive and also unaffected by external factors like road conditions, driving state of the vehicle, etc. [5].

Belal Alshaqqaqi et al. [6] model used the visual data and AI to design a driver assistance system to deal with driver drowsiness.

Arun Sahayadhas et al. [7] work deals with vehicle swaying and moving off the path to detect driver drowsiness, but very often this is too late to prevent accidents.

Mehta et al. [8] proposed a real-time driver drowsiness system using eye aspect ratio and eye closure ratio.

2 Proposed Work

First the captured image will be analyzed to check if the image is clear or not and if it is so then it will go through 68 facial landmark detection for detecting and extracting facial features. Thereafter, the facial features are screened to ensure they are clear or not. If the faces are clear, then eye extraction is carried out from the captured image. As the images are being captured in real time in case of any inconsistency (image/facial features are blurry or face/facial features are not available), the image is scrapped and the next image is processed. The flowchart of the proposed work is shown in Fig. 1.

First the captured image will be analyzed to check if the image is clear or not and if it is so then it will go through 68 facial landmark detection for detecting and extracting

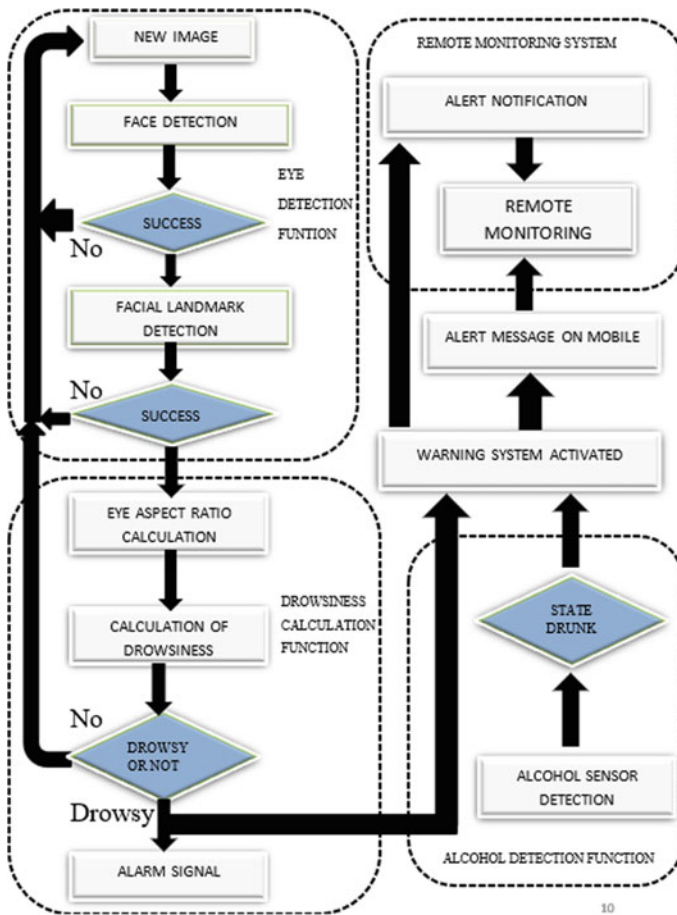


Fig. 1 Flowchart of proposed system

facial features. Thereafter, the facial features are screened to ensure they are clear or not. If the faces are clear, then eye extraction is carried out from the captured image. As the images are being captured in real time in case of any inconsistency (image / facial features are blurry or face/facial features are not available), the image is scrapped and the next image is processed. The flowchart of the proposed work is shown in Fig. 1.

Image that was captured and had its eye portion extracted will have its distance between eyelids calculated after which eye aspect ratio (EAR) will be calculated using a Euclidean formula. After this is done, then the result will be verified through pre-set perimeters whether the person is drowsy. In case the person is not drowsy, the next analyzed image will go through calculation process and inspection. There is also a feature to avoid mis-reading of drowsiness due to eye blinks as well as yawns to achieve better results.

Once the appropriate eye aspect ratio is computed, it is fed into the software for evaluation of drowsiness. The ideal EAR will be different in different lighting conditions and camera positioning and needs to be adjusted suitably.

In case the system is used in a company or transportation agency, the person in charge for that particular vehicle can remotely monitor the driver in real time using the IP address of the system. The video feed generated will be in real and can also be used to determine whether the person is following appropriate guidelines as well as emergency situations can be monitored.

The monitoring can also be done through a mobile phone or a desktop remotely as long as the one who is monitoring has a password and user ID to look into the driver. Also in case of problems warning alarm and notifications will be sent to the person who is monitoring the driver.

When the driver is found in sleepy or in drowsy state, then an alarm will ring to alert the driver. As well as an alert message will be sent to a pre-registered mobile number through a twilio account.

Before the start of travel or somewhere in between at a stoppage in the journey, the person in charge of the driver can also submit a notification for alcohol testing to the driver and the driver will put his mouth toward the sensor, while someone is monitoring remotely, and in case if alcohol is detected, the alarm will ring and a driving prohibition message will be sent.

The working of this system can be divided into two main parts:

1. Face detection
2. Facial landmark prediction.

The input image is processed, and histogram of oriented gradient (HOG) features of the image are extracted.

The facial landmark detector belonging to dlib library of Python is used to estimate the location of 68-(XY) coordinates that maps to the facial structure.

These 68-(x, y) coordinates represent the important regions of the face like mouth, left eyebrow, right eyebrow, left eye, right eye, nose and jaw. Shown in Fig. 2.

However, we only need the coordinates of the left eye and right eye.

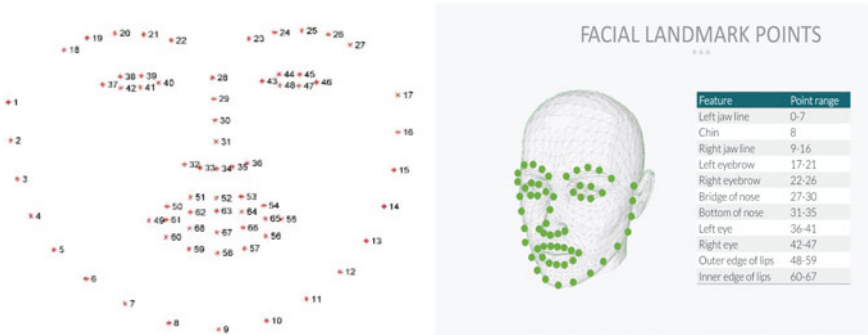


Fig. 2 Facial landmark detection and landmark points

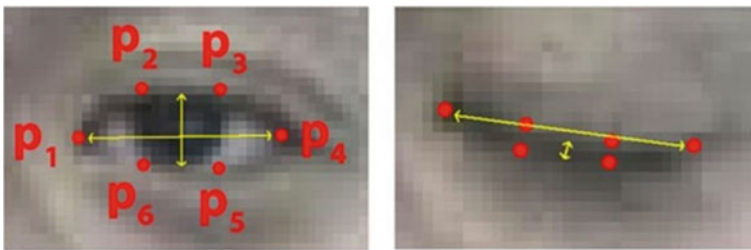


Fig. 3 A visualization of eye landmarks when eye is open (left) and closed (right)

As can be seen in Fig. 2, each eye is represented by a set of 6-(x, y) coordinates beginning at the left corner of the eye and then working clockwise around the rest of the region. Figure 3 shows an open and closed eye and their landmarks.

Eye Aspect Ratio

After obtaining the coordinates of the eyes, the eye aspect ratio (EAR) [ref:]is used to compute the drowsiness. The equation of the EAR is as follows:

$$EAR = \frac{\|p2 - p6\| + \|p3 - p5\|}{2\|p1 - p4\|} \tag{1}$$

Drowsiness Calculation

Once the facial landmark localization (FLL) is done, eye landmarks are used to determine the eye aspect ratio (EAR) to test if a person is drowsy.

A non-intrusive approach has been made, and EAR and FLL parameters detect the actual positioning and avoids the use of a complimentary database.

Accuracy of the detector has been improved by using EAR and FLL. Facial landmark localization integrates EAR and hence improves the accuracy by up to 40%. When the EAR is less than 0.25 (best result according to our test in various

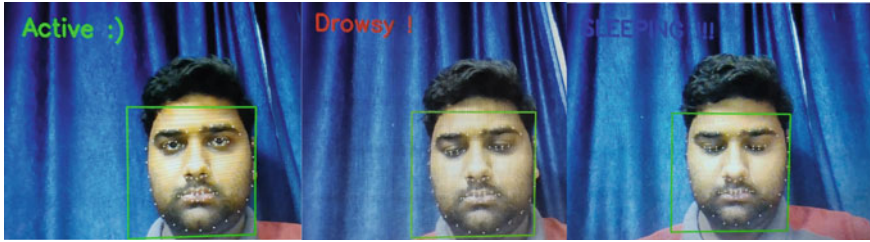


Fig. 4 Demonstration of active, drowsy and sleepy state (left to right). (The image used is that of the author)

lighting conditions) for more than 40 frames (to prevent eye blink false alarms), an alert message will appear on the screen. Along with that an alarm sound will ring to alert the driver, also an alert message will be sent to the owner of the vehicle.

Along with that there is also an alcohol alert module (based on Arduino ethanol gas sensor MQ3) which will send a warning message and alert message to owner of the vehicle or concerned authorities. For messaging an Internet messaging service, website twilio.com, is used for both drowsiness alert and alcohol alert with subscriber module - +14158624511.

The system has been developed and the application implemented in our commercial laptop for prototyping; however, the same can be implemented in Arduino which will significantly reduce the cost to lowest.

3 Results Obtained

The implemented work has been tested on several real-time images and has given accurate results in detecting sleepy eyes and drowsiness. Samples have been shown in Fig. 4.

The software developed has also been tested and is working successfully in transmitting alerts when drowsiness is detected. The snapshots of the software have been shown in Figs. 5 and 6.

4 Conclusion and Future Scope

This project can be incorporated in any vehicle as a low-cost solution for detecting drowsiness in drivers. The results have wide future scope in the automobile industry. Automatic speed control can also be integrated in the car system if driver is detected to be drowsy. This system can be implemented in the form of mobile application to reduce the cost significantly and make it more scalable. Sensors can be added to track the heart rate in order to provide physiology-based monitoring as well as check the

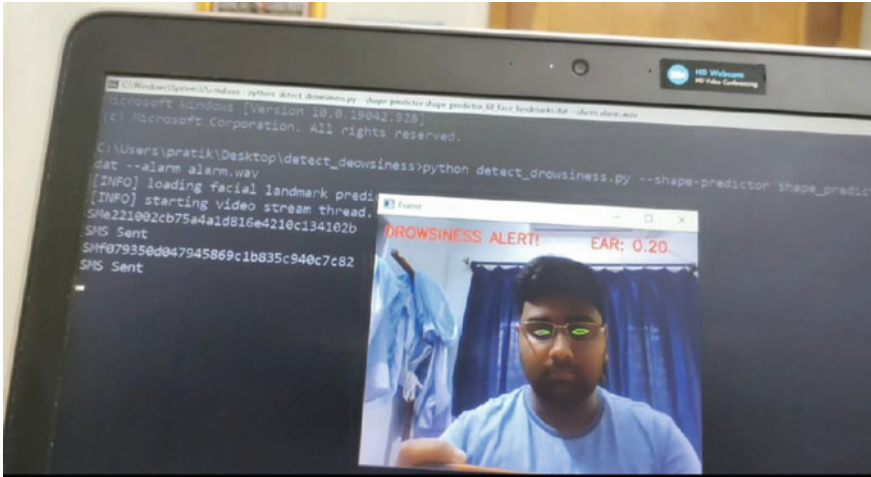
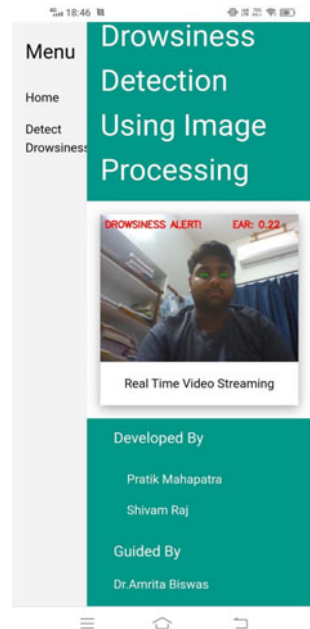


Fig. 5 Final software model (at driver’s end). (The image used is that of the author)

Fig. 6 Final software model (Remote monitoring by owner using mobile phone). (The image used is that of the author)



health conditions of drivers. Similar to alcohol sensor, infrared sensors can also be added to check for covid and other infections among drivers. The technology used here in our project can also be used in OTT platforms and virtual class platforms so that if the user leaves the view of camera or falls asleep, system can detect and close itself/notify.

References

1. Chapter - 1 (ncrb.gov.in): Date:30/10/2021
2. Drowsy Driving—Facts, Causes and Effects (medindia.net), Drowsy driving Statistics 2020—Counting Sheep Sleep Research www.countingsheep.net/drowsy-driving/ Date:30/10/2021
3. Drunk Driving | NHTSA, <https://www.nhtsa.gov/risky-driving/drunk-driving> Date: 30/10/2021
4. Sahayadhas A et al (2012) Detecting driver drowsiness based on sensors: a review. *Sensors* (Basel, Switzerland) 12(12):16937–16953. <https://doi.org/10.3390/s121216937>
5. Zhao Z, Zhou N, Zhang L, Yan H, Xu Y, Zhang Z (2020) Driver fatigue detection based on convolutional neural networks using EM-CNN. *Comput Intell Neurosci* 2020, Article ID 7251280, 11 p. <https://doi.org/10.1155/2020/7251280>
6. Alshaqqaqi B, Baquhaizel AS, Amine Ouis ME, Boumehed M, Ouamri A, Keche M (2013) Driver drowsiness detection system. In: 2013 8th international workshop on systems, signal processing and their applications (WoSSPA), pp 151–155. <https://doi.org/10.1109/WoSSPA.2013.6602353>
7. Sahayadhas A, Sundaraj K, Murugappan M (2012) Detecting driver drowsiness based on sensors: a review. *Sensors* 12(12). <https://doi.org/10.3390/s121216937>
8. Mehta S, Dadhich S, Gumber S, Jadhav Bhatt A (2019) Real-time driver drowsiness detection system using eye aspect ratio and eye closure ratio (March 20, 2019). In: Proceedings of international conference on sustainable computing in science, technology and management (SUSCOM), Amity University Rajasthan, Jaipur—India, 26–28 Feb

A Novel Deep Learning Approach for Non-invasive Blood Glucose Measurement from Photoplethysmography Signals



Gautham Reddy, Karthik K. Bhat, Umang Lunia, and Niranjana Krupa

Abstract In this paper, the authors propose a novel deep learning (DL) approach for the prediction of blood glucose levels (BGL), non-invasively, analyzing photoplethysmography (PPG) signals. Features such as Kaiser–Teager energy, spectral entropy and zero crossing rate were extracted from the PPG signals, and a random forest (RF) model was used to predict the BGL. Later, a convolutional neural network (CNN) model with 25 layers was implemented and trained for the purpose using stochastic gradient descent method. A Bland–Altman (BA) plot and the Clarke error grid (CEG) were used as means to estimate the accuracy of the RF and CNN models. By comparing the CEG and BA plots, it was observed that the accuracy of BGL measurements was superior in the CNN model when compared to the RF model. The CNN model had 84% of the points lying in region-A of the CEG and a standard deviation of errors equal to 15.67 in the BA plot.

Keywords Blood glucose · Photoplethysmography · Convolutional neural networks · Random forests · Clarke error grid

1 Introduction

Blood glucose level (BGL) indicates the concentration of glucose in the blood which is typically measured through chemical tests in a laboratory. Currently, BGL can be measured through an invasive and a non-invasive approach. The invasive approach requires blood samples, from which the BGL is measured through chemical means,

G. Reddy (✉) · K. K. Bhat · U. Lunia · N. Krupa
PES University, Bangalore, India
e-mail: gauthamreddyp@pesu.pes.edu

K. K. Bhat
e-mail: karthikkbhat@pesu.pes.edu

U. Lunia
e-mail: umanglunia@pesu.pes.edu

N. Krupa
e-mail: bnkrupa@pesu.edu

whereas a non-invasive approach does not require a blood sample but the accuracy is not on par with the invasive approach. Measuring the BGL through chemical methods is slow and tedious due to which continuous monitoring of BGL becomes difficult. Hence, it is vital to improve the accuracy of non-invasive techniques in order to continuously monitor BGL. The number of diabetic patients has risen from just 108 million in 1980 to 422 million in 2014 owing to a population increase and a rise in prevalence of the disease [1]. Thus, there is a need to measure BGL quickly and accurately to treat diabetic patients.

Photoplethysmography (PPG) is an optical technique where changes in blood volume are detected. A PPG signal reflects the alterations in volume of the blood vessels by illuminating the skin with a light source and observing the reflected or refracted light [2]. Since glucose in the blood alters the refractive and reflective properties of blood, it is established that a PPG signal is correlated to the BGL in a person [3, 4]. It is necessary to use a machine learning (ML) model to exploit this correlation and predict the BGL from a PPG signal. In this study, we use PPG signals from a database [5] to train two ML models, a random forest (RF) and a 1D convolutional neural network (CNN). The baseline RF model was trained with time and frequency domain features and tested along with the novel CNN model. The outcome of both the models was assessed with the help of the standard Clarke error grid and Bland–Altman methods.

2 Literature Survey

In existing literature, several ML models have been proposed to measure the BGL from PPG signals. Conventionally, ML models were trained with the help of certain features extracted from the PPG signals. These features include both time and frequency domain information [6] such as zero crossing rate [7], spectral entropy, autoregressive coefficients [3], etc. The extracted features are used to train several ML models. Some of the widely used ML techniques like neural networks [5] and support vector machines did not yield good accuracy for clinical purposes. Another approach, the subspace KNN [8] provided good results in classifying patients based on their BGL, but the ML model was dependent on too many feature extraction algorithms. The approach involved redundant features to be extracted. Random forests [3, 4] have been proposed for predicting BGL, and it accurately measures the BGL from PPG signals. Random forest provides the best accuracy in BGL measurement [5]. In this technique, the ML model measures the BGL from features extracted from the PPG signal. However, it is dependent on the feature extraction algorithms used prior to training the random forest. In the recent years, deep learning methods have been employed in diabetes measurement [1]. The advantage of deep learning techniques is that unlike conventional ML models, it does not contain redundant features. Also, it yields better results than conventional methods. A deep learning model was used to measure the BGL by extracting features from the signal in [9], and it provided a more robust solution to measuring the BGL.

3 Methodology

In this study, we implement the conventional random forest machine learning model and compare it with a novel convolutional neural network. An overview of the methodology is shown in Fig. 1.

3.1 Dataset

A dataset of PPG signals along with the BGL at the time of recording the PPG signals was obtained to train a machine learning model capable of predicting the BGL from the PPG signal [5]. In the dataset, PPG signals were recorded with Empatica E4 Wristband, and the BGL was measured through a Cobas-6000 laboratory test and an Accu-Chek device. The Cobas-6000 laboratory test results were used as the true or reference blood glucose values. PPG signals in the dataset were filtered and pre-processed; hence, it was not a necessity to pre-process the PPG signals before training the ML models with them. This dataset consisted of PPG signals of 217 patients and their corresponding BGL measured at the time of recording of the signal. Each PPG signal was split into two components, namely the training and testing split. The training split consisted of PPG signal samples of each patient from sample number 500–4500. The testing split consisted of samples of each patient from sample number 5000–9000. Patients with insufficient samples to perform this split were discarded. Training split was used specifically to train the machine learning (ML) models, while testing split was used to measure the BGL and test the model’s accuracy. If the ML model works perfectly, the BGL value predicted for the testing split must be the same

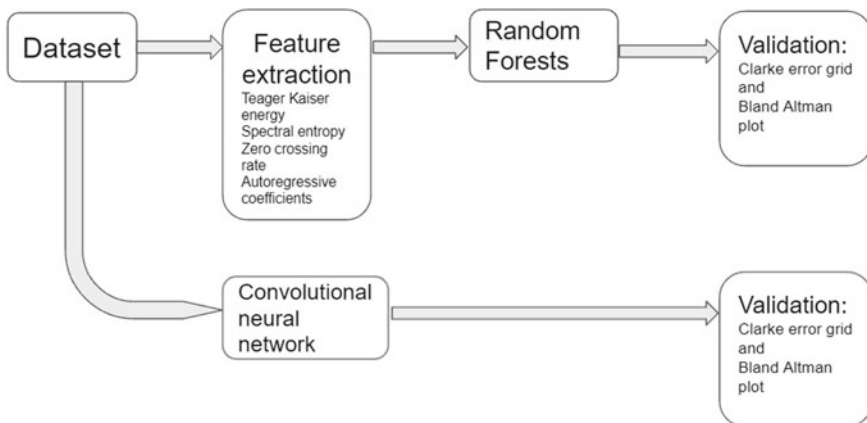


Fig. 1 Overview of the methodology for training and testing of the ML models

as the training split in the most ideal case. This is because all the 9000 samples were recorded from the same patient at the same time, due to which the same BGL must be predicted by the ML model.

3.2 Feature Extraction

A PPG signal's features or characteristics reflect biological changes within the body such as change in heart rate, breathing rate or blood pressure. Therefore, by obtaining the signal features one can determine various biological parameters that the features depend on. In the training split of the dataset, 3840 samples of the PPG signal samples (called as window from here) were segmented into frames of 320 samples. Thus, a 3840-sample window was segmented into frames of 320 sample length. Twenty-two features were extracted from these frames. Some of the important features are described. Table 1 provides all the features used for training.

Heart Rate The time interval between alternate zero crossings is known as the heart rate statistic [3]. Diabetes alters the relationship between the actual heart rate and the calculated heart rate. The mean of HR over an entire window calculated for each frame is used as the feature.

Spectral entropy Spectral entropy (H) determines the randomness associated with a probability distribution. Spectral entropy thus intuitively signifies the 'flatness' of the spectrum. A flatter spectrum will give a larger spectral entropy value, while a spiky spectrum (a spectrum with impulses) will give a lower value for spectral entropy [6].

Teager–Kaiser energy TKE of a signal can be used to represent the modulations present in the signal. The formula for KTE of a signal $x(n)$ for the $(n - 1)$ th frame

Table 1 Twenty-two features extracted from the signal

Feature	Description
Teager–Kaiser energy (TKE)	Average TKE of each frame was calculated. Four features were extracted from TKE. Mean, variance, interquartile range and skewness of all the frames were used as features
Heart rate (HR)	HR over the entire window was used as one feature
Spectral entropy (H)	Spectral entropy for each frame was calculated. Four features were extracted from spectral entropy. Mean, variance, interquartile range and skewness of all the frames were used as features
Zero crossing rate (ZR)	ZR of each frame was calculated. Four features were extracted from ZCR. Mean, variance, interquartile range and skewness of all the frames were used as features
Autoregressive coefficients (AR)	5th-order AR coefficients were calculated for the entire window of the PPG signal. Each of the five coefficients was a separate feature

is as follows:

$$\text{TKE}_{n-1}(x) = x[n]^2 - x[n-1]x[n+1] \quad (1)$$

Autoregressive coefficients Mathematically, AR coefficients (auto regressive coefficients) are numbers that can be used to predict the next value of the sample by using the previous values of the samples. The number of past sample values that it uses is equal to the order of the AR model. These coefficients can be calculated using the Yule–Walker equations. AR coefficients represent the envelope of the spectrum which is an important feature for the machine learning model. AR coefficients of order 5 were obtained through Yule–Walker equations for the entire PPG window [10].

3.3 *Baseline Random Forest Model*

Two machine learning models were trained and tested, namely the random forest (RF) model and convolutional neural network (CNN) for BGL prediction. RF is a popular ensemble learning technique that can be used for both regressions and classification. Ensemble learning models are those models which combine weaker learning elements to make them into a stronger learning model [11]. A random forest ML model with 500 decision trees was used for regression to measure the BGL. The input to this ML model consists of features as shown in Table 1 obtained from the PPG signal. These features in Table 1 were extracted from the PPG signal of each patient in the training split of the dataset and stored in feature vectors. The feature vectors were used for training the RF model.

3.4 *Blood Glucose Prediction Using 1D-CNN*

The key difference between RF and CNN is that RF requires feature extraction, that is, feature extraction must be performed separately. However, in CNN, the feature extraction is done by the network itself. CNN is a deep learning model which can be used for feature extraction and regression. For prediction of the BGL using PPG signal, a 1D CNN was implemented since a 1D CNN is ideal for a time-series signal. Since the PPG signal varies with the presence of glucose in the blood, one can provide a PPG signal window as input to a CNN model and train it to measure the BGL [12]. The PPG signal window from the training split was further segmented into smaller frames of 320 samples in length. Table 2 shows all the types of layers of the CNN. The CNN was trained using a stochastic gradient descent algorithm.

Convolution Layers Convolution layers were used to extract useful information from the signal. Similar to the use of convolution layers in images, convolution layers were used to perform feature extraction [12]. A convolution filter in a 2D

Table 2 Layers present in the CNN along with a description

Layer numbers	Layer type	Description
1	Image input	$320 \times 1 \times 1$ images with 'z-score' normalization
2, 6, 10, 14	Convolution layer	Convolution layer to extract signal information. Ten $3 \times 1 \times 1$ convolutions take place. Repeated convolution layers to extract more information from the signal
3, 7, 11, 15	Batch normalization	Batch normalization with 10 channels. It normalizes the signal to ensure zero mean and unit variance
4, 8	Max pooling	2×1 max pooling was done to reduce discretize the data or down-sample the data. Max pooling refers to selecting the maximum value within the 2×1 window
5, 9, 13, 18, 20	ReLU	ReLU activation function was used to remove negative activations. The ReLU function forces all negative activations to zero
12, 16	Average pooling	2×1 average pooling to down-sample the data. Average pooling refers to replacing selecting the average of the elements inside the window
17, 19	Fully connected	2 fully connected layers for regression. Layer is used specifically for prediction of BGL
21	Regression output	The final BGL output for 320 samples of a PPG signal

convolution layer contains a height and a width. Since this is a 1D convolution layer, it only contains one dimension called length. A smaller convolution window length is preferred as it helps with better feature extraction [13]. Thus, convolution window length was chosen to be 3. The output vector has a different length from the input. There was no padding added to the output of a convolutional layer. The number of filters was chosen to be 20 for each convolution. The shift of the convolution window is called as a stride, and a unit stride length was chosen for all convolution layers. The information extracted from convolution layers was further enhanced by other layers such as max pooling and average pooling layers.

Other Layers The other layers are used to enhance the representation of the signal for the convolution layers. ReLU or rectified linear activation unit is commonly used as an activation function in a 1D CNN. ReLU or rectified linear activation function was used to ensure that the output had only positive values [14]. Batch normalization layers were also added to ensure zero mean and unit variance in the signal [14]. The fully connected layers were trained to use the features extracted by the convolution layers and estimate the BGL. The regression for BGL measurement was performed by the fully connected layers. Therefore, feature extraction is performed by the convolution layers, whereas the fully connected layers were neural networks that used the extracted features to measure the BGL.

3.5 Model Validation

The accuracy of the ML models trained must be established. These models must also be compared with each other. In order to validate these models, a Clarke error grid and a Bland–Altman plot was created for each model. These plots and comparison metrics are explained below.

Clarke Error Grid The Clarke error grid (CEG) is specifically designed to determine the accuracy of a BGL measurement from the perspective of clinical treatments. The plot is divided into five regions, namely region A, B, C, D and E. If a measurement lies in region-A, it means that the measurement is clinically valid, i.e., the error caused in the measurement will not lead to incorrect treatment for hypoglycemia or hyperglycemia. The details of each region are mentioned below:

- (1) Region A: Does not affect the clinical treatment. Less than 20% error
- (2) Region B: Has considerable amount of error, but does not lead to harmful treatment.
- (3) Region C: Leads to no severe, but unnecessary treatment.
- (4) Region D and Region E: May lead to dangerous treatment as hyperglycemia and hypoglycemia treatment might be interchanged [15].

Bland–Altman Plot The Bland–Altman plot is used in biomedical engineering to check for the agreement between two methods of measuring a given quantity. It is a plot between the differences between method-A and method-B on the Y-axis and the average of method-A and method-B on the X-axis. Bland–Altman’s plot defines the interval of agreement or the lines of agreement. The Bland–Altman (BA) plot contains the average difference line, the $d - 2s$ line and the $d + 2s$ line where d is the average difference line and s is the standard deviation of the differences. For a 95% confidence interval, 95% or more points will lie between $d - 1.9s$ and $d + 1.9s$ [16]. A 95% confidence interval is defined as the region between the lines $d-2s$ and $d + 2s$.

4 Results

Table 3 shows the results of the CEG analysis and the percentage of points within each region. It was observed that the RF model had 83.16% of the points within region-A of the CEG but CNN performed better with 84.12% of the points within region-A. Each model had 1.57% of the data points within region-D which might be due to abnormally high BGL or erroneous data points. Thus, the CNN had a larger number of points in region-A compared to the RF model. This implies that the CNN had greater accuracy when compared to the RF model as per the CEG.

Another method of validation method used was the Bland–Altman (BA) plot. From Table.4, we can observe that the standard deviation of error of CNN is 15.67 compared to that of the RF model which is at 15.83. Since a smaller mean and

Table 3 Results of the Clarke error grid analysis

Region	RF model (%)	CNN model (%)	Interpretation: Error will cause-
A	83.16	84.21	No effect on clinical treatment
B	15.26	14.2	No harmful treatment
C	0	0	Unnecessary treatment
D	1.57	1.57	Potentially dangerous treatment
E	0	0	Dangerous treatment

The percentage of points present in each region is shown as entries in the cells

Table 4 Results of the Bland–Altman plot

Parameter	Random Forest	CNN
Mean difference	−2.60	−0.25
Standard deviation of error	15.83	15.67

The mean difference and standard deviation of the errors are shown

standard deviation of error is shown in the BA plot of the CNN, the CNN is more accurate when compared to the RF model. Thus, an improvement in accuracy was seen from existing literature.

RF Models that are the most accurate models are used to measure BGL with a mean absolute error of 8.59 [9]. Although a direct comparison cannot be made, our novel CNN has 84% of the data points lying within region-A with the same dataset being used.

5 Conclusion

From the results shown earlier, we can conclude that the CNN model performs better for BGL measurement from PPG signals with 84% of the points lying within region-A of the CEG. It is also evident that BGL is correlated to the PPG signal of an individual. The error was observed to be larger as the BGL value to be predicted increased. This can be attributed to a low number of data points with high BGL. Thus, it is possible to measure BGL from PPG signals. A more varied dataset would

yield better results since the number of data points with high BGL was less. However, from the BA plot and the CEG, it is evident that the machine learning model performs satisfactorily.

6 Future Scope

The performance of the CNN and RF model can be verified by the Bland–Altman plot and Clarke error grid, which show satisfactory results. However, these models were tested on a dataset of PPG signals which contained a smaller number of data points with BGL less than 150 mg/dl when compared to those data points greater than 150 mg/dl. Therefore, the models do not accurately measure BGL for very high glucose levels. By training the models with a more varied set of data points, the performance and accuracy can be improved. The speed of predicting the BGL was not measured or tested for both the RF and CNN models. Due to a large number of layers that must be executed in the CNN, it will be slow in measuring the BGL. These factors can be explored in further studies.

References

1. Deshpande AD, Harris-Hayes M, Schootman M (2008) Epidemiology of diabetes and diabetes-related complications. *Phys Ther* 88(11):1254–1264
2. Elgendi M (2012) On the analysis of fingertip photoplethysmogram signals. *Curr Cardiol Rev* 8(1):14–25
3. Monte-Moreno E (2011) Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artif Intell Med* 53(2):127–138
4. Habbu S, Dale M, Ghongade R (2019) Estimation of blood glucose by non-invasive method using photoplethysmography. *Sadhanā*, 44(6):1–14
5. Salamea C, Narvaez E, Montalvo M (2019) Database proposal for correlation of glucose and photoplethysmography signals. In: *The international conference on advances in emerging trends and technologies*. Springer, 2019, pp 44–53
6. Ramasahayam S, Koppuravuri SH, Arora L et al (2015) Noninvasive blood glucose sensing using near infra-red spectroscopy and artificial neural networks based on inverse delayed function model of neuron. *J Med Syst* 39:166
7. Boashash B (2015) *Time-frequency signal analysis and processing: a comprehensive reference*. Academic Press
8. Jalil M, Butt FA, Malik A (2013) Short-time energy, magnitude, zero crossing rate and auto-correlation measurement for discriminating voiced and unvoiced segments of speech signals. In: *2013 The international conference on technological advances in electrical, electronics and computer engineering (TAEECE)*. IEEE, 2013, pp 208–212
9. Zhang Y, Zhang Y, Siddiqui SA, Kos A (2019) Non-invasive blood glucose estimation using smartphone ppg signals and subspace knn classifier. *Elektrotehniski Vestnik* 86(1/2):68–74
10. Deng H, Zhang L, Xie Y, Mo S (2021) Research on estimation of blood glucose based on PPG and deep neural networks

11. Chan Y, Langford R (1982) Spectral estimation via the high-order Yule-Walker equations. *IEEE Trans Acoust Speech Signal Process* 30(5):689–698. <https://doi.org/10.1109/TASSP.1982.1163946>
12. Latinne P, Debeir O, Decaestecker C (2001) Limiting the number of trees in random forests. In: *International workshop on multiple classifier systems*. Springer, 2001, pp 178–187
13. Goh C-H, Tan LK, Lovell NH, Ng S-C, Tan MP, Lim E (2020) Robust ppg motion artifact detection using a 1-d convolution neural network. *Comput Methods Programs Biomed* 196:105596
14. Zhu T, Li K, Herrero P, Georgiou P (2021) Deep learning for diabetes: a systematic review. *IEEE J Biomed Health Inform* 25(7):2744–2757. <https://doi.org/10.1109/JBHI.2020.3040225>
15. Gopal Krishna Patro S, Sahu KK (2015) Normalization: a preprocessing stage. *arXiv e-prints*, pp. arXiv–1503
16. Kovatchev BP, Wakeman CA, Breton MD, Kost GJ, Louie RF, Tran NK, Klonoff DC (2014) Computing the surveillance error grid analysis: procedure and examples. *J Diabetes Sci Technol* 8(4):673–684

Machine Learning Approaches on Intrusion Detection System: A Holistic Review



Pinakshi De and Ira Nath

Abstract With the remarkable development of the Internet over the last few decades, network security is one of the major issues in this century. With the increasing development of malicious software or malware as well as extensive use of the Internet, the destruction and unauthorized access of the network security are increasingly vital. An intrusion detection system is basically implemented to detect the intrusion in the system and identify the different types of unauthorized access of data and information over the networks. However, to understand the research work on IDS, the survey is made on 40 papers from 2010 to 2021. The survey paper includes a basic idea of machine learning, datasets and the different algorithm used in IDS. After studying different existing approaches of IDS techniques, limitations and complexities are mentioned here.

Keywords Intrusion detection · IoT · Machine learning · Classifier · Datasets

1 Introduction

Internet is an architecture that developed in the 1970s and revolutionized data transmission among various computer networks globally. It becomes the most likely information resource about everything in the world. It becomes a major part of the IT industry, banking sector, education, health care, etc. Therefore, all the data and information must be secured. Nowadays, Internet has become an important resource for computing. Interconnectivity between machines, humans and services increased rapidly in the last decade and an advanced architecture called the Internet of Things (IoT) [1]. IoT system is a collection of compatible computing resources and manpower. It is an Internet and service-oriented footstep with the current generation (3G/4G) and next-level generation (5G/6G or beyond). Huge data are used on

P. De (✉)

The Calcutta Anglo Gujarati College, Kolkata 700001, India

e-mail: pinakshide@gmail.com

I. Nath

JIS College of Engineering, Nadia 741235, India

Internet as well as IoT also. Data are utilized in public as well as private. So, huge electronic data feel the problem of authentication, confidentiality and repudiation. We can say that not a single system is 100% secured. In addition, we can say that almost every computer system can experience different modes of attack. That is why Internet security is one of the important issues. It is very much important to design a system in such a way so that the data should be in the secured mode and authorized users can handle those data safely. But, a fully secured system is unrealistic. Network administrators are using intrusion detection system (IDS) [2] to identify different intrusions and roll back the mislaying at that moment or later. The purpose of IDS is classified into anomaly-based detection and signature/misuse-based detection. An anomaly detection method is used to detect computer and network attacks by rules, whereas signature method used a pattern to detect any type of intrusion. This technique used the known pattern to find the attack but it is failed for unknown intrusions. In signature-based IDS, environment detection rate is high and false alarm rate is low for similar pattern, whereas its detection capacity is low for an undetermined attack. But for the anomaly detection method [3], the detection rate and false alarm rate are both high.

To a great extent, significant works have been done on IDS. This survey paper is written after the corresponding analysis on IDS for the past few years. This paper enrolled all IDS-associated 40 papers from the year 2010–2021. This flow of paper is as follows: Sect. 2 details an overview of literature review regarding different techniques used for the intrusion detection systems. Sections 3 and 4 represent a statistical comparison on the algorithms and datasets for each experiment. Section 5 represents different datasets with discussion and conclusion along with future directions in IDS system using machine learning methods.

2 Literature Review

2.1 *Machine Learning (ML) Concept*

ML [4] is a data analytics application of artificial intelligence (AI) whose primary objective is to allow the system to learn automatically without human assistance. Machine learning is a subset of AI, and the primary goal of machine learning is to increase the accuracy of the algorithm after analysing the previous pattern to do specific tasks by using experience, which is derived from the training dataset [5, 6]. Training data are a set of collection sample data that are recorded from past experience or examples. Learning from this experience, machine learning can build mathematical tools for future prediction a decision-making.

Depending on the learning-feedback mechanism, machine learning approaches are classified into three categories. This categorization is important to understand the problem, analyse the data and build a proper algorithm to solve the problem.

2.1.1 Supervised Learning

The name indicates that there is a supervisor. In supervised learning [4, 7], the machine deals with labelled data, which signifies that input data are mapped with some output data. Thus, the machine learns with labelled training data and they apply the knowledge to give a new dataset.

Supervised learning approach is divided into two types: classification—in this problem, the output data variable is categorized such as ‘Yes–No’ and ‘True–False’ and regression—in this case, there is a mapping between the input data and the output data variable of a real-time entity such as length and temperature.

Several algorithms are used in supervised learning which are linear regression, K-nearest neighbour algorithm [8], multilayer perceptron [9], Naïve Bayes [10] (simple Bayes or independence), logistic regression [11] and decision trees (ID3, C 4.5, MARS) [12].

2.1.2 Unsupervised Learning

In this learning approach, different algorithms are used to draw pattern to categorized unlabelled data without any supervision. In this learning, the system analyses the data and can conclude the concealed structure from uncategorized data.

Unsupervised learning is categorized into the following: clustering—where objects are grouping into clusters according to their common properties and association—it is used to find the relationship between variables in the database.

The algorithm which we used in unsupervised learning is Apriori algorithm [13], hierarchal clustering, KNN [8], K-means clustering, neural networks and singular value decomposition.

2.1.3 Reinforcement Learning

In this learning approach, agents interact with the environment by giving actions and find errors or rewards. The trial and error method is used in this approach, and direct feedback helps to find which actions are best. Different models are used in reinforcement learning such as Q-learning, State-Action-Reward-State-Action (SARSA), Deep-Q-Net (DQN) [14] and Deep Deterministic Policy Gradient (DDPG).

2.2 Classifier

In machine learning, the classifier is an algorithm that classifies the data into one or more classes. It is a rule that is used by the trained machine to classify data. It can handle both supervised and unsupervised learning. A different algorithm is used in the classifier to perform the classification of data which are as follows.

2.2.1 Decision Tree

It used a supervised learning approach to build a tree-like model. In this algorithm, data are classified and represented by different parts of the tree such as a trunk, branches and leaves.

2.2.2 Naïve Bayes

This algorithm defines the probability regarding the classification of the given data that means in which group(s) it will be categorized.

2.2.3 K-Nearest Neighbour

It is a pattern recognition algorithm that used trained data for mapping one data with other data in a K-dimensional area.

2.2.4 Support Vector Machine [15]

It creates a three-dimensional model after classifying trained data. It gives the perfect machine learning model for its multidimensional features.

3 Comparison of Related Work

According to the survey, it is found that in the latest decade, for better intrusion detection systems, machine learning has been used but lots of improvement are needed through taxonomy and datasets. In these areas, development is needed firstly in the dataset area. Because huge work is done on KDD-cup99 [16] or DARPA 1999 [6] dataset, there is not any clear idea which data mining technique is more considerable. Secondly, based on the nature of instruction, the structure of IDS is not considered during the evolution of IDS techniques [17]. According to the detection methods, IDS is classified into anomaly based as well as signature based. The complexity of both methods is discussed in Table 1 along with what type of machine learning techniques are used.

After going through all the surveys, major consideration are as follows:

- Explanation of different IDS with different types of attack [26].
- Discussion about different available IDS datasets with their challenges [3].
- Thorough explanation of IDS evolution methods with feature selection.

After studying several research papers regarding intrusion detection system (IDS), we found that different algorithms such as Naïve Bayes, Support Vector Machine

Table 1 Machine learning for different IDS approach including data

Survey	Year	Intrusion detection system		Machine learning			Datasets
		Signature based	Anomaly based	Supervised	Unsupervised	Reinforcement	
Einy [18]	2003	Used	NA	Used	NA	NA	NA
Vinod Kumar [19]	2012	Used	NA	Used	NA	NA	NA
Jarah [20]	2014	Used	NA	Used	NA	NA	Used KDD 1999
Maseer [21]	2011	Used	NA	Used	NA	NA	Used CICIDS 2017
Zhao [22]	2017	NA	Used	Used	Used	NA	Used KDD CUP 1999
A budhammed [23]	2018	NA	Used	Used	Used	NA	Used CIDSS-001
Farzaneh [24]	2019	NA	Used	Used	Used	NA	NA
Kavitha [25]	2021	NA	Used	Used	Used	NA	Used NSL-KDD CUP 1999

(SVM), decision tree and ANN have been used among these algorithms, and it is found that the most fashionable classifiers are Support Vector Machine (SVM) and Artificial Neural Network (ANN), whereas fuzzy logic is less acceptable in the case of single learning classifiers. The detailed survey on the algorithm is depicted in Table 2.

4 Various Intrusion Detection System

Four types of IDS method are used in cloud environment, namely host-based intrusion detection system (HIDS) [41], network-based intrusion detection system (NIDS) [42], hypervisor cloud based intrusion detection system (HCIDS) [43] and distributed-based intrusion detection system (DIDS) [44].

4.1 HIDS

Host-based intrusion detection system methods are used on information which are collected from different sources. The main function of HIDS is that it analyses different activities on source in detail and also identifies the malicious activities of individual processes/users. It is basically used to analyse different functionality on a specific host. HIDS cannot be used in case of denial of service attack because in this case collection of information is held outside of the machine. In case of HIDS, a signature or string-matching method [45] is used to identify threats but nowadays a simple tool can be used to defeat signature-based mapping.

4.2 NIDS

It is a security hardware which is placed to control critical network traffic. In this method, new network packet is analysed with known pattern of attack [46]. It is an advanced and expensive method within the organization environment to detect cyber-attack, DoS and malware. Organizations collectively use different techniques and detection methods to prevent multiple cyber-attacks. NIDS cannot be implemented for weak identification and authentication in network. Signature based IDS and NIDS have high latency for detection of new attack and implementation of signature. Encrypted packets [47] also cannot be processed by NIDS.

Table 2 Reference research title regarding different algorithm

References	Year	Research title	Algorithm used
Almseidin [27]	2017	Evaluation of machine learning algorithm for intrusion detection system	Naïve Bayes
Almansob [28]	2017	Addressing challenge for intrusion detection system using Naïve Bayes and PCA algorithm	
Aksu [29]	2018	Detecting port scan attempts with comparative analysis Of deep learning and Support Vector Machine algorithm	Support Vector Machine
Boero [30]	2017	Support Vector Machine meets software defined networking in intrusion detection system	
Ahmin [31]	2019	A novel hierarchal intrusion detection system based on decision tree and rules-based models	Decision tree
Reddy [32]	2020	Analysis of dimensionality reduction techniques on big data	
Yang [33]	2019	Design of intrusion detection system for IoT based on improved BP Neural Network	Artificial neural network
Xu [34]	2018	An intrusion detection system using a Deep Neural Network with gated recurrence units	
Alsirhani [35]	2019	DDoS detection system: using a set of classification algorithms controlled by fuzzy logic system in Apache Spark	Fuzzy logic
Novaes [36]	2020	Long short-term memory and fuzzy logic for anomaly detection and mitigation in software defined network environment	
Malhotra [37]	2017	Genetic programming and K-nearest neighbour classifier-based intrusion detection model	K-nearest neighbour

(continued)

Table 2 (continued)

References	Year	Research title	Algorithm used
Salih [38]	2019	Combining best features selection using three classifiers in intrusion detection system	
Wang Dawei [39]	2011	Evolving boundary detector for anomaly detection	Detector Generation Algorithm
Mario Poggiolini [40]	2013	Application of the feature-detection rule to the negative selection algorithm	Negative selection algorithm

4.3 HIDS

It is used to run virtual machines hypervisor-based IDS on virtual machine to monitor different communication between virtual machines, hypervisor and virtual machine. The main advantage of HIDS [48] is that it uses only the available information. It is basically used to detect intrusion in virtual environment. HIDS method is used for signature detection, lie detection and row socket detection [49].

4.4 DIDS

It consists of several IDS such as HIDS and NIDS over WAN. It collects all information and transfers into a specific format to pass into the Central Analyzer [50]. It can be used for unknown as well as known attacks.

5 Various Intrusion Detection Techniques

Intrusion detection system basically uses different techniques such as signature-based intrusion detection system (SIDS), anomaly-based intrusion detection system (AIDS), ANN-based detection, fuzzy logic-based detection, SVM-based detection, GA-based detection and many more.

5.1 SIDS

SIDS is basically used for known attacks where pattern is known; it is also known as knowledge-based detection system. In this technique, when pattern is in our database

and it is matched during attack, alarm is generated. SIDS gives accurate results for known attacks but it is not working for zero-day attack [51].

5.2 AIDS

Another method is AIDS which is basically used to overcome the disadvantage of SIDS. AIDS is based on different techniques such as statistics based, knowledge based [52] and machine learning based. It can be used to detect new attack by creating intrusion signature. But the demerits are that it cannot be used for encrypted data because it creates false alarm.

5.3 ANN-Based IDS [53]

It is a pattern recognition technique. It acts like a neuron because all parameters are mapping with different host, among which some are normal and some are malicious. By ANN method, a multilayer perceptron is used to identify attack.

5.4 Fuzzy Logic-Based IDS [54]

It is powerful technique to identify uncertain reasoning attack in network security.

5.5 Association-Based IDS [55]

In this approach, missing algorithm is used for feature selection.

In the following Table 3, comparison between different IDS technologies has been discussed.

6 Intrusion Detection Datasets

Datasets are assigned for particular tasks such as classification, clustering and regression. To validate intrusion detection system (IDS) approach, datasets play a vital role. Datasets that are used to build and evaluate intrusion detection system (IDS) are explained in Table 4.

Table 3 Comparison of different IDS technologies

Intrusion detection techniques	Merits	Demerits
Signature-based detection	<ul style="list-style-type: none"> ✓ Detect intrusion by finding pattern from previously known database ✓ High detection accuracy rate ✓ Computational cost is less 	<ul style="list-style-type: none"> ✓ False alarm rate is high ✓ Cannot identify new or unknown attack ✓ Huge number of signatures has to be stored ✓ Failed to detect multistep attack
Anomaly-based detection	<ul style="list-style-type: none"> ✓ Use statistical tested data for intrusion detection ✓ False alarm rate is low for unknown attack 	<ul style="list-style-type: none"> ✓ Computational time is high to detect attack ✓ Limited detection accuracy feature
ANN-based detection	<ul style="list-style-type: none"> ✓ Detection accuracy rate is high ✓ Identify and categorized undefined network packet ✓ Efficiency of classification is done by multiple layers 	<ul style="list-style-type: none"> ✓ More time is found in training phase ✓ Less flexible in nature
Fuzzy logic-based detection	<ul style="list-style-type: none"> ✓ Efficient for quantitative data ✓ Better flexible for unknown problem 	<ul style="list-style-type: none"> ✓ Detection rate is lower in comparison with ANN
Association rules-based detection	<ul style="list-style-type: none"> ✓ Applicable for signature-based attack 	<ul style="list-style-type: none"> ✓ Not suitable for unknown attack ✓ Used only in case of misuse detection
SVM-based detection	<ul style="list-style-type: none"> ✓ It can classify intrusion efficiently for limited data ✓ It can control massive feature 	<ul style="list-style-type: none"> ✓ Cannot classify group or continuous of features
GA-based detection	<ul style="list-style-type: none"> ✓ Has better efficiency ✓ Suitable for specific case of detection 	<ul style="list-style-type: none"> ✓ Complex in nature
Hybrid-based detection	<ul style="list-style-type: none"> ✓ Efficient approach to classify data 	<ul style="list-style-type: none"> ✓ High computational cost

7 Feature Selection

The prime step for the upgradation of system performance, it is used to reduce computational complexity, remove data redundancy, increasing detection rate in the machine learning approach, and decrease false rate alarm. During the development of a predictive model, feature selection helps to diminish the number of input variables. As per taxonomic view, the following methods are used for feature selection:

Table 4 Description of different datasets

Datasets	Description
AWID [19]	Publicly available and works on 802.11 networks
Booters	There are DDoS attacks, and individual packets are not labelled
Botnet	It is a mapping of present datasets with the IP address for neural networks
CICIDS 2017	It was created in closed environment and bidirectional packet-based format in nature
DARPA	DARPA 98/99 are most popular datasets used to find volume of a network traffic
KDD CUP 99	It is a kind of DARPA datasets which is standard in nature to build predictive model
NSL-KDD [20]	This dataset is extended form of KDD and eliminate redundancy from KDD datasets
SANTA	Deals with updated data to find network traffic and intrusion which is similar to NetFlow
UNSW-NB15	It contains nine different attacks to detect varieties of intrusion

7.1 Filter Methods [27]

It used intrinsic features of univariate statistics by using the chi-square test, Fisher's score, correlation coefficient and dispersion ratio. It is comparatively faster and less costly concerning wrapper methods.

7.2 Wrapper Methods [30]

It follows a greedy search approach to access the quality of learning of all subsets and evaluates the classifier. It gives more accurate results in comparison to the filter approach. It follows forward, backward, recursive and exhaustive feature selection.

7.3 Embedded Methods [30]

It is a combination of all benefits of both the wrapper and filter methods and also provides less computational cost. It follows LASSO Regularization, Random Forest techniques.

8 Discussion, Future Direction and Conclusion

Different techniques of machine learning and AI are used to identify intrusions for maintain consistency in computer network security. Cybercriminal uses different techniques such as fragmentation, floody, obfuscation and encryption to avoid IDS. Lots of work are done in IDS but many essential things remain untouched. IDS has to be more accurate to detect and avoid the various range of intrusion with a less false alarm. The topmost challenge of IDS is to detect attacks marked by evasion techniques. The robustness of IDS regarding evasion methods needs further study.

In this paper, I have presented a detailed survey on IDS methodologies, algorithms and datasets with basic ideas of machine learning. However, these are such approaches that are facing problems during generating and updating information regarding detection rate and false alarm rate. My objective is to find suitable machine learning methods to draw IDS framework to increase accuracy, to reduce false alarm and to reduce training time for intrusion detection. Also, the acquiring techniques, output and drawbacks of different open source datasets, which are used in IDS, have been mentioned over here.

References

1. Raza S, Wallgren L, Voigt T (2013) SVELTE: real-time intrusion detection in the Internet of Things. *Ad hoc Netw* 11(8):2661–2674
2. Abubakar A, Pranggono B (2017) Machine learning based intrusion detection system for software defined networks. In: 2017 seventh international conference on emerging security technologies (EST). IEEE, 2017
3. Khraisat A, Gondal I, Vamplew P (2019) Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity* 2:20
4. Haq NF et al (2015) Application of machine learning approaches in intrusion detection system: a survey. *IJARAI-Int J Adv Res Artif Intell* 4(3):9–18
5. Dhanabal L, Shantharajah SP (2015) A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. *Int J Adv Res Comput Commun Eng* 4(6):446–452
6. Thomas C, Sharma V, Balakrishnan N (2008) Usefulness of DARPA dataset for intrusion detection system evaluation. In: Data mining, intrusion detection, information assurance, and data networks security 2008, vol 6973. International Society for Optics and Photonics
7. Osareh A, Shadgar B (2008) Intrusion detection in computer networks based on machine learning algorithms. *Int J Comput Sci Netw Secur* 8(11):15–23
8. Shapoorifard H, Shamsinejad P (2017) Intrusion detection using a novel hybrid method incorporating an improved KNN. *Int J Comput Appl* 173(1):5–9
9. Esmaily J, Moradinezhad R, Ghasemi J (2015) Intrusion detection system based on multi-layer perceptron neural networks and decision tree. In: 2015 7th conference on information and knowledge technology (IKT). IEEE
10. Muda Z et al (2011) Intrusion detection based on K-Means clustering and Naïve Bayes classification. In: 2011 7th international conference on information technology in Asia. IEEE
11. Nusinovici S et al (2020) Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol* 122:56–69 (2020)
12. Kumar M, Hanumanthappa M, Suresh Kumar TV (2012) Intrusion detection system using decision tree algorithm. In: 2012 IEEE 14th international conference on communication technology, IEEE

13. Li H, Yu N (2012) Intrusion detection technology research based on apriori algorithm. *Phys Procedia* 24:1615–1620
14. Sethi K et al (2020) A context-aware robust intrusion detection system: a reinforcement learning-based approach. *Int J Inf Secur* 19(6):657–678
15. Aslahi-Shahri BM et al (2016) A hybrid method consisting of GA and SVM for intrusion detection system. *Neural Comput Appl* 27(6):1669–1676
16. Vasudevan A, Harshini E, Selvakumar S (2011) SSENNet-2011: a network intrusion detection system dataset and its comparison with KDD CUP 99 dataset. In: 2011 second Asian Himalayas international conference on internet (AH-ICI). IEEE
17. Khraisat A, Alazab A (2021) A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. *Cybersecurity* 4(1):1–27
18. Einy S, Oz C, Dorostkar Navaei Y (2021) The anomaly-and signature-based IDS for network security using hybrid inference systems. *Math Prob Eng* 2021
19. Kumar V, Sangwan OP (2012) Signature based intrusion detection system using SNORT. *Int J Comput Appl Inf Technol* 1(3):35–41
20. Al-Jarrah OY et al (2014) Machine-learning-based feature selection techniques for large-scale network intrusion detection. In: 2014 IEEE 34th international conference on distributed computing systems workshops (ICDCSW). IEEE
21. Maseeh ZK et al (2021) Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset. *IEEE Access* 9:22351–22370
22. Zhao S et al (2017) A dimension reduction model and classifier for anomaly-based intrusion detection in internet of things. In: 2017 IEEE 15th international conference on dependable, autonomic and secure computing, 15th international conference on pervasive intelligence and computing, 3rd international conference on Big Data intelligence and computing and cyber science and technology congress (DASC/PiCom/DataCom /CyberSciTech), IEEE, 2017
23. Abdulhammed R et al (2018) Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic. *IEEE Sens Lett* 3(1):1–4
24. Farzaneh B, Ali Montazeri M, Jamali S (2019) An anomaly-based IDS for detecting attacks in RPL-based internet of things. In: 2019 5th international conference on web research (ICWR). IEEE
25. Kavitha S, Uma Maheswari N (2021) Network anomaly detection for NSL-KDD dataset using deep learning. *Inf Technol Ind* 9(2):821–827
26. Bedi P, Gupta N, Jindal V (2021) I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems. *Appl Intell* 51(2):1133–1151
27. Almseidin M et al (2017) Evaluation of machine learning algorithms for intrusion detection system. In: 2017 IEEE 15th International symposium on intelligent systems and informatics (SISY). IEEE
28. Almansob SM, Lomte SS (2017) Addressing challenges for intrusion detection system using naive Bayes and PCA algorithm. In: 2017 2nd International conference for convergence in technology (I2CT). IEEE
29. Aksu D, Ali Aydin M (2018) Detecting port scan attempts with comparative analysis of deep learning and support vector machine algorithms. In: 2018 International congress on Big Data, deep learning and fighting cyber terrorism (IBIGDELFT). IEEE
30. Boero L, Marchese M, Zappatore S (2017) Support vector machine meets software defined networking in ids domain. In: 2017 29th International teletraffic congress (ITC 29), vol 3. IEEE
31. Ahmim A et al (2019) A novel hierarchical intrusion detection system based on decision tree and rules-based models. In: 2019 15th International conference on distributed computing in sensor systems (DCOSS). IEEE
32. Thippa RG et al (2020) Analysis of dimensionality reduction techniques on big data. *IEEE Access* 8:54776–54788
33. Yang A et al (2019) Design of intrusion detection system for internet of things based on improved BP neural network. *IEEE Access* 7:106043–106052

34. Xu C et al (2018) An intrusion detection system using a deep neural network with gated recurrent units. *IEEE Access* 6:48697–48707
35. Alsirhani A, Sampalli S, Bodorik P (2019) DDoS detection system: using a set of classification algorithms controlled by fuzzy logic system in apache spark. *IEEE Trans Netw Serv Manage* 16(3):936–949
36. Novaes MP et al (2020) Long short-term memory and fuzzy logic for anomaly detection and mitigation in software-defined network environment. *IEEE Access* 8:83765–83781
37. Malhotra S, Bali V, Paliwal KK (2017) Genetic programming and K-nearest neighbour classifier based intrusion detection model. In: 2017 7th International conference on cloud computing, data science & engineering-confluence. IEEE, 2017
38. Salih AA, Abdulrazaq MB (2019) Combining best features selection using three classifiers in intrusion detection system. In: 2019 International conference on advanced science and engineering (ICOASE). IEEE
39. Wang D, Zhang F, Xi L (2011) Evolving boundary detector for anomaly detection. *Expert Syst Appl* 38(3):2412–2420
40. Poggiolini M, Engelbrecht A (2013) Application of the feature-detection rule to the negative selection algorithm. *Expert Syst Appl* 40(8):3001–3014
41. Kadar M, Tverdyshev S, Fohler G (2020) Towards host intrusion detection for embedded industrial systems. In: 2020 50th Annual IEEE-IFIP international conference on dependable systems and networks-supplemental volume (DSN-S). IEEE
42. Ahmad Z et al (2021) Network intrusion detection system: a systematic study of machine learning and deep learning approaches. *Trans Emerg Telecommun Technol* 32(1):e4150
43. Aldribi A et al (2020) Hypervisor-based cloud intrusion detection through online multivariate statistical change tracking. *Comput Secur* 88:101646
44. Meryem A, Ouahidi BEL (2020) Hybrid intrusion detection system using machine learning. *Netw Secur* 2020(5):8–19
45. Bertoli GDC et al (2021) An end-to-end framework for machine learning-based network intrusion detection system. *IEEE Access* 9:106790–106805
46. Alqahtani H et al (2020) Cyber intrusion detection using machine learning classification techniques. In: International conference on computing science, communication and security. Springer, Singapore
47. Baraneetharan E (2020) Role of machine learning algorithms intrusion detection in WSNs: a survey. *J Inf Technol* 2(03):161–173
48. Shamshirband S et al (2020) Computational intelligence intrusion detection techniques in mobile cloud computing environments: review, taxonomy, and open research issues. *J Inf Secur Appl* 55:102582
49. Satheesh N et al (2020) Flow-based anomaly intrusion detection using machine learning model with software defined networking for OpenFlow network. *Microprocess Microsyst* 79:103285
50. Kumar P, Gupta GP, Tripathi R (2021) A distributed ensemble design based intrusion detection system using fog computing to protect the internet of things networks. *J Ambient Intell Humaniz Comput* 12(10):9555–9572
51. Hindy H et al (2020) Utilising deep learning techniques for effective zero-day attack detection. *Electronics* 9(10):1684
52. Mozaffari FS, Karimipour H, Parizi RM (2020) Learning based anomaly detection in critical cyber-physical systems. *Security of cyber-physical systems*. Springer, Cham, pp 107–130
53. Gopi R et al (2021) Enhanced method of ANN based model for detection of DDoS attacks on multimedia internet of things. *Multim Tools Appl*, 1–19
54. Singh N, Virmani D, Gao X-Z (2020) A fuzzy logic-based method to avert intrusions in wireless sensor networks using WSN-DS dataset. *Int J Comput Intell Appl* 19(03):2050018
55. Sellappan D, Srinivasan R (2020) Association rule-mining-based intrusion detection system with entropy-based feature selection: intrusion detection system. In: Handbook of research on intelligent data processing and information security systems. IGI Global, pp 1–24

Human Emotion Prediction Analysis on Post-COVID-19 Crisis in Digital Media Using Deep Learning



Nikita Agarwal and Ritam Dutta

Abstract The COVID-19 pandemic has produced a significant impact on society. Apart from its deadliest attack on human health and economy, it has also been affecting the mental stability of human being at a larger scale. Though vaccination has been partially successful to prevent further virus outreach, it is leaving behind typical health-related complications even after surviving from the disease. This research work mainly focuses on human emotion prediction analysis in post-COVID-19 period. In this work, a considerable amount of data collection has been performed from various digital sources, viz. Facebook, e-newspapers, and digital news houses. Three distinct classes of emotion, i.e., analytical, depressed, and angry, have been considered. Finally, the predictive analysis is performed using four deep learning models, viz. CNN, RNN, LSTM, and Bi-LSTM, based on digital media responses. Maximum accuracy of 97% is obtained from LSTM model. It has been observed that the post-COVID-19 crisis has mostly depressed the human being.

Keywords Convolutional neural networks (CNNs) · Recurrent neural network (RNN) · Bidirectional long short-term memory (Bi-LSTM) · Long short-term memory (LSTM) · Predictive analysis · Deep learning · Post-COVID-19

1 Introduction

The novel coronavirus (nCoV-2) first originated in Wuhan fish market, China, in December 2019. Since then, this contagious disease has spread worldwide and snatched nearly 50 lacs of human lives from almost 221 countries till now [1]. This nCoV-2 coronavirus has got its pattern similarity of almost 96% with previous pandemic severe acute respiratory syndrome (Bat-SARS), which has been spread

N. Agarwal · R. Dutta (✉)

Center of Intelligent Systems and Robotics, ITER, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha 751030, India

e-mail: ritamdutta1986@gmail.com

N. Agarwal

e-mail: nikitaagarwal@soa.ac.in

from bats [2]. At initial phases, this virus has been transmitted from human to human at short distances. It is believed that the coronavirus will not have large mortality rate, rather may be similar to any flu virus. But as time progresses, the infection has got started to spread locally, became epidemic and then pandemic. Initially, WHO has categorically informed all country officials about three things, i.e., social distancing, wearing mask, and sanitizing hands since no vaccine has been officially invented and deployed to the victims. But with invention of COVID-19 vaccines, the spread of infection started facing resilience at developed countries. Recently, many variants of this virus have been evolved, that are more threatening than its older variants [3]. Even there are many instances, where people are getting post-COVID-19 symptoms, viz. breathlessness, fatigue, hearing problems, respiratory and heart problems, joint pains, and many others [4, 5], even after surviving from the disease. In this work, to predict the human emotion on post-COVID-19 crisis, many deep learning models have been studied based on several related literature surveys [6–9]. Due to the forced lockdown across the globe, people started getting more engaged toward social media as moving outside has been often restricted ordered by government officials. People are engaged in sharing thoughts rationally/emotionally on social media at larger scale [10]. In our work, substantial amount of such data has been collected to perform human sentiment analysis.

In the following sections, the methodology of the proposed work has been discussed, followed by experimental results and future work.


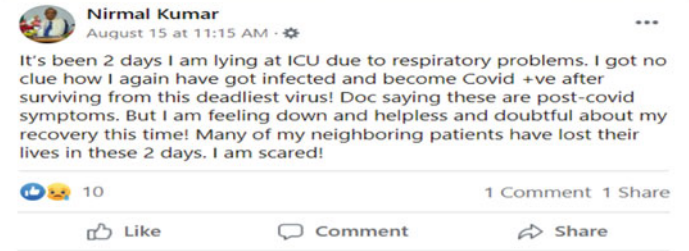
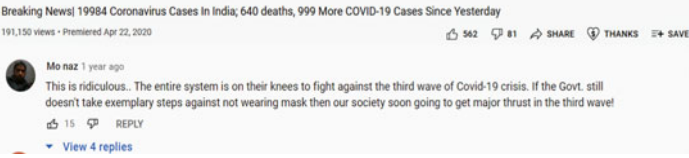
2 Methodology

In this section, we have discussed the flow of our work. Initially, the details regarding our dataset have been mentioned, followed by the data preprocessing steps, word embedding technique, and the deep learning models used.

2.1 Data Extraction

Data is the basic need of any deep learning model. The model trains itself using these data and makes predictions on unseen data. The amount of data available has a substantial impact on the performance of our model. For a deep learning model, we need substantial amount of data to train the model. Data collection has been performed from various digital sources, viz. Facebook, e-newspapers, and digital news houses in our work. We have collected 4986 samples that are related to the present COVID-19 pandemic situation. We have classified the data into three classes, namely logical, depressed, and angry. An example of the collected dataset used in our work is shown in Table 1.

Table 1 Sample dataset collected for predictive analysis

Comments	Category
	Analytical
	Depressed
	Angry

2.2 Data Preprocessing

Data preprocessing is a step where data are cleaned/preprocessed in a consistent format. Cleaning data is a mandatory step as it will remove unnecessary data and settle down the inconsistencies present in the data. It involves many steps, as explained below:

1. Converting the entire data to lowercase is a mandatory step.
2. Removal of duplicate values, white spaces, punctuations, stop-words.
3. Add contractions.
4. Lemmatization.
5. Handle missing values.

2.3 Word Embeddings

Word embedding is a technique to convert text data into its numerical vector representation to help our deep learning model analyze the text data. We have used the 200-dimensional global vectors (Glove) in our work. Liang et al. [11] has introduced Refining Affective Embedding technique from Context (RAEC) to analyze their data set. Pennington et al. [12] have introduced the Glove model. It is widely used in NLP tasks.

2.4 Deep Learning Models

In this part, we discuss the deep learning models used in our work. We have compared the performances of the following deep learning models.

2.4.1 Convolutional Neural Networks (CNN) Model

CNN is a deep learning model widely used in image and video recognition/classification tasks. Nowadays, it is also used with text data. We have used three convolution layers for feature extraction. For reducing the dimensionality of our data, we have used the max-pooling layer. Finally, the fully connected layer is added to make predictions using the softmax activation function. The parameters used in our CNN model are described in Table 2.

Table 2 Parameters used in the CNN model

Parameters	Values
Convolution layers	3
Max-pooling layers	3
Fully connected layers	2
Filters used in each convolution layer	64, 32, 32
Kernel/filter size in each convolution layer	3, 3, 3
Pool size in each max-pooling layer	3, 3, 2
Strides	2
Number of nodes in fully connected layers	100, 3
Activation function in convolution layers	ReLU
Activation function in output layer	SoftMax
Epochs	100
Learning rate	0.0001

Table 3 Parameters used in the RNN, LSTM, and Bi-LSTM models

Parameters	Values		
	RNN	LSTM	Bi-LSTM
Layers used	2	2	2
Fully connected layers used	1	1	1
Number of nodes in each layer	128, 64	64, 64	64, 64
Number of nodes in Fully Connected layers	3	3	3
Activation function	SoftMax	SoftMax	SoftMax
Epochs	100	100	100
Learning rate	0.0001	0.0001	0.0001

2.4.2 Recurrent Neural Networks (RNN) Model

RNNs are the neural networks that make future predictions based on previous knowledge. Unlike CNN's being a feed-forward network, RNNs analyze the data recurrently. It is widely used in NLP tasks. It has a memory that stores the past information that helps make predictions efficiently compared to other models. It can analyze the past data for a shorter period. Also, it suffers from vanishing and exploding gradient problems. The parameters used in our RNN model are described in Table 3.

2.4.3 Long Short-Term Memory (LSTM) Model

It is a type of RNN model that can remember past information for a more extended period than the RNN model. It removes the vanishing and exploding gradients problem of the RNN model to a large extent. The LSTM model has three gates, viz. forget gate, input gate, and output gate. The forget gate determines the information that must be forgotten/retained from our data. The input gate is used to determine the relevant information in the current step and can be used in the future time step. The output gate is used to make predictions. The parameters used in our LSTM model are described in Table 3.

2.4.4 Bidirectional Long Short-Term Memory (Bi-LSTM) Model

Bi-LSTM model analyzes the data from start to end and vice-versa. Since it analyzes the data forward and backward, it will take a little more time than the LSTM model but gives good results. It can store past information and use it to make future predictions, providing good results. The parameters used in our Bi-LSTM models are described in Table 3.

3 Evaluation Metrics

In this section, the evaluation metrics used to evaluate our models have been discussed. We have used four metrics as described below:

1. Accuracy: It represents the total number of predictions made correctly.
2. Precision: It represents the number of correctly made predictions of a class out of all predicted values for that class.
3. Recall: It represents the number of correctly made predictions of a class out of the actual values for that class.
4. F1-score: The harmonic mean of precision and recall is represented by F1-score.

4 Experimental Result Analysis

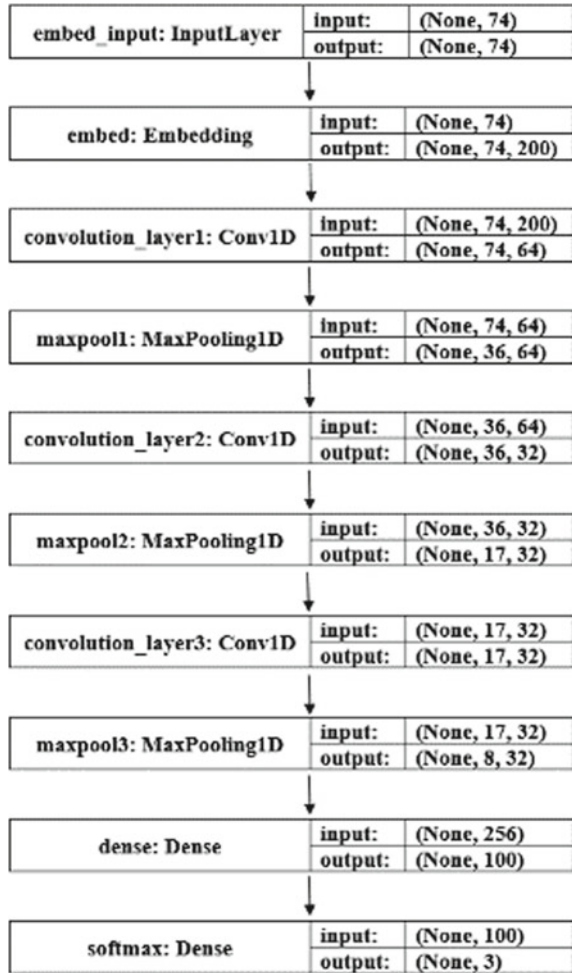
The experimental results are analyzed in this section. The parameters used for the deep learning models are shown in Tables 2 and 3. The text data is initially preprocessed to remove unwanted information and convert it into a consistent format. The Glove 200-dimensional vectors have been used to convert the text into numerical vectors so that our deep learning models can manipulate it. The deep learning models are then trained using the Adam optimizer that optimizes the models in each epoch. The architecture of the models used in our work is shown in Figs. 1, 2, 3, and 4. The parameters of the deep learning models are set using the Grid Search technique.

Fivefold cross-validation for the predictive analysis has been incorporated in this work. The mean of the k-fold results is generated to analyze the result. The performances of the models are assessed based on the evaluation metrics. The loss, accuracy, precision, and recall plots obtained for 100 epochs during the training of the deep learning models have been shown in Figs. 5, 6, 7, and 8. It has been observed that loss decrease with the increase in the number of epochs for these models. The accuracy, precision, and recall values are increasing with the number of epochs for these models. Table 4 shows the results obtained using the deep learning models. It has been observed that the LSTM model has provided a better performance compared to the other models.

5 Conclusion and Future Scope

The present pandemic had a severe impact on human lives. It has affected people, not only financially but also physically and mentally. This research work shows the behavior of the common people as they react to this global crisis. The predictive analysis has been performed using different deep learning models, viz. CNN, RNN, LSTM, and Bi-LSTM. It has been observed that the performance of the LSTM model has been better compared to other models. Moreover, a larger number of datasets

Fig. 1 Proposed CNN model



with better optimization algorithms and word embedding techniques can be used as future scope to achieve better human emotion prediction analysis.

Fig. 2 Proposed RNN model

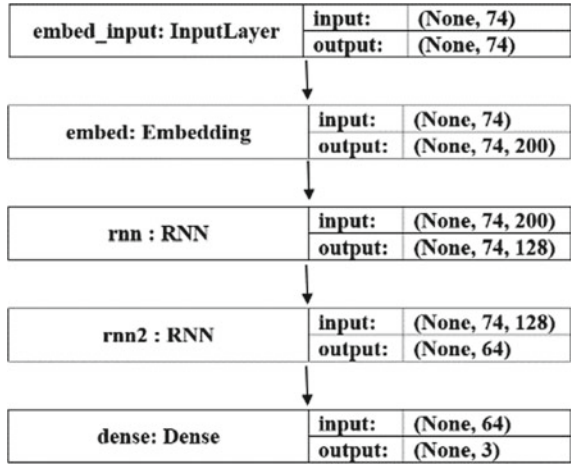
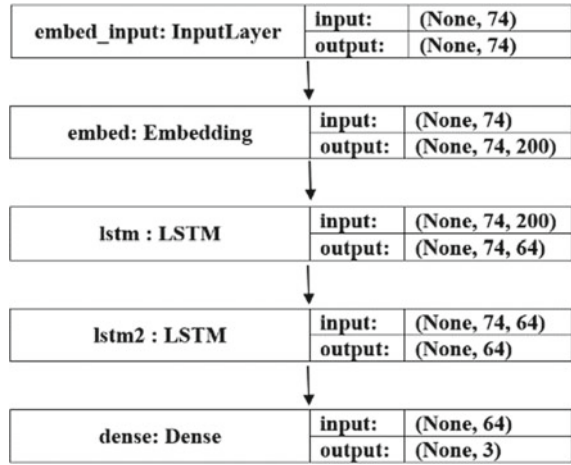


Fig. 3 Proposed LSTM model



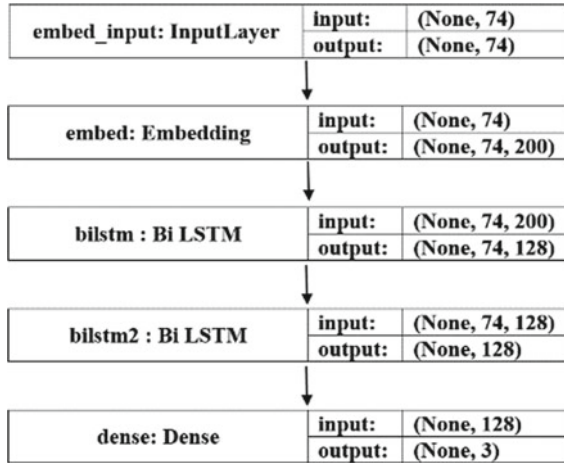


Fig. 4 Proposed Bi-LSTM model

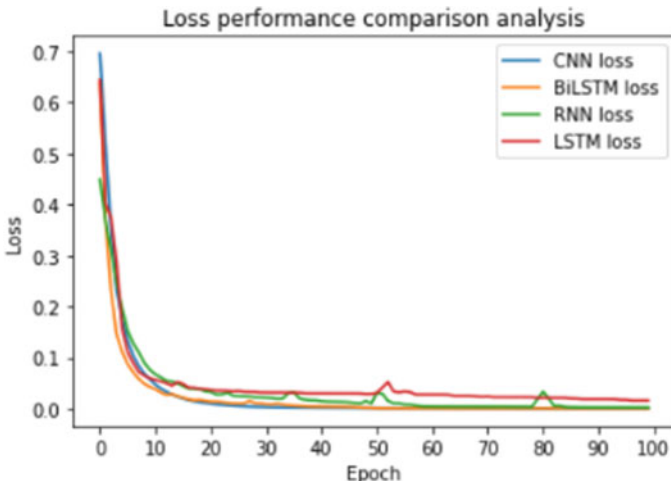


Fig. 5 Loss plot

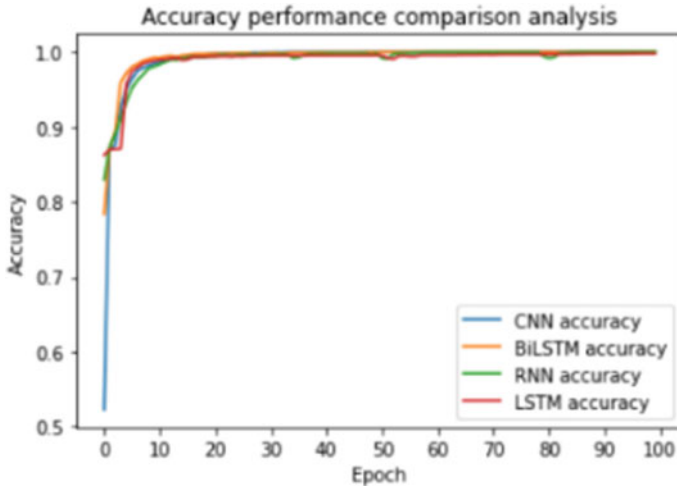


Fig. 6 Accuracy plot

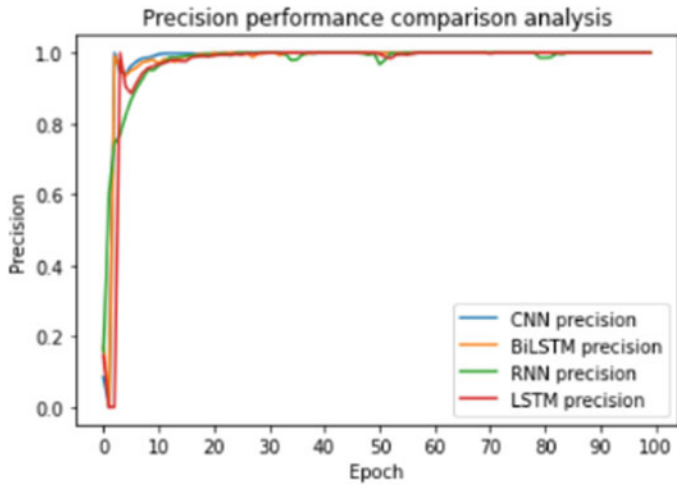


Fig. 7 Precision plot

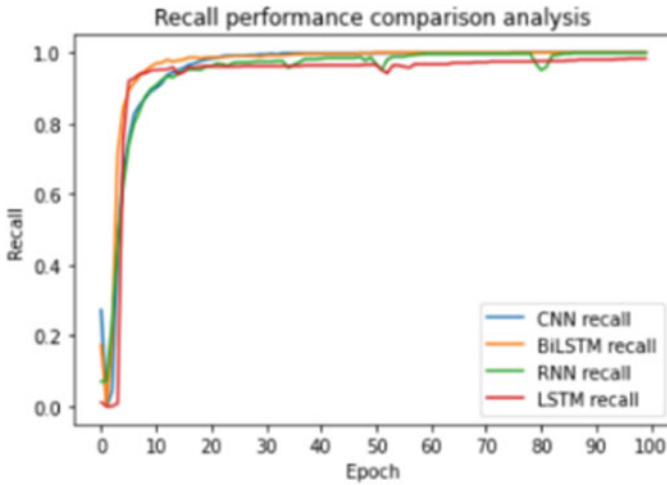


Fig. 8 Recall plot

Table 4 Comparison result on testing data

Model	Accuracy	Precision	Recall	F1-score
CNN	0.96412556	0.96323529	0.78915663	0.867549669
RNN	0.95605382	0.86335404	0.83734939	0.850152905
Bi-LSTM	0.96950672	0.897590361	0.897590361	0.897590361
LSTM	0.97309417	0.919753086	0.897590361	0.908536585

References

1. W.H.O reports on COVID-19 as on 15 Sept 2021. <https://covid19.who.int/>
2. Ross R (2021) Estimates of SARS death rates revised upward. Center for Infectious disease research and policy, 07 May 2003. Available at <https://www.cidrap.umn.edu/news-perspective/2003/05/estimates-sars-death-rates-revised-upward>. Accessed 15 Sept 2021
3. Van der Made CI, Simons A, Hoeijmakers JS. Presence of genetic variants among young men with severe COVID-19. JAMA 324(7):663–673. <https://doi.org/10.1001/jama.2020.13719>
4. Tolba M, Omirah MA, Hussein A., Saeed H (2020) Assessment and Characterization of post-COVID-19 manifestations. Int J Clin Pract 01–16. <https://doi.org/10.1111/ijcp.13746>
5. Raveendran AV, Jayadevan R, Sashidharan S (2021) Long COVID: an overview. Diabetes Metab Syndr 15(3):869–875. <https://doi.org/10.1016/j.dsx.2021.04.007>
6. Ghosh P, Dutta R (2021) Statistical machine learning forecasting simulation for discipline prediction and cost estimation of COVID-19 pandemic. In: Data science for COVID-19, pp 147–173. <https://doi.org/10.1016/B978-0-12-824536-1.00019-8>
7. Raza GM, Butt ZS, Latif S, Wahid A (2021) Sentiment analysis on COVID tweets: an experimental analysis on the impact of count vectorizer and TF-IDF on sentiment predictions using deep learning models. In: International conference on digital futures and transformative technologies (ICoDT2), pp 1–6. <https://doi.org/10.1109/ICoDT252288.2021.9441508>.
8. Dutta R, Chowdhury S, Singh KK (2021) IoT-based healthcare delivery services to promote transparency and patient satisfaction in a corporate hospital. In: Machine learning and

- internet of medical things in healthcare, pp 135–151. (2021). <https://doi.org/10.1016/B978-0-12-821229-5.00001-X>
9. El-Affendi MA, Alrajhi K, Hussain A (2021) A novel deep learning-based multilevel parallel attention neural (MPAN) model for multidomain arabic sentiment analysis. *IEEE Access* 9:7508–7518. <https://doi.org/10.1109/ACCESS.2021.3049626>
 10. Panchal N, Kamal R, Cox C, Garfield R (2021) The implications of COVID-19 for mental health and substance use. Kaiser Family Foundation
 11. Liang B, Yin R, Du J, Gui L, He Y, Xu R (2021) Embedding refinement framework for targeted aspect-based sentiment analysis. *IEEE Trans Affect Comput.* <https://doi.org/10.1109/TAFFC.2021.3071388>
 12. Pennington J, Socher R, Manning CD (2014) GloVe: global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp 01–12. <https://doi.org/10.3115/v1/D14-1162>

Role of Blockchain for Sustainability and Circular Economy



Anuj Kumar, Monika Arora, Kuldeep Bhalerao, and Meghna Chhabra

Abstract This paper aims to discuss the application of blockchain technology for achieving sustainability and moving toward the circular economy. The researchers have examined the role of blockchain technology in achieving the aims of sustainable production. While discussing the part of the blockchain for sustainability, the cases of IBM and Deloitte have been conferred to. Further, in the circular economy, the instances of Mitsui and Lablaco have been addressed. Overall, it can be concluded that the implementation of blockchain helps achieve sustainability by bringing transparency to the system and enhancing the market position. The emphasis on blockchain also helps take organizational practices toward a circular economy by generating new profit opportunities and focusing on green practices.

Keywords Blockchain · Circular economy · Sustainability · Management · Technology

1 Blockchain Technology: Introduction

The recent increase in the value of bitcoin, followed by its subsequent collapse, has re-ignited interest in the blockchain architecture that supports cryptocurrencies. Blockchain's applications, on the other hand, go far beyond virtual currency. The technology can potentially change the way people think about ownership, privacy, uncertainty, and collaboration in the digital world, disrupting industries and practices as diverse as financial markets, content distribution, supply chain management, humanitarian aid distribution, and even general election voting.

A. Kumar (✉) · M. Arora
Apeejay School of Management, Dwarka, Delhi, India
e-mail: anujsmooth@gmail.com

K. Bhalerao
Bharati Vidyapeeth's Institute of Management Studies and Research, Navi Mumbai, India

M. Chhabra
Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India

Blockchain technology uses the decentralized network where the transaction is stored and saved decentralized, available with all the entities. The use of blockchain is secure and immutable. It is also resilient as the data is stored with all the entities. If one system crashes, it can be retrieved from other devices or methods [1]. The other parameters, permissible and audibility, can also be implemented using blockchain technology. Blockchain helps in increasing trust in entities and works in the identification of risk. This risk can be mitigated at the first level without delay in other processes [2]. The blockchain uses control for decentralization, history of itself, performance in operations, the confidentiality of data, and distributed ledger technology. Distributed ledger technology aids transparency and faster transactions which help operate the supply chain processes. Blockchain has discoursed a model where artificial intelligence and deep learning can be beneficial and easily incorporated in the formation of recommender systems. This study discusses the application of blockchain technology for achieving sustainability and moving toward the circular economy. The outputs of this study will give an essential edge to researchers and practitioners for comprehending the use of blockchain in any emerging technology, and the system has been discussed for the implementation of the same in businesses, too [3]

1.1 Sustainability

Sustainability is a human-led ecosystem that uses three pillars social, economic, and environmental. Sustainable practices take care of the environment, consider the exhaustive use of natural resources, prevention of pollution and biodiversity [4, 5]; of society, focusing on the standard of living, job, and equal opportunity to everyone; on economics, focusing on growth, profit, and cost-saving [1, 6]. Environmental sustainability, social sustainability, and economic sustainability have been discussed as the various sustainability types in the business [7, 8]

1.2 Role of Blockchain for Sustainability

The systematic literature review was conducted and will help slice and dice the documents considered in the research. Tseng and Shang [9] discussed the operational efficiency to be achieved using blockchain technology in the intermediate functionality of sustainability. This will mitigate the risk and avoid conflicts [1]. The digital solution will be discussed as a shift from traditional operations. Digital technology will help in operations activities and intermediaries between social, economic, and environmental activities. The trust and transparency will be led using blockchain expertise in any business application. Varriale et al. [10] studied the adoption of sustainability in the blockchain [8]. It has discussed the positive and negative impacts of sustainability, considering all the three pillars in considerations. It has also discussed the

advantage of using blockchain and identified the parameters involved in trust in the supply chain. Many technological factors impact the company's performance using blockchain technology to develop sustainable supply chain practices. This has been discussed as the sustainable social, economic, and environment and further addressed as the sustainable and unsustainable aspects for creating blockchain technology in supply chain practices [11, 12], and Li [13] discussed the novel blockchain-based and its performances in the protection of environmental, equity of social, and efficiency of governance. Sustainability demands more transparency in intermediate processes of supply chain governance [10]. The blockchain can be integrated with big data analytics and the Internet of Things to implement blockchain in supply chain sustainability. This traceability can be preceded in a multitier global sourcing system. The challenges and issues of cost and quality are discussed in the research paper [10].

Rejeb et al. [14] discussed the use of blockchain technology in supply chain practices. They concluded that blockchain would facilitate creating data transparency and information at all levels of the supply chain [15]. Giungato et al. [16] have considered the sustainability of bitcoins and explained that the transition of traditional monetary to cryptocurrency and the whole economic system would change and maintain the virtual financial system. From the social change perspective, blockchain technologies may be used in the future and can be helpful for the implementation of sustainability [16]. Lund et al. [17] stated that sustainability could be implemented using blockchain [18]. It is a recent scientific revolution, which is promising the industry to change. With the help of intelligent grids and supply chain management systems, sustainable technological innovation has been proposed. The summary of the literature review conducted various papers on the blockchain; sustainability is described in Table 1

2 Research Methodology

The research methodology used in this paper is based on secondary data resources to present the usage of blockchain technology and circular economy practices by organizations. The researchers have used the examples of different organizations to understand how those organizations are achieving sustainability through technological means.

Case-1: Blockchain Industry Application—IBM Blockchain

IBM has worked in various sectors using blockchain. It has started creating an ecosystem for expertise in collaboration. IBM has involved multiple industries and organizations for visibility and end-to-end automation. IBM worked for sustainability and benefits the environment by performance, operation agility, speed, minimization of cost & risk, and new monetization opportunities. IBM has worked on various blockchain industry applications. Some are automotive, financial and banking, government, health care, life sciences, insurance, media and entertainment,

Table 1 Literature review for blockchain adoption and sustainability

Topic covered	Description
Operational efficiency and sustainability	It has focused the study on achieving operational efficiency by using blockchain technology. It also uses sustainability in its intermediate functionality [1]
Digital technology and sustainability	It has been stated that using digital technology, the operational activities between social, economic, and environmental movements are smoothly carried out and implemented well [1]
Trust and sustainability	The studied the adoption of blockchain and sustainability and trust in blockchain [8]
Big data analytics and IoT	The implementation of supply chain sustainability using big data analytics and the internet of things [10]
Transparency and supply chain governance	Transparency in all the intermediaries in supply chain governance can be implemented using blockchain was discussed [10]
Supply chain practices and sustainability	It discussed the use of blockchain technology in supply chain practices [15]

retail & consumer goods, telecommunication, travel and transport, supply chain, oil & gas, and manufacturing. IBM is into providing the services for blockchain success. The few of the services that are discussed in detail are as follows:

First is the use of blockchain in water management in IBM, through which the technology would help in environment protection and sustainability. It has initiated a web-based dashboard where financiers, farmers, and regulators can real-time monitor groundwater usage. This is transparent to everyone. This will regulate the market price and inform all the suppliers. The use of blockchain helps to work efficiently and mitigates asymmetric information, which will significantly contribute to blockchain-related technologies, as shown in Fig. 1. Many technologies incorporate and use blockchain technology, such as the Internet of Things (IoT) and blue-chip technology [19].

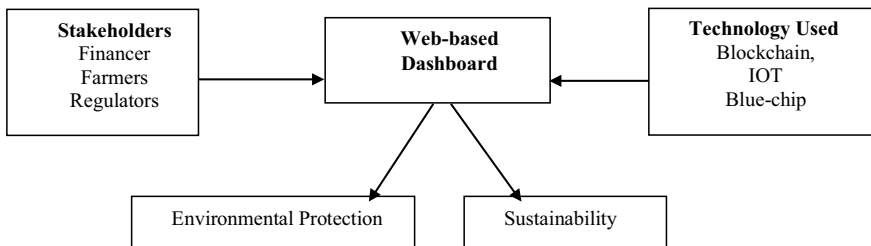


Fig. 1 IBM blockchain—water management system

Secondly, the services of IBM blockchain hyperledger fabric where the traceability has worked using the blockchain. This has helped the supply chain by utilizing the features of blockchain such as decentralization, transparency, durability, suitability, etc. The goal of IBM is to develop the solutions for the company based on company-ready solutions that will help the organizations maintain privacy, audit ability, measurability, performance, and confidentiality. It also has provided any cloud-based solutions that enable and expand the competitive edges. The system uses a public platform for the hyperledger cloth system for companies [20].

Thirdly, we discussed the use of blockchain in food ledgers. Walmart has used the same concept which IBM has operated as a food ledger. Walmart is tackling and has established the food safety collaboration center and uses the hyperledger IBM fabric system farm-to-table approach, where the promoters get the traceability of every moment. That has also discussed the challenges in implementing blockchain technology in the food supply chain. It also takes care of the temperature and other parameters while storing and traveling. This has improved the global food ecosystem, increased food safety, and reduced waste [21]. Technology adoption is the key for organizations to achieve sustainability goals [22].

Case-2: Deloitte Blockchain Services

Deloitte has developed the first blockchain lab in Dublin. The solution using blockchain was designed to validate the credentials of employees. It uses the Bitcoin ATM for its transaction. It has been accessible to all as a public. It has connected 22 companies in partnership and launched a platform for research and development for demonstration and experiments [23]. Deloitte has used the technology as sustainability, such as management of Stock in the fishing industry. It also enables a circular economy using plastics as recycling; it has reduced products that do not harm anybody. Deloitte has worked in food waste, where the reduction of waste has been performed upon.

Deloitte has developed the framework for clients to explore and develop tailored solutions to achieve end-to-end value. “Deloitte blockchain” has set the global multi-disciplinary model to assist organizations and industries in accomplishing the resolutions. They are focused on various businesses such as energy, health care, government, insurance, and entertainment, as shown in Fig. 2. It has created the overall technology strategy and helps the enterprise in blockchain development by innovation, prototyping, product development, etc. It guides the enterprises throughout the journey. The offering includes:

- “Blockchain strategy,” where the workshop, lab was created on an open platform where anyone can use it
- “Consortium strategy,” which manages the industries and enterprises
- “Platform design” includes the new business models such as vendor selection, etc.
- “Case design and implementation” consists of the test blockchain infrastructure, and one can test pilots in the market study.

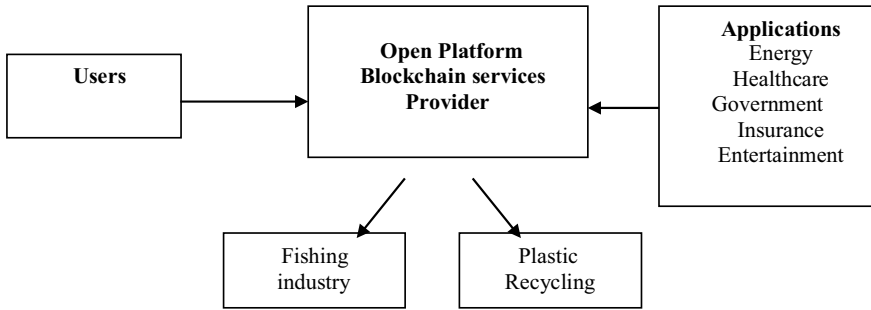


Fig. 2 Deloitte blockchain services—Open platform

2.1 Implications for Managers and Blockchain in Business

Enhancing Trust Blockchain enables real-time, which will be reducing the risk in business [24]. Managers can trust the different stakeholders connected to the business blockchain. The proof of transaction is available to all the parties and can be tracked easily. Blockchain has been increasingly used in the logistics sector. The use of blockchain has increased the trust in the system due to traceability, transparency, and cost reduction. The manual and paperwork have been replaced in paperless using blockchain implementation.

Enhancing Market Position Organizations using blockchain have an added advantage and improve their market position. It would also help the managers examine the supply chain's characteristics with excellence in products and services. The proper use of technology allows the organization to achieve sustainable goals and focuses on the efficiency and profitability of the business. However, it is also a challenge for managers to determine the best blockchain solution for their enterprise. The managers can also use the blockchain to develop human capital expertise. Disruptive technologies such as big data and cloud computing have been leveraged by the “Internet of Things” (IoT) to overcome its limitations since its conception, and blockchain could potentially improve the IoT [14].

3 Circular Economy: Introduction

A “circular economy” would turn goods at the end of their service life into resources for others, closing loops in industrial ecosystems and minimizing waste [25]. Future sustainable demands economic, social, environmental, institutional, organizational, and business ecosystems [26]. The move toward a sustainable future has a vibrant driver: The linear economy is touching to its physical limit. The circular economy has invited the attention of stakeholders like policy-makers, academicians, thinkers, and

corporate to think of the circular economy as a sustainable innovation for the well-being of the human species and the ecosystem at large. According to [27], adverse changes in climate and scarcity of resources are the concerns now and adopting a circular economy may enable sustaining position at various levels.

Pagoropoulos et al. [12] argued that circular economy is based on the principles of value creation in products and services, maximizing the life cycle of the product or service through regeneration [23]. In the transition to the circular economy, traditional businesses must rethink and redesign business definitions, business models, and business operations. In this transition, technologies play a crucial role in data collection, analysis, and integration.

However, in many cases, the technology needed for a circular economy is not affordable or expensive. This inadequacy becomes a significant barrier to the successful adoption of the circular economy.

3.1 Role of Blockchain Technology in the Circular Economy

Khan et al. [28] argued that blockchain technology plays a positive role in the circular economy [29]. Furthermore, green practices have a positive relationship with the environmental and economic performance of the organization. The study concluded that economic and ecological performance is positively influencing and boosting the performance of the organization. Kouhizadeh and Sarkis [30] concluded that blockchain technology might enable companies to exchange and trade their wastage and create value in business processes [31]. With the help of blockchain technologies, companies can interact on various waste management proposals without the involvement of third-party and enhance the profit margin of both. Blockchain-based smart contracts may also expedite the automatic exchange of waste material based on quality, volume, condition, and type of waste. In addition, e-sensors and online tracking systems are helping in tracking location, information regarding the price of waste, and making data available on blockchain ledger systems. Traceability of hazardous wastage is one of the complex tasks.

The study [32] was explored how organizations have developed the localization, agility, and digitalization (LAD) model of capability with the application of blockchain technology and circular economy principles. The study shows a positive correlation between blockchain-enabled circular economy systems (BCES) and localization, agility, and digitalization (L-A-D) capabilities. The industry in which the business is operating, and the firm's size influence the BCES and LAD. It found that artificial intelligence can enable the application of blockchain technology-based smart contracts and multi-sensor-fusion algorithms [33]. The various data fusion modes may be used for retrieving relevant parameters of plastic waste for precise separation. The application of blockchain-based technologies and artificial intelligence enables an organization to segregate plastic waste and share data regarding making available plastic waste. Various manufacturers can use this data to check the reliability of plastic waste and can be used for effective and efficient supply chain

planning and execution of purchase orders. Hence, plastic waste can be used by the manufacturer and contribute to the circular economy of plastic.

According to a study, Volvo cars are using blockchain technologies to mark out cobalt in the batteries of an electric vehicle [34]. The blockchain shares information like origin, size, and weight of cobalt, chain of a custodian of cobalt, and information ensuring that the behavior of supply chain participants is as per globally recognized guidelines. Blockchain technology can also be applied to other industries like garments, breweries, and manufacturing, enhancing transparency for their supply chain. The previous researchers also found that using blockchain technologies enables organizations to realize their objectives and achieve sustainable development [35]. Taylor et al. [1] discussed the opportunities and challenges of adopting blockchain technologies in offering more explicit information about wastes and products property rights, favoring legal and strategic goals by incentivizing sustainable management of waste and maintaining the confidentiality of individuals and institutions significant [36]. Upadhyay et al. [37] found that blockchain technology can contribute constructively to the circular economy by enabling a reduction in the transaction cost, performance improvement, ensuring the protection of human rights, maintaining patient confidentiality, and reducing carbon footprints [9]. The study also discussed challenges in blockchain technology adoption in terms of trust issues, criminal proceedings, and the threat of hacking; such problems are to be handled with a suitable legal framework and strategic decisions. The cost involved in the adoption of blockchain technologies is one of the concern areas. However, the benefits of blockchain technologies are overcoming the challenges very quickly [13].

Circular economy and blockchain technologies are giving positive results to companies, so more and more companies will adopt the circular economy [38]. Companies must take strategic benefit of advanced technologies to improve tracking and tracing of services and products after-sales service and should help in authentication, resale, and recovery of material. The present and future of the species on the earth would be significantly enhanced if the companies understood the importance of circular economy and the benefits of applying blockchain technology [39] (Table 2).

Case I: Chemical Manufacturing—Mitsui

Mitsui Chemicals and IBM Japan have decided to work jointly on applying blockchain technology to recycle plastic materials. The fundamental aim of using blockchain technology is to minimize the waste of various processes through reuse, repair, reshaping, and recycling material and components. It is significant to track the material lifecycle to ensure they are reused. Moreover, it is also essential for producers to know the accurate chemical composition used in production, which is complicated when using recycled material. Recycling various materials or waste is the basis of a prosperous circular economy.

Blockchain technology offers data about the lifecycle of material used in manufacturing and ensures its chemical composition. If data is audited appropriately, it gives confidence to regulatory agencies and environmentally conscious customer's confidence. Mitsui Chemicals and IBM have developed a technological platform

Table 2 Literature review on blockchain technology-enabled circular economy

Author	Parameter	Blockchain and circular economy
[4]	Green practices	The adopted green practices by the organization are showing significant and positive relation with environmental and economic performance
[30]	Waste management	The use of blockchain technologies enables organizations in waste management without the involvement of third-party and contributes to an increase in profit margin
[23]	Localization, agility and digitalization (LAD) model	The study found a significant and positive relation between blockchain-based circular economy and localization agility and digitalization (LAD) model
[35]	Sustainable development	The adoption of blockchain is enabling companies to meet the desired objectives and achieve sustainable development
[9]	Maintenance	The adoption of blockchain technology in a circular economy enables minimizing the cost of the various transaction, improving overall performance and human rights protection, maintaining the intellectual property, and reducing carbon footprints
[13]	Challenges and benefits	The cost involved in adopting blockchain technology is one of the significant issues, but the benefits of the blockchain are overcoming it easily in most cases
[38]	Strategic advantage	The blockchain technology-based circular economy is giving relevant and positive results to organizations. Organizations should adopt it strategically to make the world ecologically balanced

that tracks the life cycle from sourcing raw material to reuse and recycle of material. This technical platform envisages production processes for recycling raw material, physical properties of the material, quality-related information, and testing methods.

Mitsui has established a digital transformation division, aiming to speed digitization throughout the company with a primary focus on various business operations and chain of supply. This digital transformation has enhanced sustainable practices at Mitsui; this association with IBM is crucial for the improvement in divisional growth. Mitsui Chemicals has leveraged technical expertise and skills built by working with monomer and polymer and eco-friendly practices such as reusing and recycling materials. Blockchain technology has created a resource management platform that acts as a material tracing system and helps achieve a circular economy.

Case 2: Garment and Fashion Brand—Lablaco

Many entities are involved in manufacturing, stitching, and merchandising ready-made garments; the whole system becomes very complex. Therefore, the fashion industry has problems with transparency in their business commitments and transactions. Most brands have no information about who manufactured their product and where it comes from in the fashion industry. Since the collapse of the Rana Plaza factory in 2013 [17], the dark side of the inhuman working conditions of garment manufacturing companies has been revealed. The fashion industry has decided to release a transparency index. According to a study by McKinsey & Company, brands need to develop trust with their customers [40]. The data-driven business world is helping various brands to reconnect with places and people of garment manufacturing.

Laplace is one of the renowned brands in the garment industry. It has shifted to the circular economy by the application of blockchain technology. These digital technology applications enable organizations to track and record step-by-step information of the journey of material or item from farmer to consumer. This information allows the customers to make their purchasing decisions easier. This presented information creates transparency in business transactions and is accessible to all business stakeholders and develops confidence among them. With the help of a quick response (QR) code or near-field communication (NFC) chip, buyers may use the mobile scanner to know the information about manufacturers and entities involved in the supply chain and get information about how much water has been used to manufacture and how much carbon dioxide is produced during the manufacturing of the product. Blockchain applications are becoming significant in shopping which suits the ecological environment. This secure and open-source data is putting stress on manufacturers to take responsibility for their business conduct.

3.2 Implications of Circular Economy and Blockchain on Business

New Profit Opportunities Blockchain and circular economy enable organizations to lower the input cost and maximize the profit margin [31]. Due to circular economy principles, business opportunities can be created by creating a new market, minimizing waste, reducing energy usage, and continuity in the supply chain.

Green Human Resource Management' Learning and Development Recently, one of the significant responsibilities of learning and development in the organizations is to enhance the skills, knowledge, and attitude of the employee for sustainable development of an individual and organization [41]. Learning and development managers strive to create awareness about the current environmental issues, educate employees to reduce waste, and energy-saving, teaching operational employees and decision-makers about the value chain that reduces waste. The organizations' use of

circular economy principles increases awareness among employees through green learning and development initiatives.

Safeguarding of Supply The circular economy model promotes organizations to use recycled material and enables organizations to be less dependent on changing prices of the raw materials. This could also safeguard the material from probable damages due to environmental changes worldwide. In the end, the circular economy makes businesses more aware of fluctuations in prices and unexpected damages due to climate changes [37].

4 Limitations of Blockchain and Circular Economy

Economical Our economic system gives more privileges to the financial market instead of social and environmental issues [28]. There is a lack of IT- or ICT-qualified professionals [42]. The demand for recycled products is still small. Most organizations are still working on the linear economy, so developing the circular economy model in such a working environment becomes challenging [6].

Institutional Implementing and developing the circular economy is facing several barriers. Still, our economic and business systems are not prepared to adopt blockchain or a circular economy [43]. Majorly companies have short goals of value creation. At the same time, the circular economy model is a strategic value creation model.

5 Conclusion

Based on the above discussion, it can be concluded that using blockchain technology will make organizations sustainable by reducing their overall cost and maximizing profits. The transparency will be more in the system, and it also creates new profit opportunities. It will be hard to survive for the organizations without thinking about sustainable practices. For all the nations and organizations looking to achieve sustainability goals, technology adoption and usage can be a game-changer in this regard. The circular economy also focuses on using shared resources, and blockchain technology can make it possible for organizations to integrate themselves toward the goal of the circular economy.

References

1. Taylor P, Steenmans K, Steenmans I (2020) Blockchain technology for sustainable waste management. *Front Polit Sci* 15
2. Arora A, Arora M (2018) Digital-information tracking framework using blockchain. *J Supply Chain Manage Syst* 7(2)
3. Arora M, Chopra AB, Dixit VS (2020) An approach to secure collaborative recommender system using artificial intelligence, deep learning, and blockchain. In: *Intelligent communication, control and devices*, pp 483–495
4. Kumar A, Aggarwal R (2018) Sustainable development through social media tools. *J Manage (JOM)* 5(5)
5. Kumar A, Aggarwal R (2018). Sustainable development through social media tools. *J Manage* 47–51
6. Yaja M, Kumar A (2021) An empirical study of marketing of SMEs in the tourism sector. *Small Enterprise Res*
7. Kouhizadeh M, Sarkis J, Zhu Q (2019) At the nexus of blockchain technology, the circular economy, and product deletion. *Appl Sci* 9(8):1712
8. van der Heijden RECM, Coenen JWE, van Riel ACR (2017) Transitioning from a linear economy towards a circular economy: the case of the apparel industry
9. Tseng CT, Shang SS (2021) Exploring the sustainability of the intermediary role in blockchain. *Sustainability* 13(4):1936
10. Varriale V, Cammarano A, Michelino F, Caputo M (2020) The unknown potential of blockchain for sustainable supply chains. *Sustainability* 12(22):9400
11. Bakshi G, Kumar A, Puranik AN (2022) Adoption of robotics technology in healthcare sector. In: Dhar S, Mukhopadhyay SC, Sur SN, Liu CM (eds) *Advances in communication, devices and networking. Lecture notes in electrical engineering*, vol 776. Springer, Singapore. https://doi.org/10.1007/978-981-16-2911-2_42
12. Pagoropoulos A, Pigosso DC, McAloone TC (2017). The emergent role of digital technologies in the circular economy: a review. In: *Procedia CIRP*, pp 19–24
13. Park A, Li H (2021) The effect of blockchain technology on supply chain sustainability performances. *Sustainability* 1726
14. Rejeb A, Keogh JG, Simske SJ, Stafford T, Treiblmaier H (2021). Potentials of blockchain technologies for supply chain collaboration: a conceptual framework. *Int J Logist Manage*
15. Ramakrishna S, Ngowi A, Jager HD, Awuzie BO (2020) Emerging industrial revolution: symbiosis of industry 4.0 and circular economy: the role of universities. *Sci Technol Soc* 25(3):505–525
16. Giungato P, Rana R, Tarabella A, Tricase C (2017) Current trends in sustainability of bitcoins and related blockchain technology. *Sustainability* 9(12):2214
17. Lund EH, Jaccheri LLJ, Cico O, Bai X (2019) Blockchain and sustainability: a systematic mapping study. In 2019 IEEE/ACM 2nd international workshop on emerging trends in software engineering for blockchain (WETSEB). IEEE/ACM, pp 16–23
18. Kumar A, Ayedee N (2021) Technology adoption: a solution for SMEs to overcome problems during COVID-19. *Acad Market Stud J* 1–21
19. Chohan UW (2019) Blockchain and environmental sustainability: case of IBM's blockchain water management. *Notes on the 21st century (CBRI)*
20. Bhuvana R, Aithal PS (2020) Blockchain-based service: a case study on IBM blockchain services & hyperledger fabric. *Int J Case Stud Bus IT Educ (IICSBE)* 4(1):94–102
21. Kamath R (2018) Food traceability on blockchain: Walmart's pork and mango pilots with IBM. *J Br Blockchain Assoc* 1(1):3712
22. Kumar A, Ayedee N (2019) Sustainable development in SMEs through social media channels. *Int J Manage Technol Eng* 1066–1075
23. O'neal S (2019) Big four and blockchain: are auditing giants adopting yet? Retrieved from <https://cointelegraph.com/news/big-four-and-blockchain-are-auditing-giantsadopting-yet>

24. Kamble SS, Gunasekaran A, Sharma R (2020) Modeling the blockchain-enabled traceability in the agriculture supply chain. *Int J Inf Manage* 52:101967
25. Reyna A, Martín C, Chen J, Soler E, Díaz M (2018) On blockchain and its integration with IoT. Challenges and opportunities. *Future Gener Comput Syst* 88:173–190
26. Manik JA, Yardley J, Dhaka B (2013) Building collapse in Bangladesh leaves scores dead. *NY Times* 24:A1
27. Ghisellini P, Cialani C, Ulgiati S (2016) A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. *J Clean Prod* 11–32
28. Khan RAS, Yu ZS, Godil DI, Amin S, Shujaat S (2021) The role of block chain technology in circular economy practices to improve organizational performance. *Int J Logist Res Appl* 1–18
29. Welfare A (2020) The circular economy and sustainability powered by blockchain. 13 Jan 2020. Available: <https://www.forbes.com/sites/forbestechcouncil/2020/01/13/the-circular-economy-and-sustainability-powered-by-blockchain/?sh=4a39e110b8cf>. Accessed 28 Nov 2021
30. Kouhizadeh M, Sarkis J (2018) Blockchain practices, potentials, and perspectives in greening supply chains. *Sustainability* 3652
31. Korten DC (1998) When corporations rule the world. *Eur Bus Rev*
32. Nandi S, Sarkis J, Hervani AA, Helms MM (2021) Redesigning supply chains using blockchain-enabled circular economy and COVID-19 experiences. *Sustain Prod Consum* 27:10–22
33. Chidepatil A, Bindra P, Kulkarni D, Qazi M, Kshirsagar M, Sankaran K (2020) From trash to cash: how blockchain and multi-sensor-driven artificial intelligence can transform circular economy of plastic waste? *Adm Sci* 23
34. Kemp L (2020) How blockchain can power the circular economy. Retrieved 29 July 2021 from <https://blogs.oracle.com/blockchain/post/how-blockchain-can-power-the-circular-economy>
35. Bhalerao K, Mahale P, Jyothi N, Yelikar B (2021) Technology: a pathway towards sustainability. *Acad Mark Stud J* 1–9
36. Stahel WR (2016) The circular economy. *Nat News* 531(7595):435
37. Upadhyay A, Mukhuty S, Kumar V, Kazancoglu Y (2021) Blockchain technology and the circular economy: Implications for sustainability and social responsibility. *J Clean Prod* 126–130
38. Venkatesh VG, Kang K, Wang B, Zhong RY, Zhang A (2020) System architecture for blockchain based transparency of supply chain social sustainability. *Robot Comput-Integr Manuf* 63:101896
39. Pujari P, Kumar A (2021) Circular economy and environmental sustainability—The new global reality. *Wesleyan J Res* 73–77
40. Diebner R, Silliman E, Ungerman K, Vancauwenberghe M (2020) Adapting customer experience in the time of coronavirus. McKinsey & Company, pp 1–7
41. Chams N, García-Blandón J (2019) On the importance of sustainable human resource management for the adoption of sustainable development goals. *Resour Conserv Recycl* 141:109–122
42. Kumar A, Syed AA, Pandey A (2021) Adoption of online resources to improve the marketing performance of SMES. *Asia Pac J Health Manage* 16(3):137–144
43. Kumar A, Ayedee N (2021) An interconnection between COVID-19 and climate change problem. *J Stat Manag Syst* 24(2):1–21

Comparative Analysis of Machine Learning Models to Predict Stock Market Price



Shreya Sakshi, Arjun Kar, and Chitrapriya Ningthoujam

Abstract Stock market forecasting is viewed as one of the most interesting is of study for many researchers. The crucial data that can be accessed is thought to have predictive correlations with future stock performance which could provide information to investors so that they may make better decisions when purchasing equities. The paper tries to present a comparative analysis of four machine learning models to predict stock market price. The methods that we have considered are: support vector machine (SVM), artificial neural network (ANN), and hybrid models like PCA + SVM and PCA + ANN to predict stock market state. We have experimented using Vanguard Total Stock Market ETF (VTI) dataset for last 10 years which shows that SVM-based predictive model performed well among all the models for predicting the stock market status.

Keywords Stock market prediction · Principal component analysis · Machine learning · ANN · SVM · Hybrid models

1 Introduction

The fundamental position of the financial sector, as well as the other factors such as the previous day's closing price, P/E ratio, and the average price index that makes stock market prediction a difficult task. A lot of effort has been put into developing diverse data-driven algorithms for prediction. Machine learning algorithms have provided many benefits in context to the data mining approaches. Various physical, physiological, rational, and illogical variables are all important factors for study in stock market. It has been a complex task to predict stock price for many reasons. As a

S. Sakshi (✉) · A. Kar · C. Ningthoujam
Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology,
Majitar, Gangtok, India
e-mail: shreyasingh7478@gmail.com

C. Ningthoujam
e-mail: Chitrapriya.n@smit.smu.edu.in

result, understanding of both individual and market shares, financially and economically are important in addition to number of individuals and investors connected to modern technology.

2 Background

Kimoto et al. [1] and Mizuno et al. [2] used ANN to predict the Tokyo Stock Exchange index. Mizuno et al. [2] predict buying and selling signals with total forecast price of 63%. Nguyen et al. [3] used neural feed-forward network and a simple recurrent neural network (RNN) by adjusting weight factors. They have shown that RNN with modified back distribution performs better. Zahir et al. [4] and Yutis et al. [5] also used ANN, considering the parameters as: general indicators (GI), net asset value (NAV), per capita earnings (P/E), earnings per share (EPS), and share volume. Yutis et al. [5] suggested the use of integrated parameters like flexible technical and basic analysis. They have considered 18 complete inputs and applied to several neural network, later they are compared. The authors [6] have shown that the parameters of hybridized outperforms technical parameters. Waqar et al. [7] used principal components analysis (PCA) to find improvements in the performance for prediction. Stock market prediction can be done using support vector machine was shown by Iacomini [8]. While Joseph et al. [9] show that combination of SVM and PCA will shows good result.

2.1 Prediction Method Analysis

The stock market is the need of the present world, as it has a great influence on the economy. In a stock market prediction mapping, it to result and predicting it accurately is a challenging task because of its nonlinear characteristic. This section discusses the two broad parts of stock market analysis: fundamental analysis and technical analysis. The paper focus on the technical analysis part.

(a) Fundamental Analysis It involves analyzing the company's future profitability on the basis of its current business environment and financial performance.

(b) Technical Analysis It includes reading the charts and using statistical figures to identify the trends in the stock market.

2.2 Machine Learning Methods

The stock market is a need for the current world, as it has a huge impact on the economy. Predicting it is very difficult as there is a nonlinear factor. Marking it for

effect and predicting the outcome with high accuracy is a challenge. The problem arises that the percentage of accuracy should be high as it is a matter of high risk. It is a challenge to hybrid methods and come up with a solution which is better even for a large dataset.

Many researchers have used different machine learning techniques [10] which could give a better result in predicting the stock market price. Machine learning strategies such as ANN, SVM, PCA + ANN, and PCA + SVM. By estimating prices and mobility, support vector machines (SVM), and artificial neural networks (ANN) are widely used. Every algorithm has a different way of learning patterns and predicting it. The artificial neural network (ANN) is a standard tool for generating financial market predictions. In terms of pricing, support vector machines (SVM) and artificial neural networks (ANN) are widely used.

Artificial neural network (ANN) is a common tool for generating financial market predictions that include technical analysis, but hybrid models show better results in stock market predictions.

3 Methodology

See Fig. 1.

3.1 Data Preprocessing

The experiment starts by performing the basic data preprocessing of Vanguard Total Stock Market ETF (VTI) dataset for last 10 years. The prediction is done four module-wise as depicted in Fig. 1. An important requirement we faced for our experiment

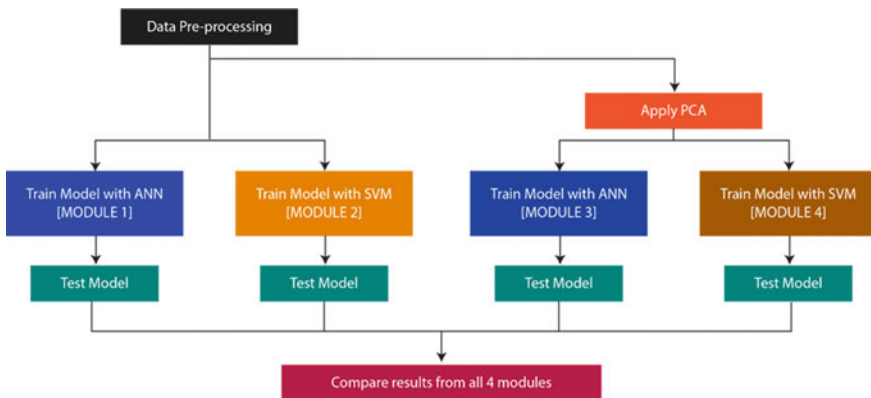


Fig. 1 Hierarchy of module

was as the dataset might contain some missing values, some outliers or some out of range values, etc., which needs to be processed before using the data to train any model. So, we considered it as an important requirement and did the following preprocessing of our dataset.

- **Data Cleaning:** We loaded the CSV file into data frame and found that no data value was missing. So, we did not face any problem related to NaN.
- **Target Variable Selection:** Basically, our project will aim to forecast the next day's return. So, we need to cross check the timing of the forecast so that we do not end up forecasting today or yesterday return.
- **Feature Extraction:** From the available features in the dataset, we have extracted few other features like initially we were having features like date, open, close, high, low, and volume and we processed those data to extract some other features like close-to-high, close-to-low, close-to-open, and close-to-close (i.e., return).
- **Scaling:** Well scaling helps in reducing the time to learn. This greatly helps in reducing time. We have scaled all features to unit variance and 0 mean to optimize training.
- **Analyzing Outliers:** We analyzed outlier of 'return' from dataset. The return lowest was 0.9606 and highest was 1.0482.

3.2 Splitting the Dataset

The dataset is split into a training and testing sets. The testing dataset helps us to get an unbiased estimate of actual performance. Often a testing database is randomly selected. But, since the stock market analysis is time series data, we will use the first 70% as a training database and save 30% as a testing set. Later, we also used cross verification using K-fold method having fold value ten to bring some improvement in the training of our ANN model (Module 1).

Choosing the right model and training is an important part of any machine learning project. We trained two models with artificial neural network (ANN) and support vector machine (SVM), respectively, in Module 1 and Module 2. In later modules, we have used PCA on our database and trained similar models with the newly generated database.

We had some datasets for testing the model that we used. We used around 30% dataset for testing the neural network model and evaluating it to get an actual unbiased view of how our dataset is performing. For our SVM model, also we have used the same proportion of data.

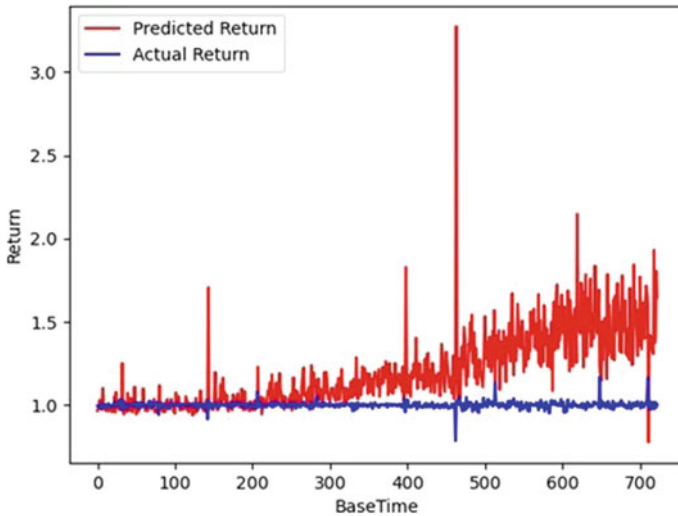


Fig. 2 Module 1 stock market prediction using ANN

3.3 Implementation Details

3.3.1 Module 1: Using ANN

Our first module is implementation of artificial neural network (ANN). For this, we have used scikit-learn library for machine learning in Python. We have used the regression ANN model having hidden layer size 200×200 , and additionally, we have used 'ReLU' activation function and 'Adam' solver/optimizer while training our model. We have divided our dataset into two sets, training set (having 70% of overall data) and testing set (having 30% of overall data). We have used time series split from scikit-learn to apply cross validation in our dataset while training for better result. The series value used is ten for our ANN model. We have used Python's pickle module to dump our ANN model after training and to load the same for testing. We are showing the results in the form of mean square error, accuracy measure, and a graphical representation of actual and predicted result is depicted in Fig. 2.

3.3.2 Module 2: Using SVM

In our second module, which is implementation of support vector machine (SVM). Since SVM can be applied for both regression and classification models. So, as our project is a regression problem, we have used support vector regressor (SVR) from scikit-learn library to implement our model shown in Fig. 3.

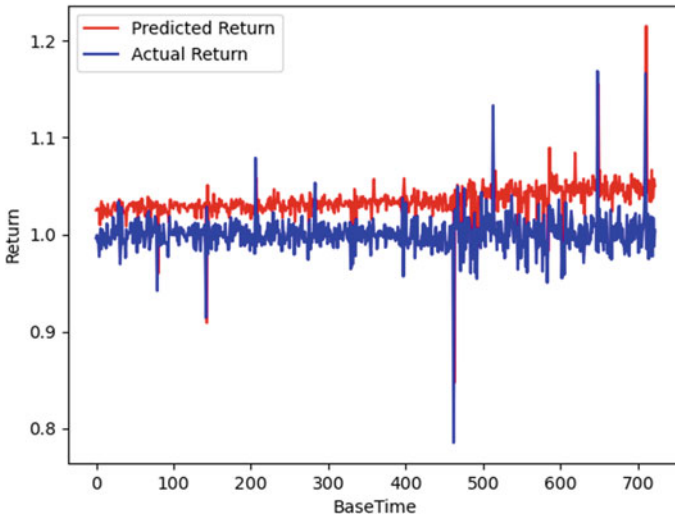


Fig. 3 Module 2 stock market prediction using SVM

3.3.3 Module 3 Using PCA and ANN

Module 3 tried to reduce the dimensionality of our dataset keeping all the details of all the features of the dataset using principal component analysis (PCA). Before implementing PCA, we again do some preprocessing with our previous dataset for better implementation of PCA. PCA consists of several steps like calculating mean of the feature set, calculating covariance, calculating eigenvalues and eigenvectors of the covariance matrix obtained in previous step, then sorting eigenvectors, and selecting the first two (because we wanted a two-dimensional dataset after reduction) features from the obtained dataset. We then trained our ANN model (from Module 1) with this newly obtained dataset expecting a better and improved result, then the previous approaches module. We also used MSE, accuracy, and plotted the graph to show the results as shown in Fig. 4.

3.3.4 Module 4 Using PCA and SVM

In our last module, we have just trained our SVM model (from Module 2) with our newly obtained two-dimensional dataset, which we got after applying PCA to our previous dataset (in Module 3). Figure 5 shows the graphical representation of predicted results.

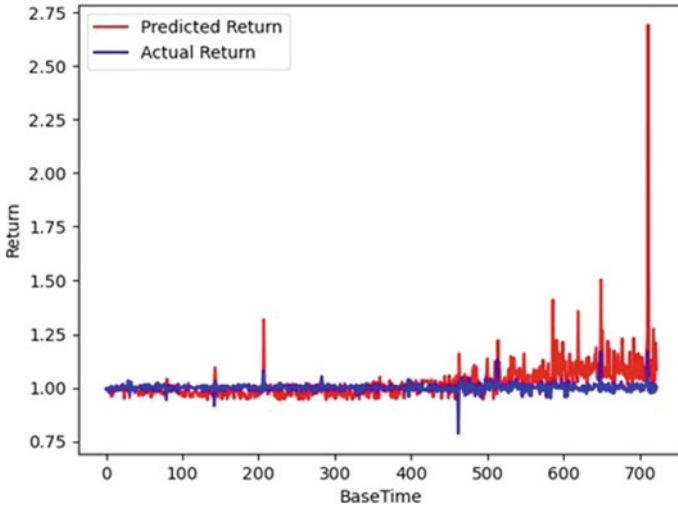


Fig. 4 Module 3 stock market prediction using PCA with ANN

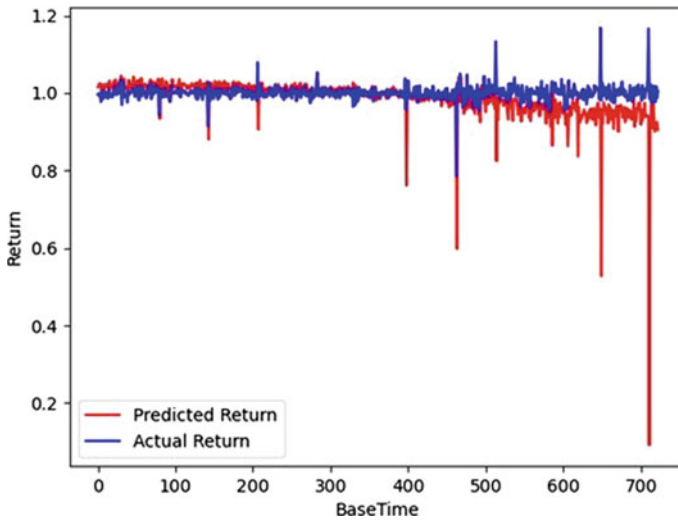


Fig. 5 Module 4 stock market prediction using PCA with ANN

4 Test Results

We can continuously check our data against expected data by using statistical rules that must satisfy. If the test and training splits are statistically different, then training data will not help predict the test data. We need to monitor the statistical properties

of the splits we make. Some of the important conditions were kept in mind while working on this project:

- The features had been scaled properly.
- Missing data records are replaced by mean or default value
- Outliers are scaled or removed.

The common metrics used for evaluating any machine learning models are:

- Mean absolute error
- Mean squared error
- R2
- Precision and recall
- Accuracy

For our model, mean squared error might be the best option to check how good our model is performing. Accuracy might also be an option as suggested by studies. So, we are showing both the results in all the four modules of our project.

We did the following for improving the performance and the outcome can be seen in Tables 1 and 2:

- Added more features apart from the features available in the dataset.
- Tune the hyperparameters by trial and error.
- Reduced the dimensions of features by using PCA in module 3 and 4.

5 Conclusion

From the above results and all the four-graphs plotted, we can conclude the following points from our analysis:

1. ANN seems to give the low accuracy compared with the other three modules.
2. When PCA is applied to reduce the dimensionality of dataset, it drastically affects the results of ANN and improves the result by a very high rate. In our case, it improved ANN's accuracy from 32 to 73%, and also reduced the MSE from 0.0902 to 0.0085 which is a huge improvement.
3. SVM is one of the best algorithms, and we found for stock market prediction as we got the best results with this.
4. Using only PCA does not necessarily improve the results. As we can see that it improved the results of ANN with a huge margin but it is not that effective with SVM it reduced the MSE but at the same time accuracy went down by 4%.

In addition to our experiment, we need a fair domain knowledge of stock market to generate new features which can improve the performance of this project like we have generated only few new features from our dataset for our project but a lot more can be generated. The more features we can add, and chances are that efficiency of our project will improve. So, we tried to include that in our project and combined

PCA with ANN and SVM. Some other approaches in this field can be worked on as well, and we can hope for better results as our results are quite satisfactory from our project.

Table 1 Test results and outcome

Test case number	Test case	Input	Expected result	Observed result
T-DPP1.1	Check for missing values	MSFT.csv	Missing values are to be replaced with mean	No missing value found
T-DPP1.2	Check for outlier	MSFT.csv	No outlier as within this time period there was no stock market crash	No outlier
T-SPLT1.1	Check if $\text{length}(\text{training}) + \text{length}(\text{testing}) = \text{length}(X)$	MSFT.csv, MSFT_X_learn.csv, MSFT_X_test.csv	No error raised	No error raised
T-TRN1.1	ANN	MSFT_X_learn.csv, MSFT_y_learn.csv	RMSE < 0.05	0.038
T-TST1.1	ANN	MSFT_X_test.csv, MSFT_y_test.csv	RMSE < 0.05 & comparable to T-TRN1.1	0.103, chances of overfitting
T-TST1.2	ANN	MSFT_X_test.csv, MSFT_y_test.csv	Accuracy > 50%	32%
T-TRN2.1	SVM	MSFT_X_learn.csv, MSFT_y_learn.csv	RMSE < 0.05	0.001
T-TST2.1	SVM	MSFT_X_test.csv, MSFT_y_test.csv	RMSE < 0.05	0.001
T-TST2.2	SVM	MSFT_X_test.csv, MSFT_y_test.csv	Accuracy > 50%	88%
T-TRN3.1	PCA with ANN	MSFT_X_P_learn.csv, MSFT_Y_P_learn.csv	RMSE < 0.05	0.005
T-TST3.1	PCA with ANN	MSFT_X_P_test.csv, MSFT_Y_P_test.csv	RMSE < 0.05	0.008
T-TST3.2	PCA with ANN	MSFT_X_P_test.csv, MSFT_y_P_test.csv	Accuracy > 50%	73%
T-TRN4.1	PCA with SVM	MSFT_X_P_learn.csv, MSFT_Y_P_learn.csv	RMSE < 0.05	0.001
T-TST4.1	PCA with SVM	MSFT_X_P_test.csv, MSFT_Y_P_test.csv	RMSE < 0.05	0.003
T-TST4.2	PCA with SVM	MSFT_X_P_test.csv, MSFT_Y_P_test.csv	Accuracy > 50%	84%

Table 2 Module-wise test results

Module	Mean square error (MSE)	Accuracy (in %)
Module 1 Using ANN	0.0902	32
Module 2 SVM	0.0018	88
Module 3 PCA + ANN	0.0085	73
Module 4 PCA + SVM	0.0013	84

References

1. Kimoto T, Asakawa K, Yoda M, Takeoka M (1990) Stock market prediction system with modular neural networks. In: 1990 IJCNN international joint conference on neural networks, 1990, vol 1. pp 1–6. <https://doi.org/10.1109/IJCNN.1990.137535>.
2. Mizuno H, Kosaka M, Yajima H, Komoda N (1998) Application of neural network to technical analysis of stock market prediction, studies in information and control , vol 7, no3, pp 111–120
3. Khoa NLD, Sakakibara K, Nishikawa I (2006) Stock price forecasting using back propagation neural networks with time and profit based adjusted weight factors. In: SICE—ICASE international joint conference 2006 18–21 Oct 2006 in Bexco, Busan, Korea
4. Khan ZH, Alin TS, Hussain Md A (2011) Price prediction of share market using artificial neural network (ANN). *Int J Comput Appl* 22(2): 0975–8887
5. Yetis Y, Kaplan H, Jamshidi M (2014) Stock market prediction by using artificial neural network. In: 2014 world automation congress (WAC), 2014, pp 718–722. <https://doi.org/10.1109/WAC.2014.6936118>
6. Adebisi Ayodele A, Ayo Charles K, Adebisi Marion O, Otokiti Sunday O (2012) Stock price prediction using neural network with hybridized market indicators. *J Emerging Trends Comput Inf Sci* 3(1). ISSN 2079-8407
7. Waqar M, Dawood H, Guo P, Shahnawaz MB Ghazanfar MA (2017) Prediction of stock market by principal component analysis. In: 2017 13th International conference on computational intelligence and security (CIS), 2017, pp 599–602. <https://doi.org/10.1109/CIS.2017.00139>
8. Iacomin R (2015) Stock market prediction. In: 2015 19th international conference on system theory, control and computing (ICSTCC), 2015, pp 200–205. <https://doi.org/10.1109/ICSTCC.2015.7321293>
9. Joseph E, Mishra A, Rabiou I (2019) Forecast on close stock market prediction using support vector machine (SVM). *Int J Eng Res Technol (IJERT)* 08(02)
10. Shah D, Isah H, Zulkernine F (2019) Stock market analysis: a review and taxonomy of prediction techniques. *Int J Financial Stud* 7(2):26. <https://doi.org/10.3390/ijfs7020026>

Hybrid Features-Based Ensembled Residual Convolutional Neural Network for Bird Acoustic Identification



Hari Theivaprakasham, V. Sowmya, Vinayakumar Ravi,
E. A. Gopalakrishnan, and K. P. Soman

Abstract Bird audio identification is one of the challenging fine-grained tasks due to various complexity in the signal. In the current work, we present a new bird audio dataset from the Indian subcontinent and propose a novel hybrid features-based ensembled residual convolutional neural network to identify bird audios from the Indian subcontinent. We utilized mel-frequency cepstral coefficients (MFCC) and melspectrogram features to train the neural network. We compared the results of the proposed model with other machine learning and deep learning models. The results show that our proposed model achieved the best accuracy of 92% and best F1-score of 91% on using modified ResNet50 model. The dataset and the experimental codes are available at GitHub (<https://github.com/Theivaprakasham/iccdn-2021-birdcall-id>).

Keywords Bird call identification · MFCC · Deep learning · Neural networks · Spectrogram

H. Theivaprakasham (✉) · V. Sowmya · E. A. Gopalakrishnan · K. P. Soman
Center for Computational Engineering and Networking (CEN), Amrita Vishwa Vidyapeetham,
Coimbatore 641112, India
e-mail: cb.en.p2cen20026@cb.students.amrita.edu

V. Sowmya
e-mail: v_sowmya@cb.amrita.edu

E. A. Gopalakrishnan
e-mail: ea_gopalakrishnan@cb.amrita.edu

K. P. Soman
e-mail: kp_soman@amrita.edu

V. Ravi
Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia
e-mail: vravi@pmu.edu.sa

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
S. Dhar et al. (eds.), *Advances in Communication, Devices and Networking*, Lecture Notes
in Electrical Engineering 902, https://doi.org/10.1007/978-981-19-2004-2_39

437

1 Introduction

Birds are a group of animals belonging to the class Aves adapted to fly using their wings. Birds play a complex essential role in our environment, providing direct and indirect benefits to humankind. Birds are known to help pollinate flowers, feed on insects and rodents and thus help increase the agricultural yield, revenue and reduce the use of insecticides. Birds also maintain the food chain and control the population of both insects and pests of agriculture. They also play an essential role in the stabilization of the ecosystem [1]. In recent decades, the bird population has been drastically reducing due to global warming, industrialization and urbanization [2]. A study by Kenneth et al. [3] reveals the net loss of nearly 3 billion birds (29%) from the bird population estimates of 1970 in North America. Monitoring birds has become essential to avoid the further decline of the bird population. The conventional bird identification uses visual cues and bird audio to manually identify the bird using field guides, online resources, DNA analysis, and morphometry studies. However, these methods are costly and highly time-consuming. The continuous monitoring of the birds is a difficult task unless the identification gets automated by intelligent systems. The recent advancements in artificial intelligence have proven solutions in various fields of wildlife monitoring. In the field, sighting a bird to observe its visual features is very difficult than hearing the bird's audio. Therefore, building an intelligent identification system based on bird audios will be beneficial in assessing and monitoring the bird population. The convolutional neural networks (CNN) [4] are extensively used in multiple domains, namely computer vision [5, 6], natural language processing, speech processing, etc. [7]. Hence, a CNN-based research on bird audio identification can be used to build an automated identification system.

The rest of the sections are organized as follows: Sect. 2 includes the literature survey, Sect. 3 discusses the proposed methodology, results and discussions are placed in Sect. 4, and finally, conclusion and future works are placed in Sect. 5.

2 Related Works

Earlier researchers have used various machine learning techniques and deep learning techniques to identify bird audio recordings. Fagerlund [8] used support vector machines (SVM) and decision trees (DT) to classify the MFCC features of bird audio recordings. Ágnes et al. [9] studied the effect of the RGB and grayscale spectrograms on CNN architecture. Kahl et al. [10] developed BirdNet, a deep CNN model to identify North American and European bird species. Hidden Markov model [11] was used to identify bird audio by analysing the spatio-temporal variations of the audio. Gupta et al. [12] used a hybrid CNN and recurrent neural network (RNN) to achieve 67% accuracy over the 100 different bird species.

The most commonly used features for bird audio identification were MFCC and spectrogram features as they have unique properties. MFCC is compressible cepstral

coefficients representation from the time-cepstrum domain, which can help identify high-level features, whereas spectrogram which contains more features also helps to visually segregate the features in the time-frequency domain. Our literature survey noted that the studies on the ensembled

In this paper, we propose to use CNN models to build a bird audio identification system. We

- Aim to build a new bird audio dataset for the Indian subcontinent.
- Benchmark the dataset on various machine learning models and deep learning models.
- Experiment and analyse the performance of proposed hybrid feature ensembled model with the traditional models.

3 Materials and Methods

3.1 Dataset

We collected bird audio recordings from the Xeno-Canto bird audio repository.¹ Xeno-Canto is an open-source audio repository where anyone from the world can record and upload bird audio recordings. Each bird audio recording is supported with meta-data consisting of the recording details, namely sampling rate, recording date and time, location of recording, genus, species, subspecies and common names of the bird. Most of the available works on bird audio recognition had been done on bird species from the regions of Europe and North America. Still, bird audio recordings from the Indian subcontinent are less studied. Hence, we selected the Indian subcontinent as our focus region and collected the bird audio recordings for the birds with more than 30 recordings. Finally, we collected 280 bird species audio recording, and the common names from each audio file's meta-data were considered the labels.

3.2 Data Preprocessing

Each collected bird audio recordings were in various lengths with multiple bird calls. We performed the following preprocessing operations to create our data for the training.

1. We separated every bird calls from the recordings.
2. We resampled the audio recordings to 16,000 Hz and normalized the recordings.
3. We padded the sampled audio with zeros to the audio, which was less than 1-s.

¹ <https://xeno-canto.org/>.

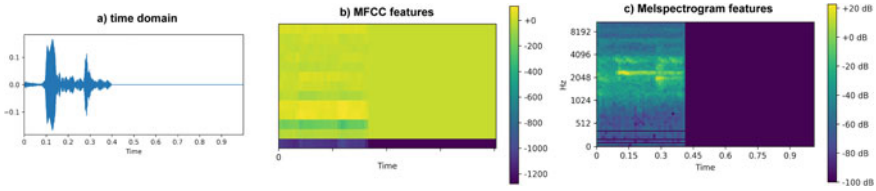


Fig. 1 a Time domain representation of the sample bird audio from the dataset; b Sample MFCC features; c Sample melspectrogram features

4. We extracted 100 MFCC features from the bird call with 512 Fast Fourier Transform (FFT) coefficients, 100 seconds of hop length, and 499 mel-filters.
5. We extracted the melspectrogram of the resampled bird calls with 512 FFT coefficients, 100 seconds of hop length, and 499 mel-filters. Figure 1 represents a sample image of the waveplot of the bird audio recording along with the extracted MFCC and melspectrogram features.
6. For training the SVM and random forest (RF) models, we flattened the melspectrogram and MFCC features into single dimensional vector.

Finally, we ended up with 67,488 recordings. We split the dataset with a ratio of 80:20 into training (53,991 recordings) and testing (13,497 recordings).

3.3 Methodology

We explored both machine learning and deep learning algorithms in this experiment. We used machine learning techniques, namely SVM and RF, to benchmark our dataset.

Random Forest: RF is an ensemble decision tree-based algorithm used for classification and regression tasks. We input the flattened features of bird audio recording into the RF model and train the model to identify the bird audio. To identify the best hyperparameters of the RF model, we performed a grid search by varying the maximum depth parameter between 5 and 200 and the number of estimators parameter between 5 and 150.

Support Vector Machines: SVM is one of the more widely used supervised machine learning algorithms that separate the data into different classes using optimization. We input the flattened features of bird audio recording into the SVM model to identify the bird audio. We selected a linear kernel function for the training process and performed a grid search to identify the optimal hyperparameters by varying the ‘C’ parameter between 0.1 and 1000 and the ‘gamma’ parameter between 1 and 0.00001.

Residual convolutional neural network: CNN algorithms have been proven to outperform classical machine learning models in many real-world applications [13, 14]. A variant of CNN named residual network [15] is known to perform better

in multiple domains of computer vision and time-series data. Residual networks mainly differ from the sequential models namely AlexNet and VGG16 models in the model architecture. The residual networks have skip-connections in between the convolutional blocks which helps to retain the memory of the previous layer and helps explore deeper high-level features from the data. This modification also helps to overcome the problem of vanishing gradient during the training of the model. In this study, we select the two pretrained architectures namely ResNet18 and ResNet50. These two models were trained on the corpus of ImageNet Dataset that consists of million of images belonging to 1000 classes.

We modified the pretrained ResNet model in two aspects: (1) We changed the number of input channels of the first 2D convolutional layer to 1 from the earlier three channels. (2) We averaged the three input channels weights of the first 2D convolutional layer to a single-channel weight.

The above two modifications enabled the use of single-channel MFCC and mel-spectrogram features to train a pretrained ResNet model.

Proposed Hybrid Ensembled CNN model The proposed hybrid ensembled CNN model consists of two residual networks. The two models are trained separately on MFCC features and melspectrogram as inputs. The inputs are simultaneously passed to both of the models. The output probabilities from the last layer of the two networks are then averaged to output the final prediction probabilities. Figure 2 illustrates the proposed hybrid ensemble residual CNN model.

Hyperparameters and Training: Hyperparameters are essential values that play a crucial part in training a neural network. We used the Adam optimizer function, the learning rate of 0.00001, batch size of 64 and the cross-entropy loss function to update the model parameters. Transfer learning technique was used to train the model. We froze the parameters of the pretrained model and trained only the final dense layer for five epochs. Then we unfroze all the model layers and trained the model for 50 epochs. We evaluated the model using the metrics namely accuracy, F1-score, precision and recall.

All the experiments were carried out on Jarvis Cloud² with 32 GB CPU RAM, 8-core Intel(R) Xeon(R) CPU @ 2.30 GHz and Nvidia A100 40 GB RAM. We used Fastai v2.3³ and Pytorch v1.10⁴ for training deep neural network and Scikit-learn v1.0 and Cuml libraries for training machine learning algorithms.

4 Experimental Results and Discussion

We performed several experiments on the dataset using SVM, RF, ResNet18, ResNet50 and our proposed hybrid ensembled CNN model.

² <https://cloud.jarvislabs.ai/>.

³ <https://docs.fast.ai/>.

⁴ <https://pytorch.org/>.

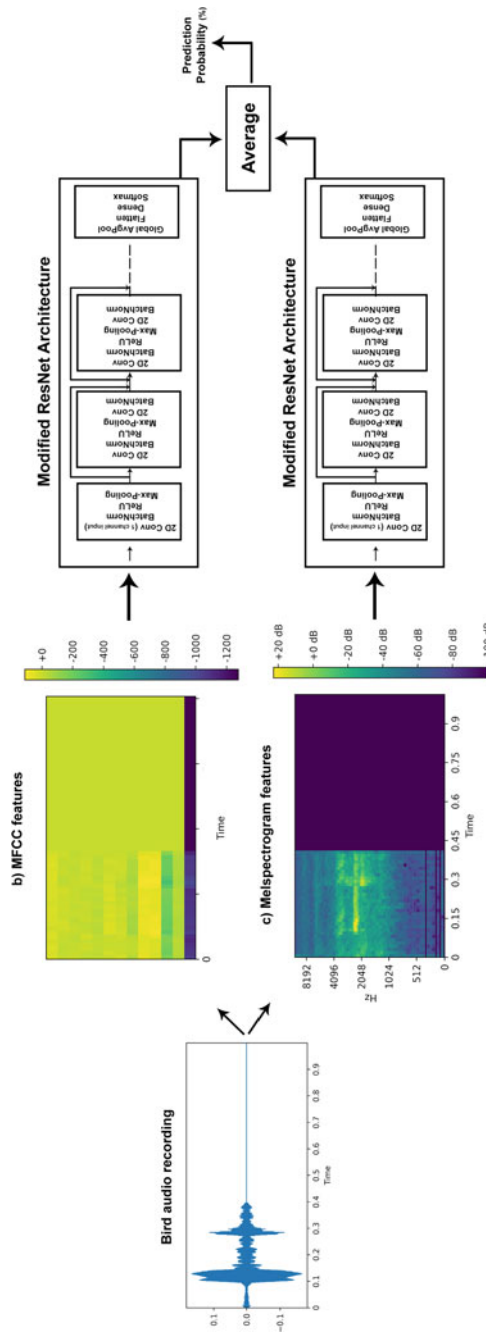


Fig. 2 Proposed Ensembled hybrid feature Residual connection-based convolutional neural network

Table 1 Performance of various model with MFCC features as inputs

Models	Accuracy	F1-score	Precision	Recall
Support vector machines	0.68	0.58	0.55	0.62
Random forest	0.67	0.64	0.62	0.66
ResNet18	0.73	0.69	0.75	0.65
ResNet50	0.81	0.77	0.75	0.81

Table 2 Performance of various model with melspectrogram features as inputs

Models	Accuracy	F1-score	Precision	Recall
Support vector machines	0.71	0.64	0.61	0.68
Random forest	0.69	0.68	0.65	0.72
ResNet18	0.89	0.88	0.92	0.85
ResNet50	0.91	0.9	0.91	0.88

In the first set of experiments, we trained SVM, RF, ResNet18 and ResNet50 models separately on both melspectrogram and MFCC features as inputs from the training dataset. In the second set of experiments, we trained our proposed hybrid ensembled model on the training dataset with: (1) Both inputs as melspectrogram with trained ResNet18 and ResNet50 models. (2) Both inputs as MFCC features with trained ResNet18 and ResNet50 models. (3) First input as MFCC feature and second input as melspectrogram on both trained ResNet18 and ResNet50 models. We evaluated the performance metrics (accuracy, F1-score, precision and recall) on the test dataset for all the experiments.

We observe that the best hyperparameters for training SVM were $C = 10$ and $\gamma = 1$. RF provided the best results with maximum depth = 50 and the number of estimators = 150. Table 1 represents the results obtained on different model using MFCC features. We observed that the ResNet50 model yielded the best accuracy of 81%, followed by the ResNet18 model. The machine learning models, namely SVM and RF, presented the least accuracy of 68 and 67%.

The results obtained on different models using melspectrogram features are tabulated in Table 2. Similar to the results of using MFCC features, the ResNet50 models performed the best with the highest accuracy of 91%, followed by the ResNet18 model with 89% accuracy. The SVM and RF models performed similarly with 71% and 69% as accuracy.

The results obtained on the proposed hybrid model, with both inputs as either melspectrogram or MFCC features, are tabulated in Table 3. When used similar feature input models as ensembles, the performance metrics remained the same as ResNet50 model without variations.

Table 3 Performance of the proposed model with same features as inputs

	Model-1	Model-2	Accuracy	F1-score	Precision	Recall
Inputs—MFCC	ResNet18	ResNet50	0.81	0.77	0.75	0.81
Inputs— Melspectrogram	ResNet18	ResNet50	0.89	0.89	0.89	0.88

Table 4 Performance of proposed hybrid ensemble model with both MFCC and melspectrogram features as inputs

Input 1—MFCC	Input 2— Melspectrogram	Accuracy	F1-score	Precision	Recall
ResNet18	ResNet18	0.9	0.89	0.92	0.87
ResNet50	ResNet50	0.92	0.91	0.92	0.88

Table 4 represents the results obtained on the proposed hybrid model with the first input as MFCC features and the second input as melspectrogram features. The results on performance metrics improved when we used differential input features for our proposed model. We observed that the ResNet50 model, when used as a feature extractor of both MFCC and melspectrogram, achieves 1% better accuracy on the test dataset than other trained models.

Overall, we observed that the melspectrogram offered better accuracy than MFCC features. However, our proposed model that ensemble both features had the best accuracy. The result shows that some bird audio recordings can be classified correctly using MFCC features, whereas other bird audios need melspectrogram features to be classified correctly. The observation from the results shows that additional features would enable the model to generalize and yield better results. We also note that increasing the depth of the residual network from 18 to 50 also increased the accuracy. This outcome indicates the efficiency of the residual network to explore deeper features from the audio recording.

5 Conclusion and Future Works

This study proposed and evaluated a novel hybrid ensemble residual connection convolutional neural network-based bird audio identification system. The MFCC and melspectrogram features extracted from the bird audio recording were used as inputs for the bird audio identification. We explored and analysed the performance of both the machine learning and deep learning models, namely SVM, RF, ResNet18, ResNet50 and our proposed model. We observed that deep learning models provided better performance than machine learning models. Compared to all the models presented in the current work, our proposed model had the best accuracy of 92% when using the customized ResNet50 model. The result clearly showed the importance of ensembling MFCC and melspectrogram features to achieve better results. Our future

works will extend the currently proposed models with wavelet filter-based features. We will also explore using transformer-based models to extract spatio-temporal data for bird audio classification and multiple bird audio detection.

References

1. Morante-Filho JC, Faria D (2017) An appraisal of bird-mediated ecological functions in a changing world. *Trop Conserv Sci* 10:1940082917703339. <https://doi.org/10.1177/1940082917703339>
2. McClure CJW, Rolek BW, McDonald K, Hill GE (2012) Climate change and the decline of a once common bird. *Ecol Evol* 2(2):370–378. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.95>
3. Rosenberg KV, Dokter AM, Blancher PJ, Sauer JR, Smith AC, Smith PA, Stanton JC, Panjabi A, Helft L, Parr M, Marra PP (2019) Decline of the North American avifauna. *Science* 366(6461):120–124
4. Lecun Y, Bengio Y (1995) *Convolutional networks for images, speech, and time-series*. MIT Press
5. Unnikrishnan A, Sowmya V, Soman KP (2019) Deep learning architectures for land cover classification using red and near-infrared satellite images. *Multimedia Tools Appl* 78(13):18379–18394. <https://doi.org/10.1007/s11042-019-7179-2>
6. Reshma R, Sowmya V, Soman KP (2018) Effect of Legendre-Fenchel denoising and SVD-based dimensionality reduction algorithm on hyperspectral image classification. *Neural Comput Appl* 29(8):301–310. <https://doi.org/10.1007/s00521-017-3145-y>
7. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444. <https://doi.org/10.1038/nature14539>
8. Fagerlund S (2007) Bird species recognition using support vector machines. *EURASIP J Adv Signal Process* 2007(1):38637. <https://doi.org/10.1155/2007/38637>
9. Ince Á, Jancsó HB, Szilágyi Z, Farkas A, Sulyok C (2018) Bird sound recognition using a convolutional neural network. In: 2018 IEEE 16th international symposium on intelligent systems and informatics (SISY), pp 295–300
10. Kahl S, Wood CM, Eibl M, Klinck H (2021) BirdNET: a deep learning solution for avian diversity monitoring. *Ecol Inform* 61:101236. <https://www.sciencedirect.com/science/article/pii/S1574954121000273>
11. Jančović P, Köküer M, Zakeri M, Russell M (2016) Bird species recognition using HMM-based unsupervised modelling of individual syllables with incorporated duration modelling. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 559–563, Mar 2016
12. Gupta G, Kshirsagar M, Zhong M, Gholami S, Ferres JL (2021) Comparing recurrent convolutional neural networks for large scale bird species classification. *Sci Rep* 11(1):17085. <https://doi.org/10.1038/s41598-021-96446-w>
13. Srinivasan S, Ravi V, Sowmya V, Krichen M, Noureddine DB, Anivilla S, Soman KP (2020) Deep convolutional neural network based image spam classification. In: 2020 6th conference on data science and machine learning applications (CDMA), pp 112–117
14. Ganesan S, Ravi V, Krichen M, Sowmya V, Alroobaea R, Soman KP (2021) Robust malware detection using residual attention network. In: 2021 IEEE international conference on consumer electronics (ICCE), pp 1–6, Jan 2021
15. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. vol 2016-Dec. IEEE Computer Society, pp 770–778, Dec 2016

Lung Disease Prediction Using Deep Learning



Debasree Mitra, Pranati Rakshit, Anjali Jha, Dristi Dugar,
and Kamran Iqbal

Abstract The evolution of deep learning has enhanced the technique of identifying and classifying lung diseases into various categories using medical images. This project aims to build lung disease detection models using deep learning to identify future potential and thus to efficiently observe and visualize the recent and upcoming trends in this domain. Identifying and discovering lung disease at an early stage has become a vital part of the medical domain because this would facilitate patient's subsequent clinical management. The project primarily focuses on pneumonia as well as considering the breathing problems of patients. Deep learning and machine learning have served the utmost significance in detecting such lung diseases at a prior stage. This enhancement has contributed much to the doctors and medical systems to provide early treatment to patients. In this project, convolutional neural network (CNN) is used to predict lung disease (pneumonia) from chest X-ray images using machine learning and deep learning frameworks.

Keywords Lung disease prediction · Deep learning · Machine learning · Chest X-ray images · Convolutional neural networks (CNNs) · Pneumonia

1 Introduction

Medical X-ray images provide a gateway to diagnose body parts such as the chest, bones, skull and teeth. Medical pioneers have used this same approach for several decades to investigate and analyse fractures or abnormalities in body organs.

Machine learning and deep learning play a pivotal role in performing the algorithms accurately to categorize the chest X-ray peculiarities besides providing a large window for the further prediction of deadly lung diseases. Over the past few

D. Mitra (✉) · P. Rakshit · A. Jha · D. Dugar · K. Iqbal
Department of Computer Science and Engineering, JIS College of Engineering, MAKAUT,
Kolkata, India
e-mail: debasree.mitra2005@gmail.com

P. Rakshit
e-mail: pranati.rakshit@jiscollege.ac.in

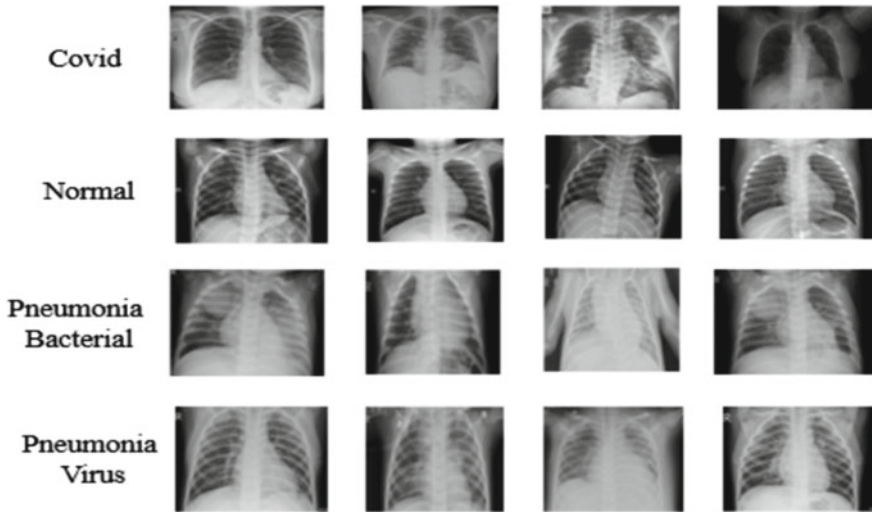


Fig. 1 Different types of lung diseases

years, digital technology has gained the utmost importance globally. Systems based on deep learning have been put into play to elevate the exactness of image classification. Deep networks showed extraordinary accuracies in such predictions. This success led the researchers to enforce the deep networks to chest X-ray images for lung disease prediction, and the classification paving their way out showed these deep networks masterfully extract beneficial features that distinguish different categories of images provided to the model. The most conventional deep learning framework used is the convolutional neural network (CNN). CNN provides a unique power to extract features from images at different levels.

This research paper may provide other doctors and researchers the right path for bringing to light the lung diseases such as pneumonia, asthma, tuberculosis, SARS, ARDS and COPD with the help of deep learning methodology as shown in Fig. 1 [1]. Several works regarding the research have been carried out on the diagnosis of lung diseases using artificial intelligence methodologies. Neural networks and artificial immune systems are already being used to diagnose fatal diseases like pneumonia and chronic obstructive pulmonary disease (COPD).

2 Literature Survey

The most severe chronic lung diseases involve chronic obstructive pulmonary diseases (COPD), pulmonary sarcoidosis, asthma, pneumonia and interstitial lung diseases, India, holding a second rank in terms of population with 1.3 billion people

residing in 29 states and seven union territories often vary widely in terms of demography, economy and ecology, through which respiratory health gets affected [2]. Table 1 [3] shows the ten most common causes of death in 2008. In India, the estimation indicates the cases of pneumonia have reduced approximately by 41% between the years 2000 and 2015, which in turn is greater than 22% reduction in pneumonia cases observed globally during the same period. Nevertheless, pneumonia still remains a crucial cause of morbidity among children country-wise [4].

Numerous researches have faced the difficulty of classifying the images through such a high degree of accuracy. Here are a few citations from our paper regarding the CNN model.

Rubin et al. [5] created a model to identify basic thoracic illness in frontal chest X-ray images. To accomplish large-scale automatic recognition of these pictures, the MIMIC-CXR dataset was employed. The dataset was divided into three sections: testing, training and validation, with a piece accounting for 20%, 70% and 10% of the entire lot. To improve overall performance, normalization of pixel was utilized.

Lakhani and Sundaram [6] developed such model to classify pulmonary tuberculosis. Chest X-ray images were also classified using transfer learning models such as AlexNet and GoogleNet. About 68%, 17.1% and 14.9% of the dataset were split into training set, validation sets and testing sets, respectively. To get the ultimate model with an AUC of 0.99, techniques like augmentation of data as well as pre-processing were used. The model’s recall and veracity were 97.3% and 100%, respectively.

Guan et al. [7] build an AG-CNN model to detect diseases related to the thorax. For the detection of such diseases from X-ray images of the chest, the chest X-ray14 dataset was applied. For categorization, a local and global branch attention-guided CNN was utilized. With an AUC of 0.868, their model outperformed the other models.

To classify chest X-ray images into pneumonia and 14 other diseases, Rajpurkar et al. [8] used chest X-ray14 dataset to train the model. They compared their ChXNet model (121-layer model) with academic radiologists in practice. Their model has an F1 score of 0.435, which was higher than radiologists’ F1 score of 0.387.

Table 1 Statistics of world health 2011

Reasons for deaths	Worldwide data	WHO European region
Ischaemic heart disease	7.3 million (12.8%)	2.40 million (24.7%)
Cerebrovascular disease	6.2 million (10.8%)	1.40 million (14.0%)
Lower respiratory infections	3.5 million (6.1%)	0.23 million (2.3%)
COPD	3.3 million (5.8%)	0.25 million (2.5%)
Diarrhoeal diseases	2.5 million (4.3%)	0.03 million (0.3%)
HIV/AIDS	1.8 million (3.1%)	0.08 million (0.8%)
Trachea/bronchus/lung cancer	1.4 million (2.4%)	0.38 million (3.9%)
Tuberculosis	1.3 million (2.4%)	0.08 million (0.8%)
Diabetes mellitus	1.3 million (2.2%)	0.17 million (1.7%)
Road traffic accidents	1.2 million (2.1%)	0.12 million (1.2%)

Krizhevsky et al. [9] trained this sort of model with five convolutional sheets, some of which were followed by max pooling layers and three fully connected layers. Such networks contained parameters over 60 million. This model earned an error rate of 17% while using dropout.

Simonyan and Zisserman [10] created a supreme model that used many tiny kernel-sized filters to reach an exactness of 92.7%. The ImageNet dataset was used to train this model and then proposed to the ILSVRC 2014 competition.

Xu et al. [11] created a convolutional neural network for brain tumour MRI classification and segmentation. This model implemented a plethora of techniques, such as data augmentation, feature selection and pooling. This model achieved a validation accuracy of 97.5% for classification and 84% for segmentation on 256 256 pixels sized frontal chest radiographs that were fed to a deep convolutional neural network to detect anomalies.

Anthimopoulos et al. [12] developed this replica with five convolution sheets, leaky ReLU, average pooling and three fully connected layers to detect interstitial lung disease patterns in a dataset comprising 14,696 images. This model had an accuracy of 85.5%.

3 Chest X-ray Data Analysis

The datasets used in the project are reported in this section. This is done to provide us with relevant information on the datasets. Dataset description and view positions of the chest X-ray images are discussed briefly in this section.

3.1 Dataset Description

As a matter of concern, public datasets have been chosen because of their availability to the public, whereas private datasets are not accessible without permission. Here, the model is exposed to two different datasets to protract a conclusion over the CNN model while visualizing the difference in results of pneumonia prediction.

- The first dataset contains 5933 images with resolutions varying from 1024×1024 to 2444×1800 .
- The second dataset contains a total of 5357 images with resolution varying from 1114×928 to 2290×2066 .

The entire dataset, one and two contain numerous classes of images with the name normal, pneumonia (virus and bacteria), syndromes such as severe acute respiratory (SARS), acute respiratory distress (ARDS), COVID-19 positive and negative, CT scan. For better visualization and accurate analysing the chest X-ray images, the diagnosis can be carried out in a more time-efficient manner by medical staff and

doctors to monitor the patient’s health accordingly. Moreover, by creating applications using intelligent machines, physicians could analyse the condition of patients better.

3.2 View Position

1. Posterior-anterior (PA) projection: The posterior-anterior (PA) projection in Fig. 2a is considered a standard position for determining matured chest radiograph as the light beams pass through the chest part of the torso from a posterior position to anterior position. The image was captured, while the patient was standing still against a film, holding his arms straight up in the air and having few deep breaths throughout the process to get good and clear pictures of both heart and lungs. These techniques ensure precise higher quality images shown in (b) and access the big heart size than the anterior–posterior (AP) position.
2. Anterior-posterior (AP) projection: The anteriorposterior (AP) projection in Fig. 3a is less considered because the light beams are restricted to pass through the chest wall. This condition occurs when the patients are unwell or too weak to even stand straight against a film. Hence, the projection is passed through the anterior to posterior position. This results in a lower quality image of the chest X-ray shown in (b), due to which capturing the figure of a clear and focussed heart becomes a tough job.

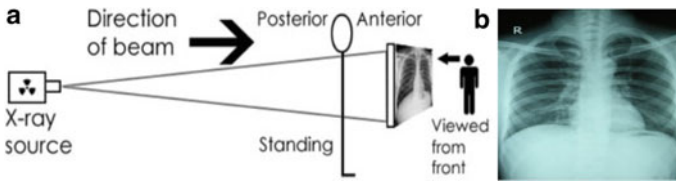


Fig. 2 a PA projection. b PA chest X-ray

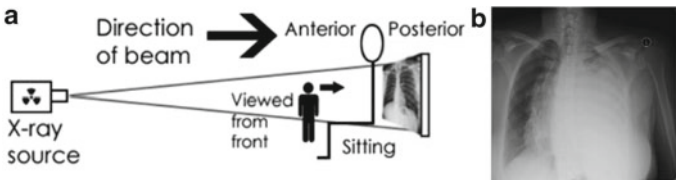


Fig. 3 a AP Projection. b AP chest X-ray

A comparative study demonstrates that it is pretty easy to find lung diseases in middle aged to elderly humans comparable to young ones. Any vague disease found in young people is easy to diagnose through primary treatments, nonetheless, the discovery of high-rated infections in middle-aged groups is a matter of concern [13, 14].

4 Description of Algorithm Used

Convolutional neural networks (CNNs) were primarily introduced by Haffner, Bengio, Bottou and Le Cun, in the late 1998s. The arrangement and anatomy of a mammal brain have played an important role towards inspiring the evolution of algorithms of machine learning, one of them being convolutional neural network (CNNs) in Fig. 4 [15], an unsupervised single layered that uses a hierarchical approach to build a neural network, a funnel-like structure that further results in a network where neurons are connected to form a fully connected network leading to training a model layer by layer to process the output. Several procedures need to be followed while classifying an image using CNN.

Firstly, a filter or kernel is made to slide over the input data (layer) pixel by pixel, the stride is considered to be one(by default) and multiplication of matrices is performed to obtain a feature map (matrix). The size of the filter may vary. This process is carried out on each convolution until a final feature map is achieved.

Secondly, the images are resized before the model undergoes training. This is done with the help of pooling and flattening layers. Pooling refers to continuous reduction in dimensionality to eliminate the computation in the network and the number of parameters. This leads to handling overfitting with less training time. Max pooling is the most frequent type which calculates the maximum value in each and every window. Hence, decreasing the size of the feature map while retaining the chief information. Flattening helps in converting original data into a 1-D array which would further be used as an input layer in the next step. This process is carried out to

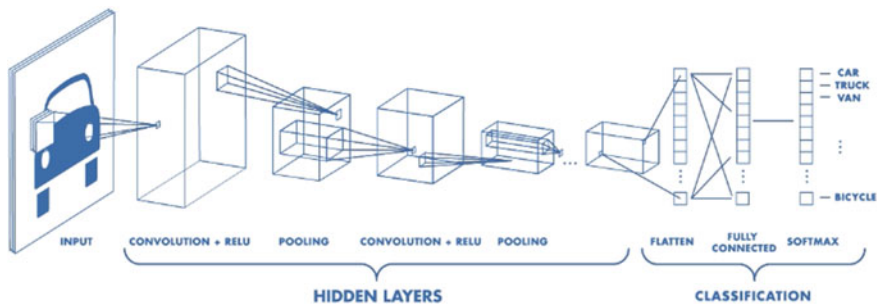


Fig. 4 Architecture of a CNN

extract significant features. Flattening assists in getting an output of fully connected neural networks after the complete convolution is done.

Lastly, to prevent overfitting of the model, batch normalization is considered to be the most efficient as it regularizes the output of the previous layers which in turn makes the output layers learn independently. Here, activation functions highly contribute when it comes to scale the input. Most widely used activation functions in neural networks are ReLu, Leaky ReLu, softmax Function, sigmoid function and so on. The choice of using a particular activation function depends on the features provided by a particular activation function. In this paper, the activation functions used are rectified linear unit (ReLu) and softmax activation. The advantage of implementing ReLu over any activation function is that it prevents the problem of vanishing gradient. The former function does not activate all the neurons at the same time, hence, it performs the nonlinear transformation of the input data. Hence, the output layer achieved is a fully connected layer. Every neuron of the latest layer is attached to all the neurons in previous film, thus forming a densely connected layer.

5 Methodology

5.1 Data (Image) Pre-processing Phase

To detect any lung disease, image dataset is considered to be the most relevant when it comes to using deep learning. These include medical photocopies of CT scans, X-ray images and many more. These medical images are available in the raw form which may not be understandable by machines. So, in order to make it machine readable, these images need to be pre-processed before moving on towards training.

In order to achieve a high training phase, these X-ray images are rescaled and the size is reduced by 255 NumPy array matrices. The images are then transformed from greyscale image to RGB for training purposes. Furthermore, various modifications are made on the resultant images and after carrying out data augmentation on this dataset to gain more relevant data out of it. Unnecessary data are removed, and the complete dataset is filtered and pre-processed. Feature extraction is then performed on this processed data to draw out the most cardinal features so as to prepare the data finally to be trained on the chosen model.

5.2 Training Phase

Since, in this paper, the dataset used for predicting lung disease is X-rays and CT scan images, deep learning algorithms fit best when it comes to training image datasets. There are many deep learning algorithms available that perform differently for different datasets. To compute and train large amounts of data, deep learning

introduces many multiple layers of neural networks. Few well-known deep learning algorithms of neural networks are convolutional (CNNs), recurrent (RNNs), multi-layer perceptron (MLPs), deep belief (DBNs), long short-term memory (LSTMNs) and so on. These algorithms are chosen accordingly based on the types of data to be dealt with and which model fits best to give desired output to the given problem statement. When the dataset contains time and date series, RNNs and LSTMNs are the algorithms that are considered the most. RNNs remember previous output and use it as current input while having hidden layers in between them. LSTM networks are a type of RNNs that works on sequence prediction problems. MLP networks can be used to deal with data that are nonlinearly separable, and in computational neuroscience, it contains multiple layers of neurons, and the use of back-propagation is a major factor here. To deal with image and video sequences of data, deep belief networks are mostly preferred since it uses a greedy algorithm approach. CNNs are used to get acquainted with image datasets as it takes tensors as input, and after passing filters over multiple layers of neurons and hidden layers, it succeeds in finding relationships between each pixel in the images. The datasets used in this paper are grey-scaled images that minutely need to be spatially connected among the pixels. Hence, CNN is used in this regard for generating the best confusion matrix. During training, a technique named dropout selects neurons randomly and drops them, thus regularizing deep neural networks. There is a fundamental difference between application of dropout to fully connected layers and to convolutional layers. To avoid overfitting, regularization techniques are used to appropriately fit the model by reducing the error. With the intent of standardizing the input to a layer, batch normalization is applied to the deep learning network. This aims to put a stop to the overfitting of model by reducing the number of training epochs. The stabilization attained by the method early stopping targets to improve model performance by stopping the training and specifying the increased number of training epochs.

5.3 Classification Phase

The problem statement of this paper is to predict lung disease and to determine whether a patient is suffering from pneumonia or not. The chosen algorithm, convolutional neural network, is a classification algorithm that predicts the output classifying it, corresponding to which it gives high accuracy. The two classes are normal (the value assigned is '0') and pneumonia (the value assigned is '1'). The extracted features are arranged into layers, and according to the working principle of CNN, filters are supposed to slide over these input layers as well as some hidden layers until it reaches and becomes a fully connected network. Learning rate, being a configurable hyperparameter, determines the step size. It serves a paramount role in optimizing and training the model. If the learning rate of a neural network model is too high, then the loss function may show undesirable divergent behaviour of the loss function. A desirable learning rate could be a prime difference between a model that presents

accurate results and a model that does not learn. Hence, reduction of learning rate becomes a momentous step in optimizing the model's performance. Therefore, an ideal learning rate should be in the range of 0.1–0.001.

The model is then fitted and has used a certain number of epochs until the desired accuracy is reached. Then, the model is tested over a certain amount of testing data and checked for how the model performs after getting trained on the selected features. This process results in classifying images into the respective classes it belongs to. As an outcome, the classified images are tagged with accuracy percentage which supports their belonging to that particular class (normal or pneumonia).

6 Result and Discussion

The overall implementation of the proposed CNN model is tested and presented in this section.

6.1 *Model Validation and Evaluation*

After training our model, a validation set was used for estimation purposes. The performance of the model was visualized by plotting various graphs to observe the general behaviour that our model learnt. The graphs between training loss versus training validation loss and training accuracy versus validation accuracy are represented for both datasets. The training of the model was performed in two ways on two different datasets.

The first, without the use of data augmentation. Figure 5a, b depicts the graphs without the use of data augmentation for dataset 1; (c) and (d) show the graphs with the use of data augmentation for dataset 2.

The second, with the use of data augmentation; (e) and (f) depict the graphs with the use of data augmentation for the dataset 1; (g) and (h) depict the graphs with the use of data augmentation for the dataset 2.

6.2 *Justification*

For training and validation of the model, two different datasets (chest X-rays images) were taken into consideration to demonstrate a comparative study which ensued in a better visualization on how the chosen algorithm (CNN) works. The model trained without data augmentation has resulted into overfitting in Table 2.

This model when introduced with data augmentation as it is about fabricating more data from the data we actually got, adding variance without losing the information the data carries. Doing this reduces the risk of overfitting, and generally, the accuracy on unseen data can be improved [16] in Table 3.

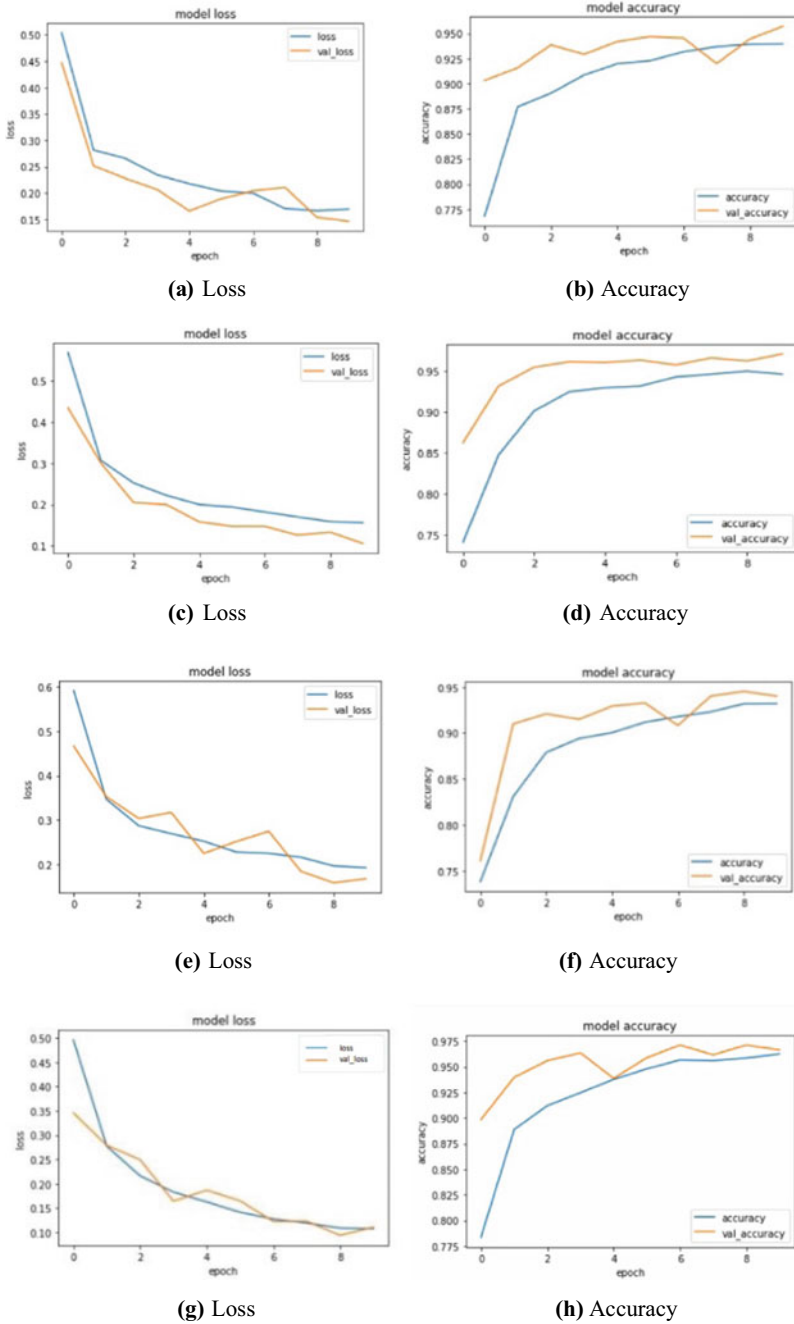


Fig. 5 a Loss. b Accuracy. c Loss. d Accuracy. e Loss. f Accuracy. g Loss. h Accuracy

Table 2 Model performance without data augmentation

S. No.	Name	Train accuracy (%)	Train loss (%)	Test accuracy (%)	Test loss (%)	Validation accuracy (%)	Validation loss (%)
1	Dataset 1	94.15	15.85	95.71	14.6	95.71	14.60
2	Dataset 2	94.65	15.77	97.03	10.5	97.03	10.57

Table 3 Performance of the model with data augmentation

S. No.	Name	Train accuracy (%)	Train loss (%)	Test accuracy (%)	Test loss (%)	Validation accuracy (%)	Validation loss (%)
1	Dataset 1	92	19.90	95.74	14.67	95.75	14.67
2	Dataset 2	96.07	10.90	96.64	11.05	96.65	11.05

6.3 Free-Form Visualization

In this paper, to evaluate the trained model, we tested it on five hundred random images from the test dataset. The rationale behind is to interrogate how our model performs when introduced to unseen dataset (images). The final results from both the datasets are below as Fig. 6a–d.

7 Conclusion

In the current paper, pre-training and the detection of lung disease became a great aid using convolutional neural network (CNNs). CNNs have three layers designed for the ease of detecting diseases at medical diagnostics. Hence, sliding filters over the input multiple times make the process worth determining the results earlier. According to the research, a lot of papers have been issued on lung diseases. The emergence of various lung diseases such as asthma, COPD and infections like influenza, pneumonia, lung cancer and so on has made it essential for the improvement and amplification of the technology and designing such early disease prediction models using deep learning.

TensorFlow and Keras have also played an influential role in supporting backend of the CNN model. Since, raw medical images (X-ray images) contain a lot of trivial data; hence, best algorithms were used for feature extraction and data pre-processing. The use of two datasets gave proper understanding of how the results vary on a large scale such as hospitals. The choice of suitable algorithm made the task more efficient as CNN is considered to be one of the best algorithms that targets to achieve best results in detecting lung diseases. Within less iterations, CNN accomplished a good precision model also with less number of epochs. The results so far obtained

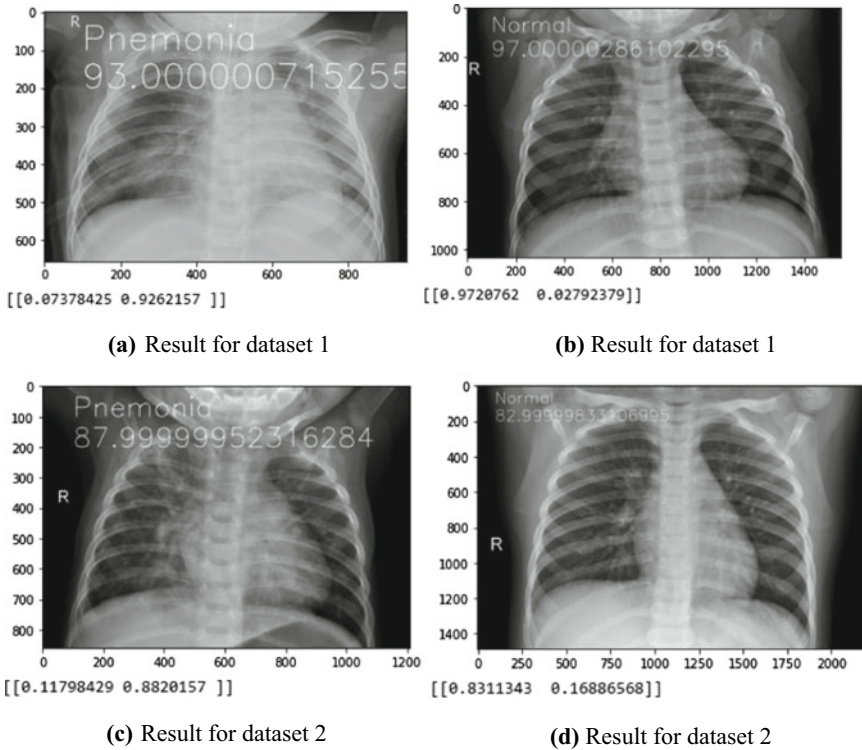


Fig. 6 **a** Result for dataset 1. **b** Result for dataset 1. **c** Result for dataset 2. **d** Result for dataset 2

demonstrate a fair performance of the CNN model used for this project. So, as a means to make the proposed CNN model through this paper to be actively used in hospitals and other medical diagnostics centres, further improvements can be made to achieve the accuracy to reach a peak point and also to detect lung diseases at a more precise level. This would serve a lot and is definitely going to make a difference in early diagnosis of lung diseases through its applications.

References

1. <https://medium.com/analytics-vidhya/classification-of-chest-xrays-using-pytorch-d50edd9ebb0>
2. [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(18\)30409-1/fulltext](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30409-1/fulltext)
3. <https://www.erswhitebook.org/chapters/the-burden-of-lung-disease/>
4. [https://www.thelancet.com/pdfs/journals/lanchi/PIIS2352-4642\(20\)30129-2.pdf](https://www.thelancet.com/pdfs/journals/lanchi/PIIS2352-4642(20)30129-2.pdf)
5. Rubin J, Sanghavi D, Zhao C, Lee K, Qadir A, Xu-Wilson M (2018) Large scale automated reading of frontal and lateral chest X-Rays using dual convolutional neural networks. arXiv preprint [arXiv:1804.07839](https://arxiv.org/abs/1804.07839)

6. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284(2):574–582
7. Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L, Yang Y (2018) Diagnose like a radiologist: attention guided convolutional neural network for thorax disease classification. arXiv preprint [arXiv:1801.09927](https://arxiv.org/abs/1801.09927)
8. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpan-skaya K, Lungren MP (2017) Chexnet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint [arXiv:1711.05225](https://arxiv.org/abs/1711.05225)
9. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
10. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
11. Xu Y, Jia Z, Ai Y, Zhang F, Lai M, Eric I, Chang C (2015) Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In: *2015 International conference on acoustics, speech and signal processing (ICASSP)*, pp 947–951
12. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 35(5):1207–1216
13. https://www.physio-pedia.com/Chest_X-Rays
14. https://www.radiologymasterclass.co.uk/tutorials/chest/chest_quality/chest_xray_quality_projection
15. <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>
16. <https://towardsdatascience.com/how-i-got-1-better-accuracy-by-data-augmentation-2475c509349a>

Implementation of Different Classification and Prediction Models on Skin Cancer Using Deep Learning Techniques



Debasree Mitra and Pranati Rakshit

Abstract Skin cancer is widely menacing forms of cancer in North America and South East Asia and some part of Australia also. The main reason of skin cancer is caused by damaged deoxyribonucleic acid (DNA) in skin cells of human body which is inherited from genetic disorder or mutations on the skins. Skin cancer is to gradually spreading over other body parts with acute pain and it is only curable in initial stages. That is always recommended to detect at early stages of cancer as we know that there are four stages of cancer. The skin cancer has high mortality rate all over the globe as compare to other types of cancer and its treatments are very expensive. This paper presents a detailed review of deep learning techniques like convolutional neural network for the early detection of skin cancer and their types. We have develop a 2-D CNN model and evaluated the model with different parameters and finally evaluated model with data augmentation and predict the incorrect probability for different types of Skin cancer for HAM 1000 datasets.

Keywords Skin diseases · Deep learning · Dermoscopic images · Convolutional neural networks (CNN) · Regularization · Data augmentation

1 Introduction

Skin cancer is one of the dangerous types of cancers. Many recent surveys show that the rate of patients infected by skin cancer is increasing every year as compared to other cancer types. Skin is the largest organ of the human body [1]. There are two types of skin cancer, one is melanoma and other is nonmelanoma skin cancer. Melanoma is very dangerous and rare deadly type of skin cancer. According to worldwide statistics of different studies shows that melanoma skin cancer cases are

D. Mitra (✉) · P. Rakshit
Department of Computer Science and Engineering, JIS College of Engineering, MAKAUT,
Kalyani, India
e-mail: debasree.mitra2005@gmail.com

P. Rakshit
e-mail: pranati.rakshit@jiscollege.ac.in

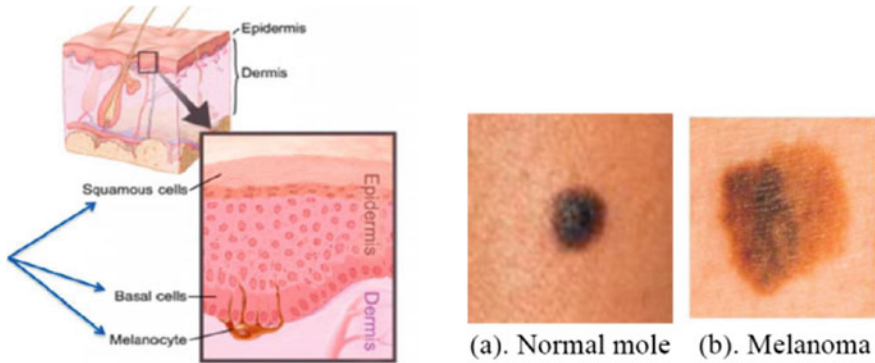


Fig. 1 Cross section view of human skin and samples of normal mole and melanoma

only 1% of total cases but they result in a high death rate [2]. There are many types of skin cancers. Melanoma have the power to increase or control melanocytes cell which are the skin surface of human body and get darker with different dark shades. It has different color shades like rosy pink, royal purple, azure, colorless as well. It has the ability to spread rapidly and create the situation more dangerous and fatal. The risk factors of skin cancer disease can be reduced by detecting on correct time or in an early stage. The significant reduction in the mortality rate can be achieved by detecting skin cancer in its early stages. Therefore, the identification and classification of this disease in its initial stages are important [1]. Melanoma is the deadly situation when healthy melanocytes cells are growing out of control and which creates a cancerous tumors. Melanoma may appear and affect any parts of the human body. Melanoma usually appears on those areas exposed to sun light. If it will be diagnosed at early stages otherwise the spreading is less to other body parts. Skin cancer death are very painful death. There are different types of melanoma skin cancer such as Nodular melanoma, Superficial spreading melanoma, Acral lentiginous, and Lentigo maligna. Most of the cancer cases lie under the umbrella of nonmelanoma categories, such as BCC: basal cell carcinoma, SCC: squamous cell carcinoma, and SGC: sebaceous gland carcinoma. BCC, SGC, and SCC are formed in the middle and upper layers of the epidermis [3]. These cancer cells have a low tendency of spreading to other body parts. Nonmelanoma cancers are not so harmful but treated easily as compared with melanoma cancers (Fig. 1).

2 Literature Survey

In 2015, Dr. Hannon, Baidaa M. ALSafy proposed to diagnose melanoma in early stages by the help of the image processing technique is used and succeeded the initial process to detect the pigment skin lesion is benign or malignant, and the later to recognize malignant melanoma skin type. They took the help of segmentation and

feature extraction. Diagnosis is done between benign and malignant melanoma categorizing different type of malignant melanoma such as the superficial, the nodular, the lentigo, and the acral melanoma [4]. Later researcher concentrated on identification of skin cancer by the help of pre-processed images using combinations of traditional machine learning and some popular image processing techniques to detect the stage of cancer. Images of the affected area were captured with the help of dermoscopes. Several algorithms like SVM, Naïve Bayed, and Neural Network have been proposed to detect skin cancer but most of the inputs are fed manually and which were very time consuming. Their main objectives were to develop a machine learning algorithm which requires minimal involvements of human [1, 4]. The first effective results on skin cancer classification was done by a pre-trained GoogLeNet Inception V3 CNN model came from Esteva et al. They have used 129,450 clinical skin cancer images including 3,374 dermoscopic images. The reported accuracy of classification was 72.1 ± 0.9 [5]. Deep neural networks is a set of interconnected nodes similar to the human brain in terms of neuronal interconnectedness and all these nodes work cooperatively to solve particular problems efficiently. Neural networks are getting trained for certain tasks and subsequently tested. Deep learning has huge capacity to deal with the large data sets, whereas machine were suffering some problems to deal with the huge and large datasets. The deep learning is considering as most sophisticated sub division of machine learning. All the deep learning algorithms are motivated by the functions and structures of the human brains. As human brain can think, visualize and take decision and act on some instances and similarly deep learning techniques are also capable to do all the activity like human. There are different application areas of deep learning such as in the research area of speech recognition, pattern recognition, medical healthcare, and bioinformatics. The image analysis and image classification has become very standard and sophisticated now a days and the challenge provided by ISIC in 2018 has made research in skin cancer more advanced and sophisticated now. Some researcher proposed automated system or mobile apps can detect skin cancer or other skin related diseases and now it is important to detect more accuracy in propose a good model in deep learning. Image classification got a new dimension when convolutional neural network (CNN) was introduced by Fukusjima in 1988 and later improved by Le Cunn in 1990. CNN is mainly mimic the human recognition system by providing plethora of image processing, computer vision, and classification techniques [6].

3 Skin Cancer Data Set

In this research paper, we have implemented the model using deep convolution neural networks or CNN and for this research purpose we have used **HAM10000** (“**Human Against Machine with 10,000 training images**”) dataset. In this datasets there are 10,015 **dermoscopic** images which are released as a training set for academic machine learning purposes and those are publicly available through the International Skin Imaging Collaboration (ISIC) archive [7]. It has seven different classes

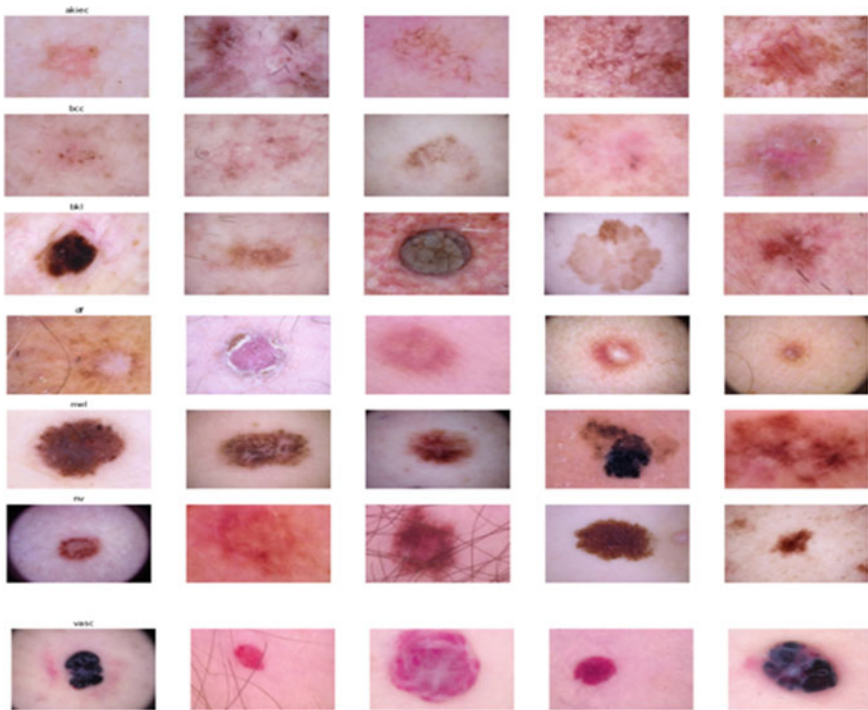


Fig. 2 Skin cancer data set HAM10000

of skin cancer: (a) Melanocytic nevi (nv), (b). Melanoma (mel), (c) Benign keratosis-like lesions (bkl), (d) Basal cell carcinoma (bcc), (e) Actinic keratoses (akiec), (f). Vascular lesions (vasc), (g). Dermatofibroma (df) [3] (Fig. 2).

In MNIST HAM10000 datasets, all dermoscopic images are have their own classes and features. The classes and features were stored in a formatted file Comma Separated Values (.csv). The MNIST HAM10000 datasets has seven data features with their own names such as lesion-id, image-id, dx, dx-type, age, sex, and localization. The lesion-id feature is an ID that characterizes the patient's identity while image-id is a numbering ID that characterizes dermoscopic images [7] (Fig. 3).

4 Description of Algorithm

4.1 Convolutional Neural Network (CNN) Architecture

The convolutional neural network (CNN) is one kind of deep neural architecture which has achieved great performance capacity in tasks like image recognition. Different form of traditional neural networks like back-propagation, fully connected

Dataset	License	Total images	Pathologic verification (%)	akiec	bcc	mel	nv	vasc
PH2	Research&Education*	200	20.5%	-	-	-	40	160
Atlas	No license	1024	unknown	5	42	70	20	275
ISIC 2017 ^b	CC-0	13786	26.3%	2	33	575	7	1019
Rosendahl	CC BY-NC 4.0	2259	100%	295	296	490	30	342
ViDIR Legacy	CC BY-NC 4.0	439	100%	0	5	10	4	67
ViDIR Current	CC BY-NC 4.0	3363	77.1%	32	211	475	51	680
ViDIR MoleMax	CC BY-NC 4.0	3954	1.2%	0	2	124	30	24
HAM10000	CC BY-NC 4.0	10015	53.3%	327	514	1099	115	1113

Fig. 3 Summary of some publicly available datasets of skin cancer

neural network, which contains input layer, hidden layers, and output layer, with and Max pooling, dropout in hidden layers.

Convolutional layers contain weight matrix called kernel and the size of the kernel is 3×3 and different features were mapped through this kernel. Kernels perform convolutional operation with every feature maps in previous layer $L1$, then we sum them up and output is the pixel value in layer $L + 1$. There are three convolutional layer and in each layer is decreased exponentially with the number of hidden nodes and layers. When a feature has been detected, its exact location becomes less important. Max pooling layers performs sub-sampling operation on different feature maps. For every four pixels, we only retain the max value, so that the size of feature maps will be half of the original size. Sub-sampling reduces the resolution of feature maps and reduces the sensitivity of the output to shifts and distortions, so that the model will be more robust. Max pooling operation can be incorporated into convolutional layers, and we do not need additional layers to do sub-sampling. The deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters, i.e., kernels, pooling, fully connected layers (FC) and apply softmax function to classify an object with probabilistic values between 0 and 1. Figure 4 is a complete flow of CNN to process an input image and classifies the objects based on values.

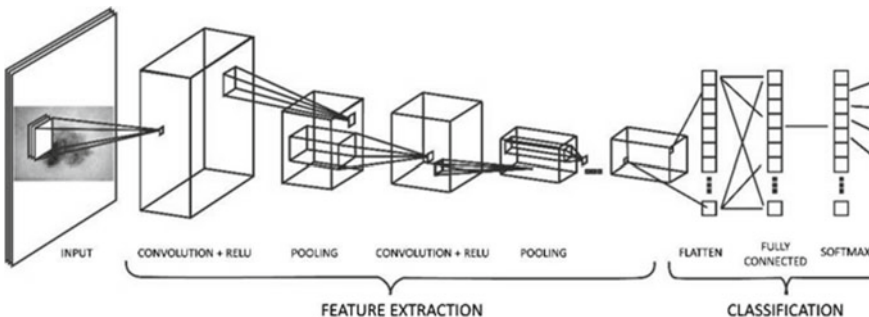


Fig. 4 CNN model

5 Methodology

We have followed for model building and evaluation which are as follows (Fig. 5):

Step I: Importing Essential Libraries

Step II: Data Preprocessing and analysis

Step III: Loading and Resizing of images

Step IV: Train Test Split of datasets

Step V: Apply Normalization

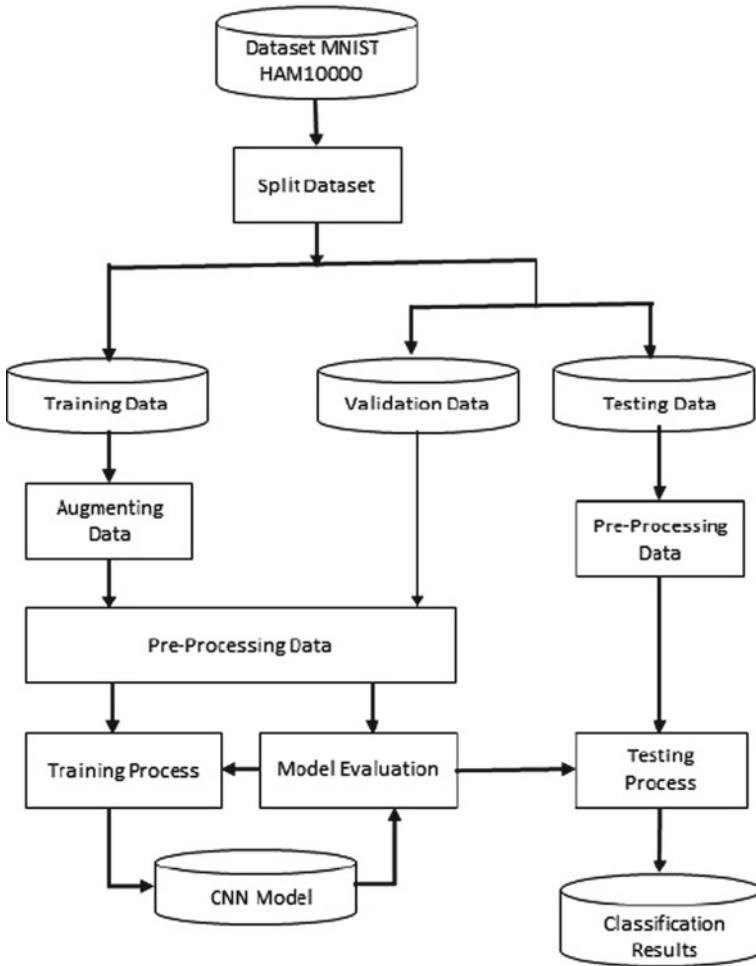


Fig. 5 Proposed model implementation

Step VI: Label Encoding process

Step VII: Train validation split of datasets

Step VIII: Model Building (CNN)

Step IX: Setting Optimizer

Step X: Fitting the model

Step XI: Model Evaluation

At first we have consider the process of training mode that need to determine the size of batch size, training steps and number of epochs. Batch size is the number of data that needs to trains that we wanted to send as input across the neural network in each iteration in training steps. In this training process we have considered the batch size 128 in general model building and 10 in model building after data augmentation. Train steps is the amounts of samples in that model that are the initial amount of data train before the data augmentation process we have considered. Epoch is a series of round of training steps needed to training a neural network by which we can found errors found to be close to zero. In this case we have considered 10 epochs on the training model process and 50 times epochs in data augmentation, respectively. Input size of the images are $75 \times 100 \times 256$ and kernel size is 3×3 . Here we have used three convolutional layer with max pooling, 30% dropout in each layer and in final stage we have used the flatten and two fully connected neural later with softmax function which produced seven different level of classification with probabilistic values.

6 Results and Model Evaluation

From Fig. 6 it is clear that in the model we have implemented there are over-fitting of data and hence data augmentation required. Data augmentation is the process to add data so that the number of data to be processed is increasing the performance parameter. After data augmentation we got the following results in Fig. 7 and Table 1.

From the previous process, it is obtained that the number of dermoscopic image for the training data is 8000 images and the number of images for validation data and test data are 1000 images. In the data augmentation process, new data are only added to the data training. The new data are generated by modifying spatial properties of the images such as horizontal or vertical flips, rotations, variations in the brightness of images, horizontal or vertical shifts, and zoom to existing data. Figure 8 is showing that after data augmentation we get better accuracy and loss of the model are very less. Figure 9 is showing the incorrect prediction of different types of skin cancer type through this proposed 2-D CNN model.

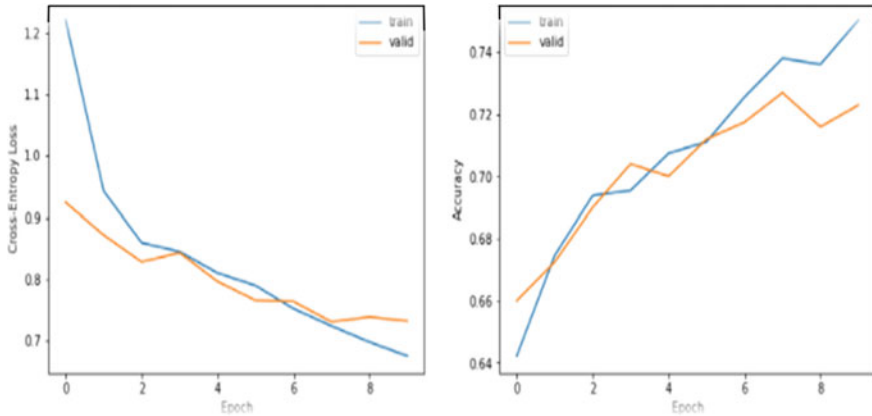


Fig. 6 Model evaluation for cross entropy loss and accuracy for convolution 2-D neural network

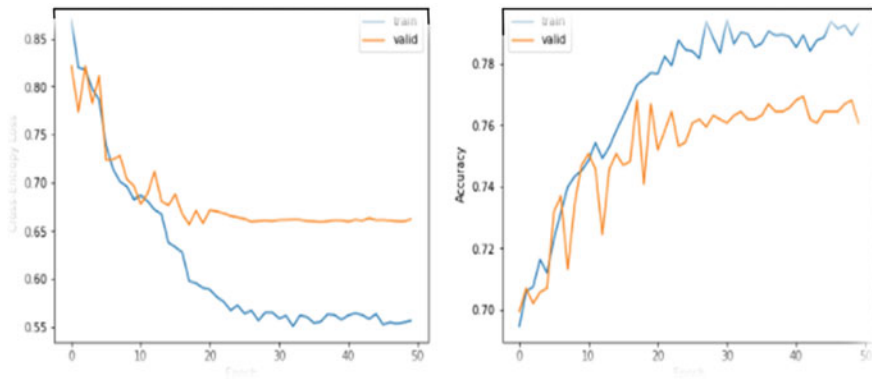


Fig. 7 Model evaluation after data augmentation for cross entropy loss and accuracy for convolution 2-D neural network

Table 1 Train–Test results for the proposed model

Method	Train-Acc	Train-Loss	Test-Acc	Test-Loss	Val-Acc	Val-Loss
General 2-D CNN	74.46	68.52	73.21	72.29	72.29	73.21
General 2-D CNN after data augmentation	79.32	55.81	88.39	76.33	88.06	66.22

7 Future Scope and Conclusion

The proposed research methodology will study the performance on auto-encoder-decoder CNN architectures, another new approach Kohonen self-organizing neural networks (KNN), and most recently widely used generative adversarial neural

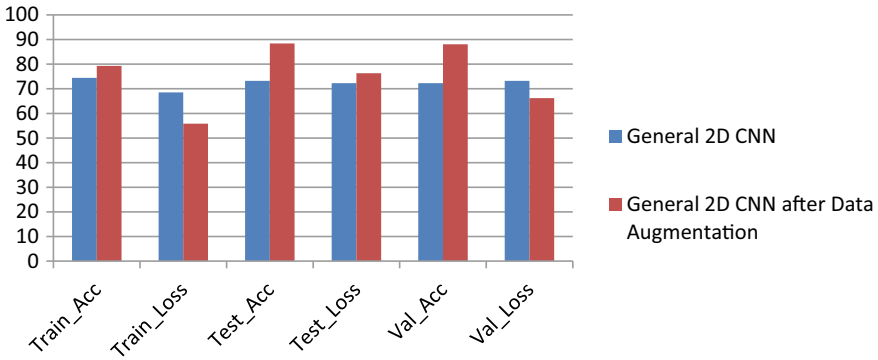
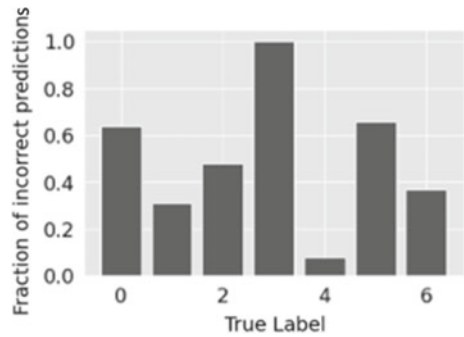


Fig. 8 Accuracy is increasing after data augmentation and loss is decreasing after data augmentation

Fig. 9 Incorrect prediction in the proposed model



networks (GAN) for skin cancer detection. The effective reviews have been done through various neural network techniques for skin cancer detections and classification techniques in research review. All of these approaches are painless and fast which provide results more accuracy. Most of the research works were related to skin cancer detection focuses on whether a given lesions images are cancerous based on some previous trained image cancer trained model. However, when a patient is diagnosed with different symptoms of skin cancer appears on any part of human body then the all the machine learning oriented research cannot provide a suitable answer. Next proper and suitable feature extraction algorithms are required and some image balancing techniques, batch normalization are also important to get the better model. Hence, the proposed research methodology will study the performance on different CNN architecture for skin cancer detection.

By applying different Optimizer like SGD, RMSprop, and Nadam optimizer in the future work in data augmentation.

References

1. Gaana M, Gupta S, Ramaiah NS (2019) Diagnosis of skin cancer melanoma using machine learning. Available at SSRN: <https://ssrn.com/abstract=3358134> or <https://doi.org/10.2139/ssrn.3358134>
2. Naeem A, Farooq MS, Khelifi A, Abid A (2020) Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities. *IEEE Access* 8:110575–110597. <https://doi.org/10.1109/ACCESS.2020.3001507>
3. Purnama KE et al (2019) Disease classification based on dermoscopic skin images using convolutional neural network in tele dermatology system. In: 2019 International conference on computer engineering, network, and intelligent multimedia (CENIM), 2019, pp 1–5, <https://doi.org/10.1109/CENIM48368.2019.8973303>
4. Alasadi A, ALSafy BM (2015) Early detection and classification of melanoma skin cancer. *Int J Inf Technol Comput Sci* 7(12):67–74. <https://doi.org/10.5815/ijitcs.2015.12.08>
5. Esteva A, Kuprel B, Novoa R et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118. <https://doi.org/10.1038/nature21056>
6. Kadampur MA, Al Riyae S. Skin cancer detection: applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. <https://doi.org/10.1016/j.imu.2019.100282>
7. Tschandl P (2018) The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. <https://doi.org/10.7910/DVN/DBW86T>, Harvard Dataverse, V3, UNF:6:/APKSsDGVDhwPBWzsStU5A== [fileUNF]

Segmentation and Classification of Skin Cancer Using K-means Clustering and EfficientNetB0 Model



Vatsala Anand, Sheifali Gupta, Deepika Koundal, Soumya Ranjan Nayak, Jana Shafi, and Akash Kumar Bhoi

Abstract Nowadays, skin cancer is the widely recognized cancer all over the world. As the spreading rate of skin cancer is increasing day by day, so, there is a need to develop a technique that can detect skin cancer at an early stage. These days, deep learning has attained outstanding success for the detection and diagnosis of cancers. In this paper, a transfer learning-based EfficientNetB0 model is improved by adding one average pooling layer, one dropout layer, one batch normalization and one dense layer with softmax activation function. The proposed model has been simulated using the Kaggle database. The training and calculation are done with different hyper parameters such as batch size, optimizer and epochs. The data augmentation technique is applied to solve the problem of less amount of images. The proposed model has attained 87% accuracy on Adam optimizer with 32 batch size and 30 epochs.

V. Anand · S. Gupta

Institute of Engineering and Technology, Chitkara University, Chitkara University, Rajpura, Punjab, India

e-mail: vatsala.anand@chitkara.edu.in

S. Gupta

e-mail: sheifali.gupta@chitkara.edu.in

D. Koundal

Department of Systemics, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India

e-mail: dkoundal@ddn.upes.ac.in

S. R. Nayak (✉)

Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India

e-mail: nayak.soumya17@gmail.com

J. Shafi

Department of Computer Science, College of Arts and Science, Prince Sattam Bin Abdul Aziz University, Wadi Ad-Dwasir 11991, Saudi Arabia

e-mail: j.jana@psau.edu.sa

A. K. Bhoi

KIET Group of Institutions, Delhi-NCR, Ghaziabad 201206, India

Directorate of Research, Sikkim Manipal University, 737102 Gangtok, Sikkim, India

Keywords Classification · K-means · Benign · Malignant · Data augmentation · Clustering · Skin cancer · Kaggle · EfficientNetB0

1 Introduction

Skin is the outer covering of body that protects human body from direct radiations of sun. Skin always acts as a safeguard against all the harmful radiations. But the spreading rate of skin cancer is increasing day by day. Skin cancer is caused because of the abnormal growth of cells in skin. Benign and malignant are the two types of cancer [1, 2]. Benign nodules or lesions are non-cancerous and does not spread, whereas malignant lesions are harmful and can spread to other parts [3–5]. So, there is a need to develop a technique that can detect skin cancer at an early stage [6]. Different researchers had worked and presented techniques to overcome the issues, but it lags in case of accuracy. Also, most of the time there are issues like inter-class and intra-class variations. In the recent studies, deep learning models are used for the identification of skin cancer. Garnavi et al. [7] had used an algorithm using 30 high resolution dermoscopy images for segmentation. Lu et al. [8] had proposed a novel segmentation method for skin histopathological images. The dataset consists of a total of 28 skin tissue and had shown sensitivity rate of 80% and positive prediction rate of 70%. Antari et al. [9] had used HAM10000 dataset and worked on different disease classes using transfer learning-based models and had achieved accuracy of approximately 85%. Hekler et al. [10] had used ResNet50 model on ISIC archive with 75.03% accuracy. Mahbod et al. [11] used fusion approach on ISIC archive and had obtained an accuracy value of 86.2%. Masni [12] had used different transfer learning models on ISIC 2017 dataset and had achieved an accuracy of approximately 81%. The major contributions of the study are as follows:

1. Initially, segmentation has been performed on skin dermoscopy images using K-means clustering for region-of-interest (ROI) extraction. Then, classification is performed on segmented images using modified EfficientNetB0 model.
2. The EfficientNetB0 model has been improved by adding one average pooling layer, one dropout layer, one batch normalization and one dense layer with softmax activation function.
3. To increase the number of images in dataset, data augmentation technique has been applied.
4. The proposed model has been analyzed with Adam optimizer on 32 batch size and 30 epochs.

The rest of the paper is as follows. Section 2 shows the proposed framework, followed by results and discussion in Sect. 3. Conclusion is given in Sect. 4.

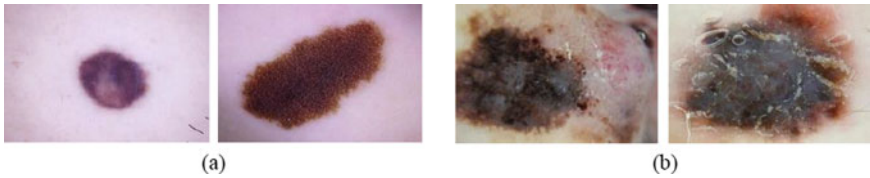


Fig. 1 **a** Benign skin cancer and **b** malignant skin cancer

Table 1 Dataset description

	Total	Training	Test	Validation
Benign	1800	1534	186	80
Malignant	1497	1284	144	69
Total	3297	2818	330	149

2 Proposed Framework

In this section, proposed model is explained for the classification of skin disease.

2.1 Dataset

For the evaluation and validation of the proposed model, a set of 3297 skin cancer images are used that are collected from the Kaggle database [13]. This dataset consists of two skin cancer classes namely benign and malignant with 1800 and 1497 images, respectively. Figure 1 shows the original images of benign and malignant class.

Table 1 gives description about the dataset. It shows that the number of total skin cancer images in the dataset is 3297 in count from which 1800 images are of benign class and 1497 images are of malignant class. The splitting of the dataset is done in such a way that approximately 10% of benign and malignant images are used for the testing purpose. From the remaining count of images, 5% of the skin cancer images are used for validation purpose. Rest dataset is used for training the model.

2.2 ROI Extraction Using K-means Clustering

In this section, jpg image is transferred into an array image. After that image is resized, segmented and append. Figure 2 shows the images after ROI extraction using K-means clustering. This is an unsupervised approach that depends upon the value of k [14, 15]. It allocates data points to clusters as shown in Eq. 1. The working steps are as follows:

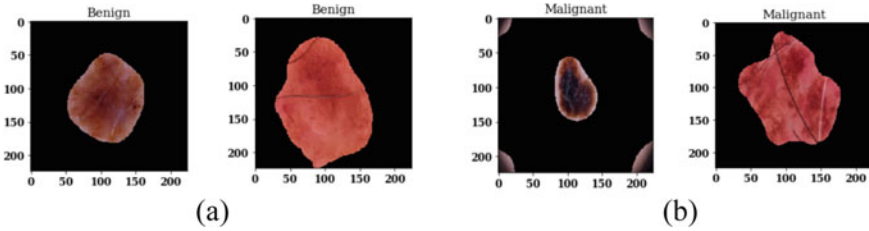


Fig. 2 ROI extraction **a** benign image after ROI extraction and **b** malignant image after ROI extraction

- Step 1: Specify the value of K .
 - Step 2: Initializing the centroids with dataset shuffling.
 - Step 3: Randomly select the K data points.
 - Step 4: Continue the iterating until and unless there is no change in the centroids.
 - Step 5: Analyze the sum of squared distance between the centroid and data points.
 - Step 6: Assign each data point to the closest cluster.
 - Step 7: Calculate the cluster centroids by taking the average of all the data points.
- In this paper, the value of K is two.

$$J = \sum_{i=1}^m \sum_{k=1}^K w_{ik} \|x^i - \mu_k\|^2 \tag{1}$$

where $w_{ik} = 1$ for data point x^i if it belongs to cluster k ; otherwise, $w_{ik} = 0$. Also, μ_k is the centroid of x^i 's cluster.

2.3 Augmentation of Image

In the deep learning, a large quantity of dataset is needed for best accuracy of the model. The Kaggle database consists of a less number of images. Therefore, to solve this issue, data augmentation techniques are applied [16]. The augmentation includes rotation range of 360 degrees, horizontal flip, vertical flip, zooming at a level of 0.75 followed by brightness range of 0.75–1.25. Figure 3 shows the segmented and augmented images after applying augmentation techniques in sequential order. Figure 3a–c shows the segmented image 1, 2 and 3, respectively, whereas Fig. 3d–f shows the augmented image 1, 2 and 3, respectively.

Table 2 shows the description of benign and malignant skin cancer class after augmentation. After augmentation the total images increased to 5636.

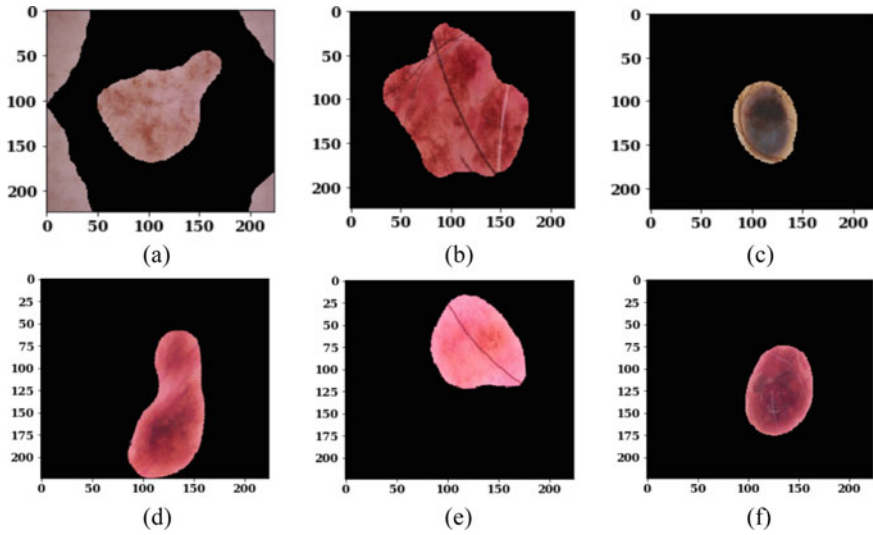


Fig. 3 Images after segmentation and augmentation. **a** Segmented Image 1, **b** segmented Image 2, **c** segmented Image 3, **d** augmented Image 1, **e** augmented Image 2 and **f** augmented Image 3

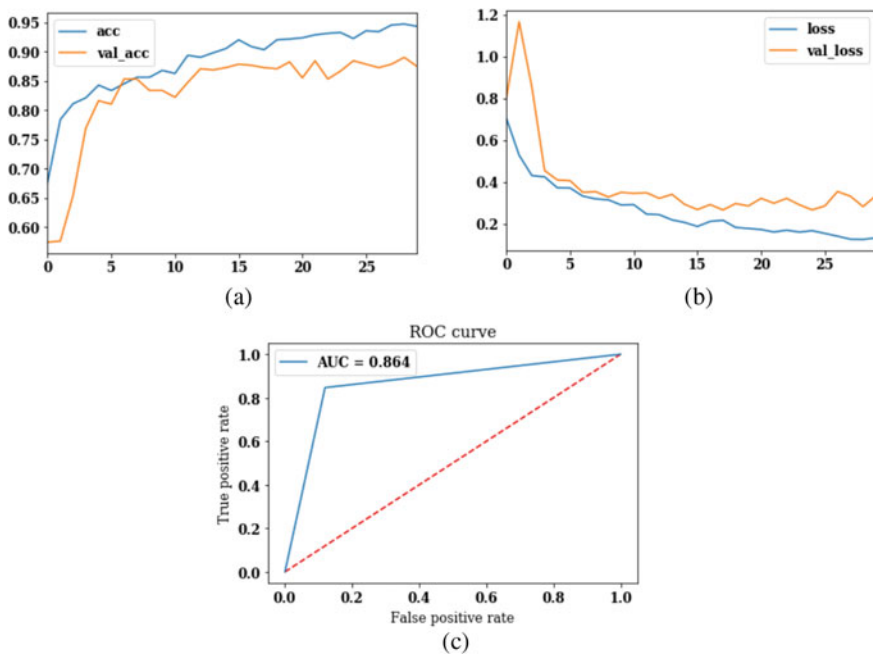


Fig. 4 Analysis of training and validation accuracy and loss. **a** Training and validation accuracy, **b** training and validation loss and **c** ROC curve

Table 2 Dataset description after augmentation

	Training images before augmentation	Training images after augmentation
Benign	1534	3068
Malignant	1284	2568
Total	2818	5636

2.4 Fine-Tuned EfficientNetB0 Model

In this section, fine-tuned transfer learning-based EfficientNetB0 [17] model is used. EfficientNetB0 model is a convolutional neural network that is trained on millions of images from the ImageNet dataset. This model architecture can easily classify 1000 categories of images. The authors had developed a full family of EfficientNets for B1-7 and the baseline model is B0. The working of the proposed model is done by adding one average pooling layer, one dropout layer, one batch normalization and one dense layer with softmax activation function. Table 3 shows the parameters of the proposed model. The EfficientNetB0 model shows a total of 4,049,564 parameters whereas the total number of parameters obtained after fine tuning of the model are 4,057,246. A total number of 4,012,670 trainable parameters are obtained and the total number of 44,576 non-trainable parameters are obtained.

3 Results and Discussion

This section is comprised of all the results obtained by using proposed model. The proposed model is simulated using Kaggle database on Adam optimizer and 32 batch size with 30 epochs. Training and validation loss and accuracy curves are attained, and the experimental analysis is carried out on different performance parameters such as precision, sensitivity, specificity and accuracy.

Table 3 Parameters of the proposed model

Layer (Type)	Output shape	Parameters
EfficientNetB0	7, 7, 1280	4,049,564
Global average pooling	1280	0
Dropout (0.5)	1280	0
Batch normalization	1280	512
Dense	2	256
Total parameters		4,057,246
Trainable parameters		4,012,670
Non-trainable parameters		44,576

Table 4 Hyper tuning Parameters

S. No.	Parameter name	Value
1	Batch size	32
2	Optimizer	Adam
3	Epochs	30

3.1 Hyper Parameter Tuning

The hyper parameter tuning is performed using different parameters such as epochs, optimizer and batch size. The analysis is performed using Adam optimizer on 30 epochs with 32 batch size. Table 4 shows the hyper tuning parameter values.

3.2 Analysis of Loss and Accuracy on Training and Validation

This section shows the values of loss and accuracy. Table 5 shows the analysis using 30 epochs, and results are shown on epoch value 1, 10, 20 and 30. It can be seen that as the epoch value is increasing the value of validation accuracy is increasing and the training loss value is decreasing. On the 30th epoch, the value of validation accuracy is 87.50%, training accuracy is 94.37% and the value of training loss is 0.1328.

Figure 4 shows the accuracy and loss curves. Figure 4a shows the curve of training accuracy and validation accuracy. The value of validation accuracy is approximately 87% on the 30th epoch value. Figure 4b shows the curve of training loss and validation loss. The maximum validation loss value is greater than one. Figure 4c shows the AUC curve of the proposed model. The AUC curve is the measure for distinguishing between different classes. The value of AUC is 0.86 as shown in figure.

Table 5 Analysis on training and validation

Epoch	Training loss	Training accuracy (%)	Validation loss	Validation accuracy (%)
1	0.7047	67.69	0.8016	57.42
...
10	0.2909	86.83	0.3509	83.40
...
20	0.1785	92.20	0.2860	86.28
...
30	0.1328	94.37	0.3345	87.50

Fig. 5 Confusion matrix

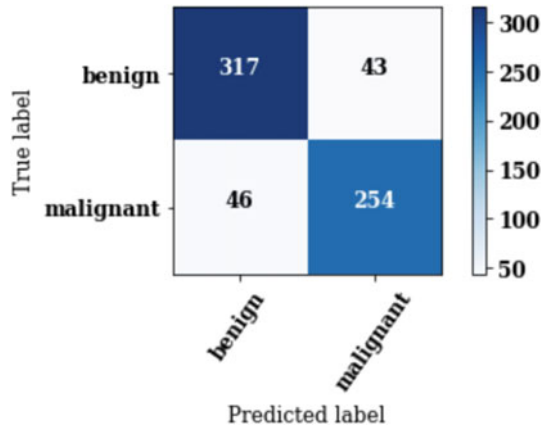


Table 6 Analysis of confusion matrix parameters

	Precision (%)	Sensitivity (%)	Accuracy (%)
Benign	87	88	88
Malignant	86	85	85
Macro average	86	86	86
Weighted average	87	87	87

3.3 Analysis of Confusion Matrix Parameters

Confusion matrix shows the clear view of incorrect and correct predictions. Figure 5 gives the confusion matrix for proposed EfficientNetB0 model. It gives the true and predicted labels. Here, true labels are shown vertically and predicted labels are shown horizontally.

Confusion matrix parameter analysis is done by attaining the values for precision, sensitivity, specificity and accuracy. Table 6 shows the values obtained on benign, malignant, macro average and weighted average. In case of benign class, the value of accuracy is 88%, whereas for malignant class the value of accuracy is 85%.

3.4 Classification Results

Figure 6 shows the classification results of images using Adam optimizer on 32 batch size and 30 epochs. Figure 6a shows the predicted and actual results of benign class, whereas Fig. 6b shows the predicted and actual results of malignant class. It is observed that the predicted class is same as the actual class.

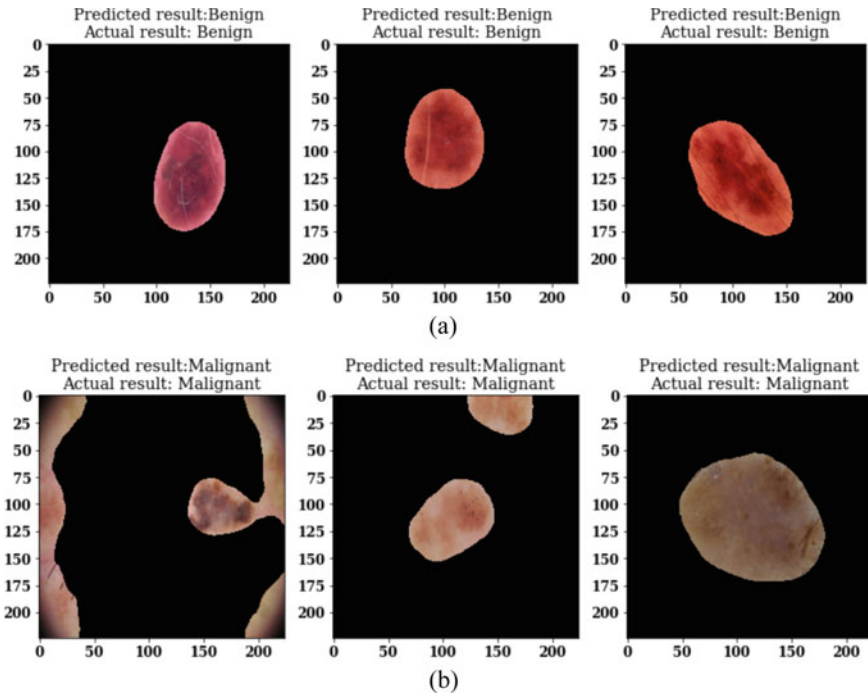


Fig. 6 Classification results

Figure 7 shows the misclassification results. Figure 7a shows the predicted result of the image as benign but in actual the image is malignant, whereas Fig. 7b shows the predicted result of the image as malignant but in actual the image is benign.

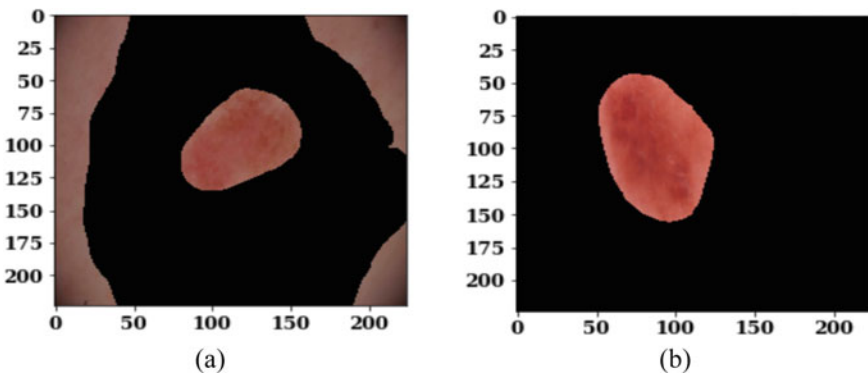


Fig. 7 Misclassification results

4 Conclusion

In this paper, a transfer learning-based EfficientNetB0 model is improved by adding one average pooling layer, one dropout layer, one batch normalization and one dense layer with softmax activation function. The proposed model has been simulated using the Kaggle database. The training and evaluation of the proposed model is done using different hyper parameters such as batch size, optimizer and epochs. The data augmentation technique is applied to increase the count of images. The proposed model has attained 87% accuracy on Adam optimizer with 32 batch size and 30 epochs.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries'. CA: Cancer J Clin 68(6):394–424
2. Anand V, Gupta S, Koundal D (2022) Skin disease diagnosis: challenges and opportunities. In: Proceedings of second doctoral symposium on computational intelligence. advances in intelligent systems and computing, vol 1374. Springer, Singapore
3. Bauer A, Adam KE, Soyer PH, Adam KWJ (2020) Prevention of occupational skin cancer. In: Kanerva's occupational dermatology, pp 1685–1697
4. Silveira M, Nascimento JC, Marques JS, Marçal AR, Mendonça T et al (2009) Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. IEEE J Sel Top Signal Process 3(1):35–45
5. Celebi ME, Iyatomi H, Schaefer G, Stoecker WV (2009) Lesion border detection in dermoscopy images. Comput Med Imaging Graph 33(2):148–153
6. Emre Celebi M, Wen Q, Hwang S, HItomi, Schaefer G (2013) Lesion border detection in dermoscopy images using ensembles of thresholding methods. Skin Res Technol 19(1):252–258
7. Garnavi R, Aldeen M, Celebi ME, Bhuiyan A, Dolianitis C, Varigos G (2010) Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels. Int J Med Med Sci 1(2):126–134
8. Lu C, Mahmood M, Jha N, Mandal M (2013) Automated segmentation of the melanocytes in skin histopathological images. IEEE J Biomed Health Inform 17(2):284–296
9. Al-antari MA, Rivera P, Al-masni MA, Valarezo E, Gi G, Kim TY, Park HM, Kim TS (2018) An automatic recognition of multi-class skin lesions via deep learning convolutional neural networks. In: Conference: ISIC2018: skin image analysis workshop and challenge 2018
10. Hekler A, Kather JN, Krieghoff-Henning E, Utikal JS, Meier F, Gellrich FF, Belzen J, French L, Schlager JG, Ghoreschi K, Wilhelm T (2020) Effects of Label noise on deep learning-based skin cancer classification. Front Med 7(1):1–7
11. Mahbod A, Schaefer G, Wang C, Dorffner G, Ecker R, Ellinger I (2020) Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. Comput Methods Programs Biomed 193(1):1–9
12. Masni MA, Kim DH, Kim TS (2020) Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. Comput Methods Programs Biomed 190(1):1–12
13. Kaggle database. <https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign>
14. Likas A, Vlassis N, Verbeek JJ (2003) The global k-means clustering algorithm. Pattern Recogn 36(2):451–461

15. Chowdhury K, Chaudhuri D, Pal AK (2021) An entropy-based initialization method of K-means clustering on the optimal number of clusters. *Neural Comput Appl* 33(12):6965–6982
16. Connor S, Khoshgoftaar T (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1):1–48
17. Hoang V-T, Jo K-H (2021) Practical analysis on architecture of EfficientNet. In: 2021 14th international conference on human system interaction (HSI) IEEE, pp 1–4

Comparative Assessment of Performances of Various Machine Learning Algorithms in Detection of Liver Ailments



Dwaipayan Saha, Indrani Mukherjee, Jesmin Roy, and Pranati Rakshit

Abstract Liver is a crucial interior organ of the human body whose primary tasks are to eliminate waste which is produced by the organs. In human body, most of the difficult tasks are being performed by the liver, and any abnormal activities happening due to malfunctioning of the liver will result into life-taking reason. Disease prediction in the human being has always been a long tenure procedure in early days. As the days have passed, computer-based diagnosis has become an important role in the medical world for prognosis, analyzing, and storing medical information with their related images. The liver disorder can cause various fatal and life-taking diseases, which also includes liver cancer. Early diagnosis and treating the patients beforehand can be helpful to reduce the risk of those lethal and fatal diseases. As the diagnosis of liver disease is quite expensive and sophisticated, numerous research have been performed using machine learning (ML) methods for classifying liver disorder cases. In this present work, we have categorized the liver patients on the basis of liver patient dataset using various machine learning techniques and approaches. In our work, we have classified our dataset with different classifiers such as k-nearest neighbor, random forest, support vector machine, and Extra Trees. After analysis, we concluded that after comparing all the accuracies, we found that Extra Trees classifiers gave the highest accuracy compared to other classifiers. We got the highest accuracy of 91.67%.

Keywords Liver diseases · Classification · Data preprocessing · Accuracy · Random forest

1 Introduction

Liver is an essential and the largest gland of our human body. It carries out most of the vital and essential tasks. The liver's main function is to purify and filter the blood which comes from the digestive tract, before passing it to the rest of the areas [1].

D. Saha · I. Mukherjee · J. Roy · P. Rakshit (✉)
Department of Computer Science and Engineering, JIS College of Engineering, Kalyani, India
e-mail: pranati.rakshit@jiscollege.ac.in

The liver clarifies the chemicals and metabolizes drugs, while detoxifying chemicals, the liver produces bile juice that gets accumulated back in the intestines. The liver also produces proteins which are important for blood clotting mechanism and other functions [1].

Any disorder related to liver functions is considered as liver disease. Long-term damage to the liver from any cause can lead to permanent blemish, called liver cirrhosis; thus, the liver then is unfit to function properly. After cirrhosis, there are high chances of occurring liver cancer, hepatocellular carcinoma which are life-taking diseases. As cirrhosis results in the leakage of fluids from the liver into the belly, the fluids become distended and heavy and causes problem. There are many factors such as bilirubin, albumin, alkaline phosphatase, aspartate aminotransferase, and many more for which such ailments occur [1, 2].

In the preliminary stage of the liver problem, diagnosis will enlarge the survival rate of the patient. Classifications main ambition is to do predict upon the training data and the test data and analyze the dataset accordingly. So, we have taken a dataset which contain the details of 538 liver patients. We have used various classifiers on the dataset and have analyzed our results.

2 Related Work

El-Shafeiy et al. [1] tested and trained the dataset, out of 7000 patients details they worked with twenty-three attributes. In their proposed methodology, feature selection technique with C5.0, SVM, and NB are employed as an ensembled model. They stated that their proposed technique resulted higher in accuracy and sensitivity while comparing with the result of applying C5.0, SVM, and NB on the dataset single-handedly.

Singaravelu et al. [2] preprocessed a dataset in their research paper of liver disease with the eleven parameters. Discretization of data was performed for all the columns to group the values based on the similarities and algorithms such as Naive Bayes, K-Star, random tree, and J48 are selected for implementation. The confusion matrix of all the algorithms was then analyzed, and the analysis showed that random tree algorithm gave the highest accuracy of 74.2% among them, and the execution time was the highest 0.05 secs.

Sazzadur Rahman et al. [3] presented a comparative report of implementing six different machine learning algorithms logistic regression, random forest, DT, support vector machine, KNN, and Naïve Bayes on the 583 patients dataset that includes 75.64% of male patients and 24.36% of female patients, and it contained 11 particular parameters, whereas 10 parameters are features for analysis and 1 parameter is the target class. The performance of different classification techniques was evaluated on different areas such as precision, f-1 score, and specificity, and after analyzing the report showed that logistic regression achieved the highest accuracy.

Gulia et al. [4] involved classification in their proposed work, with the usage of Bayesian Network, SVM, J48, multi-layer perceptron, and RF. Hybrid model

construction and comparative analysis were implemented for improving the accuracy rate in three phases. Different classifiers were executed on the original dataset collected from UCI repository. Once the three-phase analysis is done, it was concluded the RF is the best one with an accuracy rate of 71.87%.

Kumar et al. [5] highlighted the rule-based classification model. It was notified that the efficiency of all the algorithms decreased without the usage of rule classification technique. Their proposed work idea indicated that out of all the rules 20 rules were selected and were used for the classification of the diseases. DT gave the best performance using rule-based classification, and the accuracy rate decreased when rule-based was not implemented.

Singh et al. [6] predicted liver disease by using software engineering approach. The authors in their research work have used various classification techniques such as LR, SMO, random forest, Naive Bayes, J48, and KNN to find the optimum best outcome. The results of the proposed work were investigated into two halves, one using feature selection technique and the other one without using feature selection techniques, and it was concluded that the best result was achieved using logistic regression with incorporation of feature selection techniques, and execution time of different classifiers are developed and it is found lower after the implementation of feature selection.

3 Methodology

3.1 Data Preprocessing

Data preprocessing [7] is primarily a vital step toward the classification process. It is used to clean up the data and make it in readable and understandable form. Data preprocessing helps in extraction of important insights from the raw data. Data preprocessing involves four important steps:

- Data cleaning—it includes cleaning of the raw data and removing all unnecessary data to make it to usable form.
- Data integration—after clean up, the data that are scattered all over are integrated or combined to again make it to usable format.
- Data reduction—after the integration process, data reduction is carried out. In this phase, the amount of total data is reduced to a particular amount which is actually required. Also, it eliminates all redundant data from the dataset.
- Data transformation—after the data is reduced to its required size, it then undergoes data transformation. In this, the format of the data, the structure or the values of data are changed accordingly.

We have worked with a dataset in which the health observations of approximately 600 patients have been recorded. Several attributes like Gender, Age, Aspartate_Aminotransferase, Total_Protiens, Albumin, Alkaline_Phosphotase, Albumin

Alamine_Aminotransferase, Total_Bilirubin, and Albumin_and_Globulin_Ratio [7] are observed hence recorded. Data preprocessing is an essential and primary task which requires heed and has to be executed before we do the training and testing operations on the data. This helps to reduce errors and minimize them.

In the present work, we have initially checked whether there are any null values or not. Then, we checked if any of the rows had missing values or not. On checking for the same, we found that there were some null values [8]. We found that some of the values were missing for Albumin_and_Globulin column. Hence, we proceeded further with data preprocessing.

Since Albumin_and_Globulin column [8] had null values, we could not neglect them without knowing whether that would impact the result of our work or not. Hence, we required to check the correlation of Albumin_and_Globulin_Ratio with other columns. Thus, we created a heatmap.

From the above heatmap (Fig. 1), it was observed that Albumin_and_Globulin_Ratio is highly correlated with albumin. Hence, we applied binning to albumin and have filled the values in Albumin_and_Globulin_Ratio using median of the bin value.

Then, we plotted some of the statistics from the dataset. For example, we plotted the statistics of the number of male and female recorded to investigate whether they have liver disease or not [8, 9]. Next, we plotted the number of people in total who

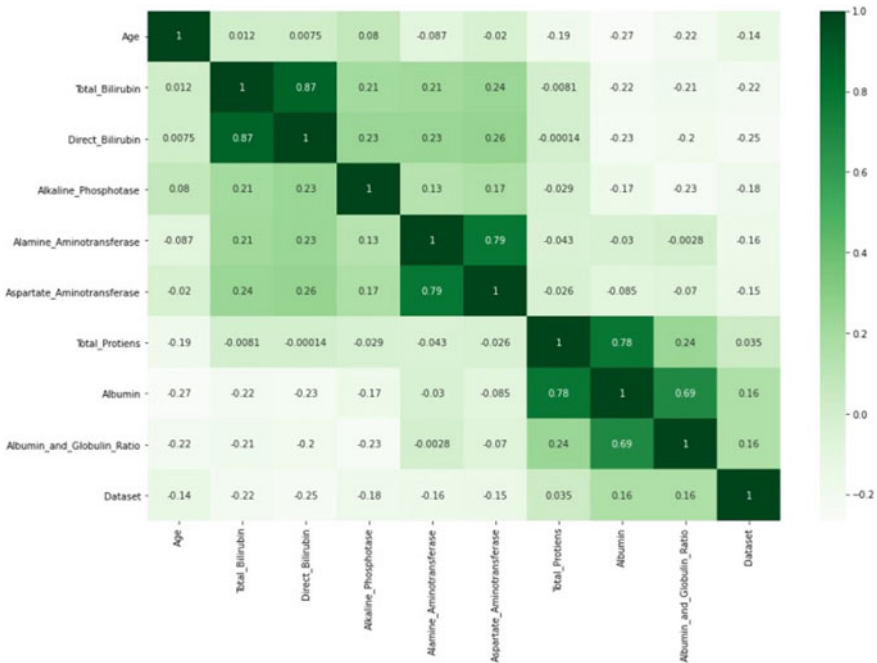


Fig. 1 Heatmap to check the co-relation

belonged to respective classes of having liver ailment and not having liver ailment. We have shown the results obtained in later section of this work. After seeing the result, we found that there was an imbalance and less than 30% data belonged to the class 2. So, first, we had to balance the data so that we could obtain more precise predictions. For this, we have used over sampling.

3.2 Classification

Several techniques which are used classically for classification have been used in the process of classification. The results obtained from the different classifiers have been compared keeping the accuracies as the criteria.

(a) K-Nearest Neighbor K-nearest neighbor (KNN) [9, 10] is usually considered as some of the simplest and the easiest ML algorithms which is based upon supervised learning technique. The KNN algorithm considers the similar features between the given case data and available case data and will give the verdict in favor of the case category which is the most similar in comparison to the available ones.

KNN works by calculation of the distance, which is Euclidean, of a variable number of neighbors which is considered as K, and then, we consider the k-nearest neighbors after calculating the Euclidean distance. Among all the k-neighbors, counting is done, and the data are given to the category after comparing where the maximum number of neighbors have been calculated. Hence, the model is ready to give the accuracy. This classifier is simple to implement and proves effective for large dataset but always need to estimate the correct value of K can be considered as a complex work.

(b) Naive Bayes Naive Bayes [9–11] classifiers are a category of classification algorithm which is solely based on the Bayes' theorem. All the algorithms of this group of classification share a common principle.

Bayes' theorem is a theorem which tells us the way of finding a probability when some certain are known. Bayes theorem is an approach to figure out the conditional probability. Conditional probability can be defined as the occurrence of an event happening, keeping the condition in to consideration that it has some relationship to one or more other events.

Naive Bayes classifier gives more accurate results when it is used for textual data analysis.

(c) Random Forest It is one of a kind of supervised ML algorithm. The forest that the algorithm builds, it is an ensembling dealing with decision trees and is generally an application of bagging method. One of the biggest advantages of random forest classifier is that the algorithm can be implemented for classification as well as regression problems. With the help of random forest classifier, one can also deal with regression tasks by using the algorithm's regressor [9, 10]. Another biggest advantage of random forest classifier is its versatility nature. This classifier can be

used in a dual purpose as it serves for both regression and classification tasks. Random forest is also a very friendly algorithm because the default hyperparameters of these classifiers that is used often generates a good prediction result. One of the biggest problems in machine learning algorithm is its overfitting nature, but this overfitting problem can be solved by the use of random forest classifier.

(d) Extra Trees Classifier Extra Trees classifier [9–11] is a widely used ensemble method which is basically based on the extended use of decision trees. This classifier, like random forest, randomizes the obtained decisions and forms subsets of data which as a result minimizes overlearning of the data and save the classification from overfitting.

This algorithm basically creates several decision trees from the dataset which is given for the purpose of training. And the predictions hence are made accordingly by finding out the average of the prediction of the decision trees in the case of regression and uses the technique of majority voting for the purpose of classification. The optimum number of trees has to be found until the model performance is stabilized. It can be considered that more number of trees can lead to overfitting. Another hyperparameter for this abovementioned algorithm is the number of features used in the model of the ensemble.

(e) Linear Support Vector Machine SVM [9, 10] is a model which is linear and is useful for problems related to both regression and classification. This linear SVM model separates the data with hyperplane by drawing a straight line which is linear in nature, and thus, it is called a linear SVM. Earlier, SVM was used to solve only classification-based problem, but due to emerging technologies, now, SVM is also used to solve regression-based problems. The main objective of the abovementioned algorithm is to generate the best line so that the dimensional space can be split into different classes. Linear SVM proved to be highly efficient in dealing with huge large size datasets. It also provides solution for multiclass classification problems consisting of any number of classes. Linear SVM can work with any number of high-dimensional data in both sparse and dense format of dataset. As the training dataset is linearly separated, the region bounded by the two hyperplanes is known as margins, and linear SVM has two types of margins: soft margin and hard margin. Linear SVM does not perform very well when the dataset is more noisy.

The very foremost and initial part of our work involved the data preprocessing procedures, during which the raw data were first collected and were followed by the cleaning of the dataset in order to achieve a better accuracy rate. After the data were cleaned and preprocessed, the irrelevant parts of the dataset were omitted and removed on account of the fact that they have no significant impact on the target objective, and those data don't influence the result. After which the rest of the impactful data was selected and from the selected data, some of the data was transformed as required for the usage and need of the data.

3.3 Implementation

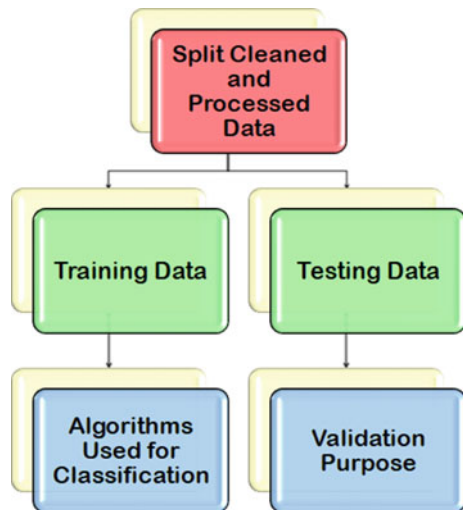
Successive to the cleaning and preprocessing which is shown in Fig. 2, the data were splitted into two halves. The first half comprised of 80% of the data, which was used for the purpose of training the models, and the rest half contained the 20% data for testing and validation purpose. We have used various multimodal approaches to classify the data. We have used a number of existing classifiers, namely KNN classifier, Naïve Bayes, random forest, linear SVC, and Extra Trees classifier. Workflow diagram is shown in Fig. 3.

After the completion of training and testing, we have found the accuracy obtained from each classifier along with respective confusion matrix. After this, we have made a comparison among the different classification techniques. And we have discussed about the results in the later section of this work.



Fig. 2 Data preprocessing

Fig. 3 Workflow diagram



4 Result and Discussion

Liver is a very crucial part of our digestive system. Damages caused to it due to various reasons can even cause death of a person [12]. So, we should take due care of liver to eradicate liver diseases and maintain a healthy lifestyle. In the project that has been done, the dataset consists of 584 records of people of various age groups. The age group is in range of 5–85 years. It also consists of various attributes which are very useful in determining whether a patient is having liver disease or not [13]. Some of the most important determining attributes are, gender, albumin to globulin ratio, total bilirubin, direct bilirubin, etc. In this project, we have taken various classifiers, like, KNN, Naïve Bayes, random forest, Extra Trees, and linear SVC [9–12]. In the above section, we had discussed about the results obtained after preprocessing the data[12, 13]. Here, below are some of the results that we got after data visualization and data preprocessing.

It is seen from the Fig. 4 that almost 324 male patients and 92 female patients are affected with liver disease; whereas, only, 117 male patients and 50 female patients are healthy or having non-liver diseases. Then, after studying the data, it is seen that the dataset is imbalanced. Since it consists of two classes, class 1 and class 2, but, less than 30% of data only belongs to class 2.

Thus, the data need to be balanced first to get more precise predictions. For this purpose, oversampling is done; and data preprocessing is carried out. After this, oversampling is done; it was seen clearly that all the classes got balanced.

It was inferred that Naïve Bayes has produced the lowest accuracy rate (71.42%) among all the used classifiers that have been implemented for this work. Linear SVC has shown better result than Naïve Bayes has produced and has given percentage of 73.81 where accuracy is concerned. KNN has performed better than the previous

Fig. 4 Statistics of male and female affected by liver ailments

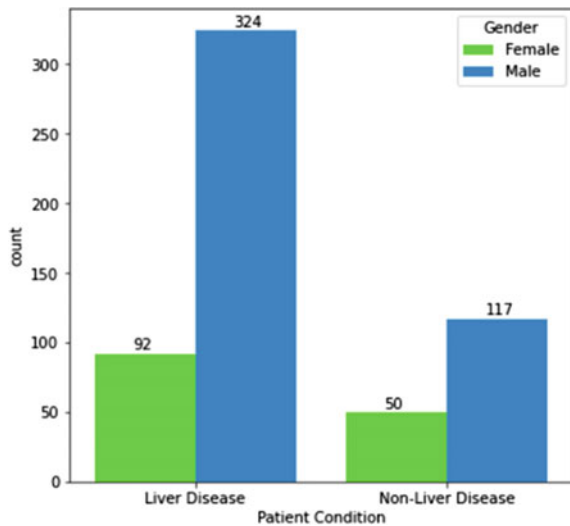


Table 1 Accuracies obtained with different classifiers

Classifier	Accuracy obtained (in %)
Naive Bayes	71.42
Linear SVC	73.81
KNN	80.95
Random forest	90.47
Extra Trees	91.67

one, and percentage is 80.95 where accuracy is the measure. Extra Trees classifier has given the highest accuracy of 91.67%, whereas the random forest has attained an accuracy of 90.47% (Table 1).

Various classifiers are used for creating the required model for detecting liver diseases. It was seen that many classifiers did not give very good output, whereas some gave expected output. There were variations in the output of various models based on their respective algorithms they use and also based on the type of data they use for their calculations [13]. After overall analysis, the accuracy the Extra Trees classifier is the highest. In general, majority of the classifiers have given pretty decent accuracies, but Naïve Bayes has the least accuracy among them.

5 Conclusion and Future Scope

Liver is an essential and the largest gland of our human body. It carries out most of the vital and essential tasks. Any disorder related to liver functions is considered as liver disease. Liver can get permanently damaged, and thus, the liver then becomes unable to function properly.

Classification is the methodology in which items that are variable are allotted to choose classes and categories. Its main aim is to do predictions based upon the training data and the test data. So, we have taken a data which contain the details of over 600 liver patients. We have used several classifiers on the dataset of liver patient and have analyzed our results.

Various classifiers (shown in Fig. 5) are used for creating the required model for detecting liver diseases. It was seen that many classifiers did not give very good output, whereas some gave expected output. There were variations in the output of various models based on their respective algorithms and also based on the type of data they use for their calculations. The Extra Trees classifier has given highest accuracy of 91.67% for this present work.

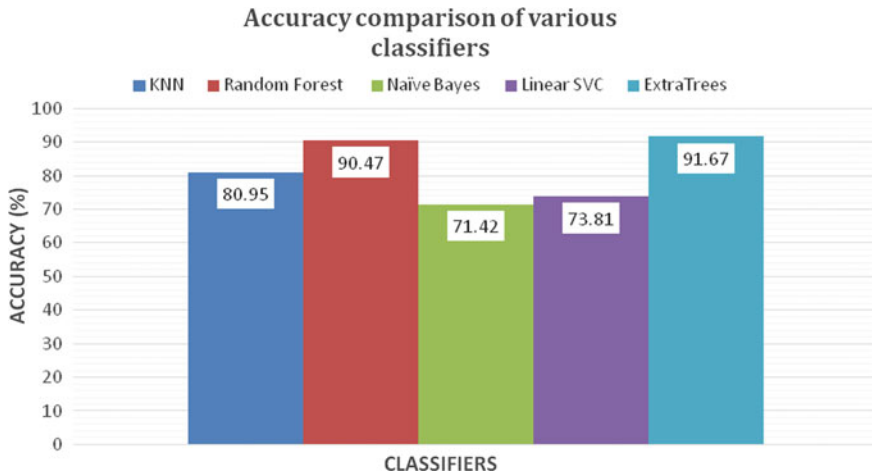


Fig. 5 Accuracy comparison of various classifiers

References

1. Elshafeiy E, El-Desouky AI, Elghamrawy S (2018) Prediction of liver diseases based on machine learning technique for big data. In: The International conference on advanced machine learning technologies and applications (AMLTA2018), pp 362–374. https://doi.org/10.1007/978-3-319-74690-6_36
2. Singaravelu M, Rajapraksh S, Krishnan S, Karthik K (2018) Classification of liver patient dataset using machine learning algorithms. *Int J Eng Technol* 7(3):323–326. <https://doi.org/10.14419/ijet.v7i3.34.19217>
3. Sazzadur Rahman AKM, Javed Mehedi Shamrat FM, Tasnim Z, Roy J, Hossain SA (2019) A comparative study on liver disease prediction using supervised machine learning algorithms. *Int J Sci Technol Res* 8(11). ISSN 2277-8616
4. Gulia A, Vohra R, Rani P (2014) Liver patient classification using intelligent techniques. *Int J Comput Sc In Technol (IJCSIT)* 5(4):5110–5115
5. Kumar Y, Sahoo G (2013) Prediction of different types of liver diseases using rule-based classification model. *Technol Health Care* 21:417–432
6. Singha J, Baggab S, Kaurc R (2020) Software-based prediction of liver disease with feature selection and classification techniques. In: International conference on computational intelligence and data science (ICCIDS 2019), vol 167
7. Takkar S, Singh A, Pandey B (2017) Application of machine learning algorithms to a well-defined clinical problem: liver disease. *Int J E-Health Med Commun (IJEHMC)* 8(4)
8. Ramana BV, Prof. Surendra Prasad Babu M, Prof. Venkateswarlu NB (2011) A critical study of selected classification algorithms for liver disease diagnosis. *Int J Database Manage Syst (IJDMS)* 3(2).
9. Fatima M, Pasha M (2017) Survey of machine learning algorithms for disease diagnostic. *J Int Learn Syst Appl* 09(01)
10. Singh A, Pandey B (2014) Intelligent techniques and applications in liver disorders: a survey. *Int J Biomed Eng Technol* 16
11. Arsene CTC, Lisboa PJ (2012) Bayesian neural network applied in medical survival analysis of primary biliary cirrhosis. In: IEEE 14th international conference on computer modelling and simulation

12. Bucak IO, Baki S (2010) Diagnosis of liver disease by using CMAC neural network approach. *Expert Syst Appl* 37(9)
13. Lin R-S, Lee F-Y, Lee S-D, Tsai Y-T, Lin HC, Rei-Hwa, L, Wan-Ching H, Cheng-Chun H-S, Wang K-J (1995) Endotoxemia in patients with chronic liver diseases: relationship to severity of liver diseases, presence of esophageal varices, and hyperdynamic circulation. *J Hepatology* 22(2)

Sentiment Analysis of Twitter Data Using Deep Learning



Pranati Rakshit, Pronit Sarkar, Debosmita Ghosh, Shubhankar Roy, Subhadip Talukder, and Partha Sarathi Chakraborty

Abstract In today's world, the use of social networking websites is in the next level. People express their thoughts and opinions on any brand, product or any social events via these sites. Sentimental analysis often uses natural language processing (NLP) to obtain a sentiment behind the text, tweet or comments. In this modern world, sentiment analysis has become one of the most efficient way to mine the public emotions, opinions based on their particular topic of their interest. We have used datasets which consists of tweets from Twitter which contains tweets from various domains. This paper described an approach where a stream of tweets is pre-processed then classified based on the emotion within the text. We used sequential model, with long short-term memory (LSTM) as a layer to train our model. This trained model can be used to analyse any tweets or blogs to obtain sentiments behind that. We have achieved an accuracy of 81.5% from our developed model.

Keywords Sentimental analysis · Tweet data · NLP · Long short-term memory · Accuracy

1 Introduction

Nowadays, social media become an effective path to people for collecting various kind of information [1] as well as a platform for all those people who wants to share their opinions through these sites without any hesitation [2]. One of them is Twitter which has millions of users who always stay active. It is said that at least 500,000 new tweets are posted in Twitter on every single day [3]. Now, Twitter become most used social site for public to share their views, thoughts and opinion on particular topic they are interested in and also Twitter has become a platform for marketing of brands [4].

These social media platforms especially Twitter work as a place to analyse the opportunity for a new product or a business, also as a place to promote or analyse

P. Rakshit (✉) · P. Sarkar · D. Ghosh · S. Roy · S. Talukder · P. S. Chakraborty
Department of Computer Science and Engineering, JIS College of Engineering, Kalyani, India
e-mail: pranati.rakshit@jiscollege.ac.in

existing products to enhance their business [5, 6]. As more and more people are getting connected with the Internet, any product marketed there can be reached to the great number of people. These platforms also generate tons of data from huge amount of users all around the globe on various domains. Data collected by these platforms is also said to be better sample compared to data collected by others means, as they have data on every domain, from all geographical locations, of all age groups. Sentiment-rich data in form of tweets, blogs, comments, reviews, posts, etc. can be used to analyse for the betterment of products and services, which as a result is used for the growth of the business [7].

Sentimental analysis actually analysed the emotions or sentiment behind any blogs, posts, reviews or comments. It usually helps users by gathering all information for a particular product and tell us whether we should buy it or not. This analysis is a huge process of using of pre-processing [8] and text analytics to analysis the sentiment of tweets [9]. Natural language processing plays a major role behind this sentimental analysis [10].

Feature level sentimental analysis is a part of analysis which provides a fine-grained sentiment analysis on certain opinion targets and has wider range of applications on E-business.

Sentimental analysis is also known as opinion mining [11]. It is a very effective tool that is used in any business field or any social media. In any online website, when a company wants to sell a product, it is very important to know about how a customer react about the product for further progress. Sentimental analysis is one among the foremost common ways which analyses an approaching message and tells whether the elemental estimation is positive, negative or unbiased [12]. Sentiment interpretation is one among the toughest challenges within the production of tongue, and sometimes individuals fail to properly analyse emotions (Fig. 1).

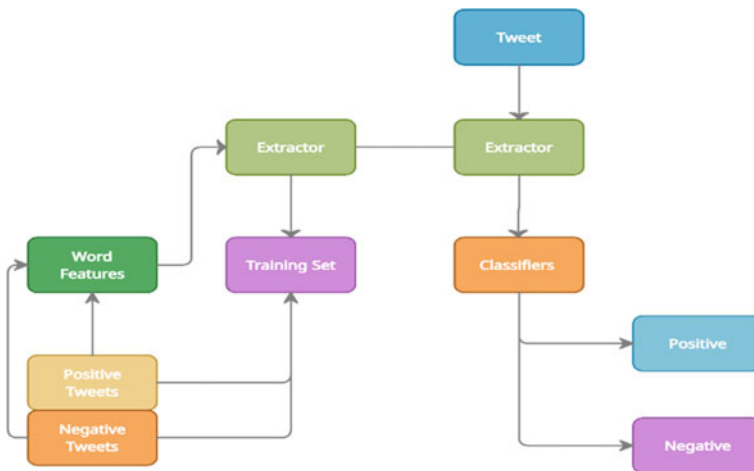


Fig. 1 Sentimental analysis architecture

- (a) Sentimental analysis has some real-life example like reputation management, brand sentimental analysis, etc. Essentially, the application of sentimental analysis gives analysis flexibility and insight into the awarding the organisation and its products.
- (b) Sentimental analysis helps to solve the problem of dealing with huge volume of data. Users can easily find out the emotional tone of any reviews or comments.
- (c) Sentimental analysis helps to mine any data and extract the emotions that underlie social media conversations.
- (d) Sentimental analysis helps to differentiate the positive or negative comments by analysing all data.

Section 2 describes the literature survey, Sect. 3 describes methodology, and result and discussions are depicted in Sect. 4, whereas conclusion is expressed in Sect. 5 with future direction.

2 Literature Survey

From the past few years, we recommended opinion of users, domain experts for making a decision in today's life. According to business changing nowadays, we have to be more effective to our customer's need, e.g. which brand is more popular and effective for the certain product, whether the market value of the product is high or efficient, whether the current series are good or not. According to Appel et al. [13] and Stephen et al. [14], opinion mining can be called as sentimental analysis which plays an important role in this process. It is the objective of emotions like sentiments, expressions that are basic properties in natural language processing by Liu [15], Pak and Paroubek [16], Vinodhini and Chandrasekaran [17], Maks and Vossen [18]. There are some techniques which applied to omit emotions from unstructured data. The analysis of sentiment deals with the identification and classification of opinions or feelings that are present within the source text. According to Zvarevashe and Olugbara [19] and Saberi and Saad et al. [20] within the sort of tweets, status updates, reviews and blog posts, etc., social media generates an enormous amount of sentiment-rich data. In understanding the opinion of the gang, sentimental analysis of this user-generated data is extremely helpful by Feldman [21] and Madhoushi et al. [22]. A path that makes us understand the essential sentiment of the users that they want to express by commenting on our blog by Neethu and Rajasree [23]. Basically, we will be making computer capable of understanding the contents of the text, including the sentiment within them. It will help us to extract all details and observations found in the text and to categorise and arrange the record themselves according to Chowdhury [24] and Indurkhya and Damerou [25]. As a Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches, Twitter sentiment analysis is difficult compared to basic sentiment analysis, thanks to the presence of slang words and wrong spellings by Patel et al. [26], the utmost character limit allowed on Twitter

is 140. For the analysis of feelings from the text, a machine learning approach is often used according to Kharde and Sonawane et al. [27].

3 Methodology

Here, we use sequential model with LSTM as a layer in order to instruct computers to try to learn whatever comes obviously and gain knowledge from encounter. Calculations using machine learning make use of computational strategies to simply “learn” data from data without counting on a preordained condition.

3.1 *Data/Text Cleaning*

In our dataset, there are several types of tweets like comments, review on merchandise, posts concerning one thing, live tweets on matches, etc.; thus, it is important to clean our dataset and each text so as that its square measure usually processed simply. So, cleaning each text was our first initiative. It is done in order to make it easier to process. Filtering out the unwanted symbols, tags, that we do not need so that they do not need to be looked at or processed. At first, we tokenized each sentence in order that it is easier to process. Tokenization is basically the method of splitting or tokenizing a string, text into a list of tokens. Then we clean the text of varied symbols and website marks like @, https, .com, etc. Next, we worked on “stopwords” from each text. In NLP, useless or words with very little meaning are mentioned as stopwords, some commonly used stopwords are “the”, “a”, “an”, “in”. These kinds of words have been programmed to ignore, both when indexing entries for searching and when retrieving them as the results of a query.

3.2 *Split and Train*

While operating with datasets, associate formula of machine learning typically works in two steps. We have a tendency to typically split information around 80–20% train-test ratio. Here, we have around 1.6 million tweets from which we have taken a sample of 5 lac tweets. We train the model using data which are typically referred as training set or training data. Training data is that data which already consists of actual value that must have the ground truth values and thus the algorithm creates changes to the value of the parameters to interpret for training the model. From the 5 lac tweets from the sample, 4 lac tweets are used for training the model and 1 lac for testing purpose.

We can apply the trained model to evaluate any text now on the basis of their score after evaluation by the model and check with the sentimental threshold and classify

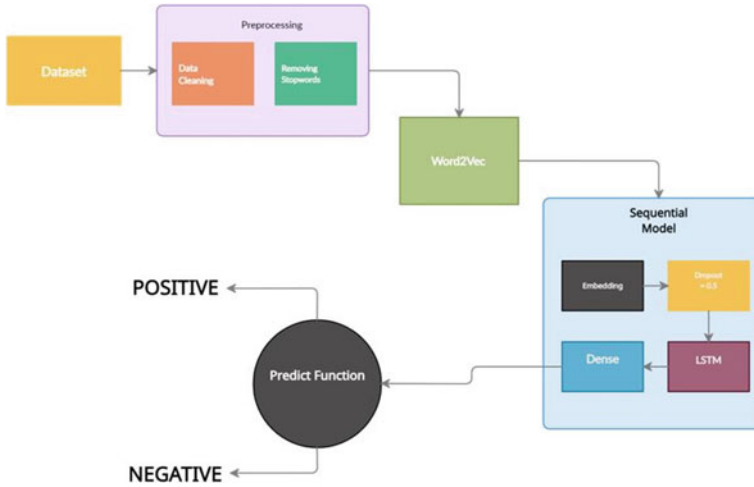


Fig. 2 Workflow model

them as positive, negative or neutral. This model can also be connected in back-end of a blogging site to analyse the sentiments of the blogs automatically, without any interaction (Fig. 2).

3.3 Embedding Layer

As an input to the embedding layer, tweets are given. Any tweet is split into tokens. Each token is transformed into a vector of fixed-sized words, often known as embedding by phrase.

Pre-trained word embedding Word2vec is used for generating the word vector for each token. Among the pre-trained word embeddings, different word embeddings are developed using tweets using the Word2vec algorithm.

The embedding layer output is fixed-sized word vectors that are supplied to the Sequential model as an input for training the model.

4 Result and Discussion

It is very common to remove the stopwords (such as “did”, “doing”, “an”, “the”, “and”) in the pre-processing step as these words tend to have no or very little meaning.

Nevertheless, sometime these stopwords may also have a high impact on the result of certain text. So, to test that, we have conducted an experiment to measure the impact of stopwords removal on the sentimental analysis. We have compared the results of

sequential model with LSTM, both with and without removing the stopwords. It was observed that the model trained with stopwords has outperformed the model trained without stopwords. Hence, we can conclude that removing the stopwords should be avoided, especially for sentimental analysis. We have calculated the accuracy of our model with and without stopwords. After applying the model on our tweet dataset, we have a tendency to get accuracy of seventy nine percent to eighty one point five percent (79% to 81.5%). During this method of analysing the sentiment of tweets, we have used LSTM along of the layers in our ordered layer together with the embedding layer, dropout layer to prevent overfitting, to teach our model. One in every of the foremost uses of LSTM is that this sentimental analysis. The foremost distinction between feedforward neural networks and LSTM is that it is feedback affiliation and it will method entire sequence of data that makes it extremely economical. Sentimental analysis square measure usually machine controlled and should supply resolution to sizable quantity of issues and judgments square measure usually taken on the thought of a giant quantity of information rather than plain intuition that is not correct. It can also be associated integral part of the approach to plug analysis and client service. You may not solely see what individuals believe your own merchandise or services; you may additionally see what your rivals take into account them. With sentiment analysis, the final shopper expertise of your customer's square measure usually discovered simply.

Accuracy (without stopwords)	Accuracy (with stopwords)
79%	81.5%

5 Conclusion

In this present work, we have taken a dataset consisting of tweet data which have various sorts of tweets associated with quite different topics and created a model to analyse each tweet's sentiment. Using this model, we will analyse the other text and determine the sentiment behind it. So, we have got the highest accuracy of 81.5%. As a future direction of this work, we will work with larger dataset and will try to optimise the hyperparameters.

References

1. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. Sentiment analysis of Twitter data. Department of Computer Science, Columbia University, New York, NY 10027 USA
2. Kharde VA. Sentiment analysis of twitter data: a survey of techniques. Department of Computer Engineering, Pune Institute of Computer Technology, Pune University of Pune (India)

3. Liao S, Wang J, Yua R, Sato K, Cheng Z. CNN for situations understanding based on sentiment analysis of twitter data
4. Shah S, Kumar K, Saravanaguru RK. Sentimental analysis of twitter data using classifier algorithms software developer. School of Computing Science and Engineering, VIT University, Vellore, India
5. Adwan OY, Al-Tawil M, Huneiti AM, Al-Dibsi RH, Shahin RA, Abeer A. Twitter sentiment analysis approaches: a survey. Abu Zayed University of Jordan, Amman, Jordan. <https://doi.org/10.3991/ijxx.vx.ix.xxxx>
6. Alayba AM, Palade V, England M, Iqbal R (2017) Arabic language sentiment analysis on health services, pp 114–118
7. Neethu MS, Rajasree R (2013) Sentiment analysis in twitter using machine learning techniques. In: 2013 Fourth International conference on computing, communications and networking technologies (ICCCNT), pp 1–5. <https://doi.org/10.1109/ICCCNT.2013.6726818>
8. Haddi E, Liu X, Yong S. The role of text pre-processing in sentiment analysis
9. Kabir AI, Karim R, Newaz S, Hossain MI. The power of social media analytics: text analytics based on sentiment analysis and word clouds on R
10. Shrestha H, Dhasarathan C, Munisamy S, Jayavel A. Natural language processing based sentimental analysis of Hindi (SAH) script an optimization approach
11. Kumar P, Jaiswal UC. A comparative study on sentiment analysis and opinion mining
12. Mehra R, Bedi MK, Singh G, Arora R, Bala T, Saxena S (2017) Sentimental analysis using fuzzy and Naive Bayes. In: 2017 International conference on computing methodologies and communication (ICCMC), pp 945–950. <https://doi.org/10.1109/ICCMC.2017.8282607>
13. Appel G, Grewal L, Hadi R, Stephen AT (2020) The future of social media in marketing. *J Acad Mark Sci* 48(1):79–95
14. Stephen A, Hadi R, Grewal L, Appel G (2019) The future of social media in marketing. *J Acad Market Sci* 48
15. Liu B (2012) Sentiment analysis and opinion mining. *Syn Lect Human Lang Technol* 5(1):1–167
16. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. *LREc* 10:1320–1326
17. Vinodhini G, Chandrasekaran RM (2012) Sentiment analysis and opinion mining: a survey. *Int J* 2(6):282–292
18. Maks I, Vossen P (2012) A lexicon model for deep sentiment analysis and opinion mining applications. *Decis Support Syst* 53(4):680–688
19. Zvarevashe K, Olugbara OO (2018) A framework for sentiment analysis with opinion mining of hotel reviews. In: 2018 Conference on information communications technology and society (ICTAS). IEEE, pp 1–4
20. Saberi B, Saad S (2017) Sentiment analysis or opinion mining: a review. *Int J Adv Sci Eng Inf Technol* 7:1660–1667
21. Feldman R (2013) Techniques and applications for sentiment analysis. *Commun ACM* 56(4):82–89
22. Madhoushi Z, Hamdan AR, Zainudin S (2015) Sentiment analysis techniques in recent works. In: 2015 Science and information conference (SAI). IEEE, pp 288–291
23. Neethu MS, Rajasree R (2013) Sentiment analysis in Twitter using machine learning techniques. In: 2013 Fourth International conference on computing, communications and networking technologies (ICCCNT). IEEE, pp 1–5
24. Chowdhury GG (2003) Natural language processing. *Ann Rev Inf Sci Technol* 37(1):51–89
25. Indurkha N, Damerau FJ (eds) (2010) Handbook of natural language processing, vol 2. CRC Press
26. Patel AP, Patel AV, Butani SG, Sawant PB (2017) Literature survey on sentiment analysis of Twitter data using machine learning approaches. *IJRST-Int J Innov Res Sci Technol* 3(10)
27. Kharde V, Sonawane P (2016) Sentiment analysis of twitter data: a survey of techniques. arXiv preprint: arXiv:1601.06971

Stun Gun PCB Circuit with Arduino Power Shield



Mehaboob Mujawar, D. Vijaya Saradhi, and K. Swetha

Abstract The project demonstrates PCB design for an Arduino-powered stun gun which transforms low current pulses into high voltage signals. Stun guns are interactive weapons that need direct contact with the subject to work efficiently. Generally, 3–5 s of contact can cause muscle disruption. This project involves a 12 V power supply, Arduino, power shield, 555 timers, a MOSFET transistor, a transformer, and a voltage doubler to model the stun gun. It paralyzes an attacker with fewer penalties and allows the user to pursue more safety precautions. Not all stun guns are created the same, they come in different sizes and voltages to perform differently.

Keywords Arduino · Power shield · 555 timer · MOSFET transistor · Voltage doubler · Voltage regulator

1 Introduction

Stun gun is a device used for self-protection against criminals. It stops the normal functioning of the target. The stun gun immobilizes the target by sending shock waves. The waves affect the nervous system of the target. The stun gun circuit is designed to convert low current pulses into high voltage signals. The input of the transformer is fed with low current pulses, and the transformer outputs high voltage signals; these are further fed to a voltage multiplier to increase the voltage even more. There are several applications of stun gun such as: (1) Stun guns are used by individuals for security purposes from intruders. (2) Tasers are used by police in many countries including Canada, Brazil, New Zealand, Australia, France, Germany,

M. Mujawar (✉)
Goa College of Engineering, Goa, India
e-mail: mehaboob311134@gmail.com

D. Vijaya Saradhi · K. Swetha
Malineni Perumalu Educational Society Group of Institutions, Andhra Pradesh, India
e-mail: saradhi.gitam@gmail.com

K. Swetha
e-mail: swethakatragadda88@gmail.com

and UK. (3) Stun guns are used for protection against wild animals. (4) They can be chosen as equipment for modern warfare. The main advantages of stun guns include: (1) Unlike bullet guns, stun guns do not cause any long-term harm to the target. It paralyzes the target for a short period without causing any serious injuries. (2) They are light in weight hence easy to carry. (3) They are cost-effective. (4) The shock waves can penetrate even through clothing and immobilize the target; hence, they are reliable and effective. Every device has its pros and cons [6], so does the stun gun. The implementation of a stun gun circuit on hardware has to be done with uttermost care and precautions as in high voltage pulses [3]. The output of the stun gun should not be touched with bare hands as the shock signals can disrupt the nervous system. Although the stun gun is said to not cause any serious injuries, it can turn out to be lethal in the case of people with cardiac issues or neurological disorders. The stun gun works effectively only with it comes in contact with the targets body. Stun guns are made of different shapes and sizes. Fake Smartphone: Stun gun, disguised as a smartphone, is less obvious to the attacker. Flashlight: Investigate with a beam of light. Batons: Stun guns in the shape of a stick. Tasers: These have probes to shoot electrodes from a safe distance. Knuckle Stun guns: These can be worn on the finger like a ring. Mini Stun guns: These are mini devices fitted in flash drives and other everyday objects.

2 Literature Survey

In paper [1], author constructs a handheld stun gun with components such as a 9 V battery, IC 555 timer, a MOSFET IRF530 transistor, a transformer with a turn ratio of 1:10, diodes, and capacitors with a voltage rating of 2500 V to handle high input voltages. When the switch is pressed, the timer IC 555 goes into action; the IC 555 is connected in such a configuration that it acts like an A-stable multivibrator. A-stable multivibrators do not require any external triggering to change their state from Level 1 to Level 0 or from Level 0 to Level 1. The author has configured the timer to produce a low current, low-frequency wave of about 50 Hz with a duty cycle of 75%. Since the timer output frequency is low, the output of the timer is given to the IRF530 MOSFET transistor. This acts as a switch for the transformer producing a pulsed DC voltage across the transformer. The transformer has a turn ratio of 1:10 and as the transformer receives a 9 V input to the transformer's primary the voltage across the secondary of the transformer comes to about 1000 V. The output across the transformer's secondary is given to 5 voltage doubler stages which increases the voltage across the output pins by 10 times giving a 10 kV output. Similarities between our project and paper [1] are, it also uses 555 timer as a stable multivibrator, transformer with the 1:10 turn ratio, voltage doubler arrangement and specifications. Differences between our project and paper [1] are instead of a 9 V power supply directly from the battery to the 555 timers, we will be using an Arduino with a power shield as a power supply to the stun gun. Arduino is used for future expansion to add modules like GPS and GMS modules and for Bluetooth for direct compatibility

with a mobile application. The author in paper [2] illuminated how the stun gun works on the principle of electric shock gun effect, which generates high voltage with low current inputs. It states that the subject should be in direct contact with the stun gun to have its effects on it, which affects the nervous system making the subject paralyzed. The model used two main components in the design: (1) A-stable multiplier to produce high frequency. (2) Voltage doubler circuit to produce high DC voltage from low AC power. 9 V supply was used as an input, and output was obtained with the help of a multiplier and a voltage doubler circuit. The author further defined it to be a safety or security purpose weapon to protect oneself from animals. The author designed a shock gun that gave a noticeable spark and a crack sound, giving a signal that the attacker has been captured. The author further explained the power supply and design. The author further explained that no external triggering was required to make the multivibrator switch between its two unstable states. The author used the Cockcroft–Walton principle for voltage doubler to increase voltage levels. The model preferred half-wave regulators since high voltage outputs were required. The following equation was obtained: $V_{\text{output}} = (2V_{\text{input}} + 1.414) S$. Here, pulses are produced using an A-stable multiplier, while we use 555 timers for the same. 9 V battery is used here, while we are using 12 V input, Arduino, and power shield as a power source. In paper [3], authors have designed an electrical stun gun that can majorly be used as a safety weapon by women for self-defense. The stun gun immediately effects the sensory and motor nerves, hence resulting in strong muscle contractions due to which the movement of the assaulter is resisted for a period of time. In various countries including in India, the use of stun gun is strictly prohibited, but when the stun gun is well equipped with the latest modern technology like global system for mobile communications (**GSM**), Global Positioning System (**GPS**), and radio frequency identification module (**RFID**), it could be safely used. For proper authorized use of the stun gun, RFID module is used. RFID plays a crucial role in identifying and tracking objects. To get exact location of the user, a GPS module is used which can be easily monitored and GSM module in turn transfers entire information to the nearest police station; also, text messages are sent to the close contacts of the user. In paper [4], the authors have aimed at designing a safety dress for women based on stun gun technology with GPS to control crime against women. The various components used in this project are an ATMEGA controller, GPS module, ESP module, boost converter, and buzzer. The ATMEGA controller acts as the CPU of the entire circuit. It is programmed to take inputs from the panic key and the ESP module. It also consists of an android application; the safety dress and mobile phone are synchronized with the help of the ESP module. The ESP module is used to connect the microcontroller to a Wi-Fi network. The boost converter does the work of the stun gun by sending shock pulses [5]. The input voltage is stepped up by this DC power converter. The buzzer is used as an alerting device in the system. The GPS module gives the location of the woman wearing the safety dress once the panic key is pressed.

Table 1 Components used and its specifications

Name of components	Specification
DC supply input	Barrel Jack 12 V power supply connector (01)
555 timer	LM555 timer (01)
Linear voltage regulator	Adjustable positive linear voltage regulator LM317 (01) fixed linear voltage regulator LM7805 (01)
Switching transistor	IRF740
Capacitor	10 μ F (11), 1 μ F (2), 0.1 μ F (2)
Diode	BA159 (10)
LED	02
Connectors	01X01 male connector (02), 01X05 male connector (04), 01X05 female connector (04)
Resistor	750 Ω (01), 330 Ω (01), 560 Ω (02)
Potentiometer	01
Switch	01
Transformer	1:10 ratio transformer (01)

3 Design Methodology

See Table 1.

4 Stun Gun Circuit

The below circuit represents the input circuit of the stun gun; the input circuit consists of a 12 V power supply which is provided to the circuit by an Arduino power shield. A LED is used to indicate if the power supply is ON. This input is given to the LM555 timer which is configured to work in A-stable multivibrator mode, The A-stable multivibrator provides a rectangular wave/pulsed DC output of about 50 Hz; this wave is given to the MOSFET transistor which acts as a switch for the current flowing through the transformer producing a 9 V pulsed DC output across the primary of the transformer.

Timer IC 555 in A-stable Mode: Initially, at time zero, capacitor has no voltage drop across it, so the output of the IC is high, within the IC itself there is a transistor that helps capacitor to charge from resistor R2 and R1. As capacitor C is charging through resistor R2 and resistor R1, and as the voltage around the capacitor becomes greater than two-third of the input supply (+VCC), a LOW state is given at the output, and the discharge transistor is turned on given a path for the capacitor to discharge through resistor R2. As the capacitor is discharging and the voltage across the capacitor goes below one-third the input supplies, a HIGH state is given at the

output of the IC [7]. The charging and discharging periods are controlled by the values of the resistors and the capacitors producing a free-running multivibrator.

The output of the transformer's secondary is about 1000 V; this is given as input to the voltage doubler stages; there are five voltage doubler stages which bring the output to about 10 kV. Working of voltage doubler circuit: **Let's first consider the negative** half cycle where the D1 is conducting, and C1 gets a path for charging itself to a max voltage value equal to the max value across the secondary of the transformer and as the voltage starts to drop across the secondary of the transformer (in magnitude) D1 becomes reverse biased, and there is no discharge path for the charge across C1. Now, as C1 is in series with the input voltage, it acts like an additional voltage source connected in series. When D1 is reverse biased at the same time, D2 becomes forward biased, and C2 starts charging. As the +ve half cycle starts, only, D2 is conducting, and D1 is not conducting leaving no path for capacitor C1 to discharge itself. Now, conductor C2 starts to charge itself through diode D2, but since capacitor C1 now acts like an additional voltage source connected to the transformers secondary, so now, C2 charges to two times the peak value across the transformers secondary. As the negative half of the next cycle approaches, D3 starts conducting, and C3 starts charging through D3, and for the positive cycle, C4 starts charging since D4 starts conducting; by this time, the capacitor C4 has four times the peak value across the transformers secondary. This continues till the capacitor C10 was charged to ten times the peak value across the transformers secondary.

Arduino power supply shield: Arduino shields are the boards, which are placed over the Arduino board to increase its functionalities. A variety of shield is created for different tasks like Arduino motor shield, Ethernet shields, and power supply shield. Arduino power supply shield is needed since multiple output voltage power supply is most important while developing electronics projects or circuits. Different sensors, modules (GPS, GSM) need different voltage and current as input for them to run efficiently. The main input voltage to the shield is given through the Barrel jack (12 V approx.). To ensure a steady constant voltage supply through all operational conditions, two voltage regulator IC's (LM7805_TO220 and LM317L_SOT-89) are placed which will convert 12–5v output and 12–3.3v output, respectively. Resistors R3 (330 ohms) and R4 (560 ohms) are used as voltage divider circuit to convert 12v to 3.3v. The output capacitors C14 and C15 are important and are used to filter out the varying DC output voltage so that at the output, a constant voltage (DC) is supplied (to reduce ripple voltage). These output voltages are in turn fed to the stun gun circuit (Figs. 1 and 2).

5 Conclusion

The objective of the project is to design a stun gun with an Arduino power supply. The schematic, footprint, and PCB design of the stun gun were done successfully using Ki-CAD software. The stun gun can be used by people to protect themselves from

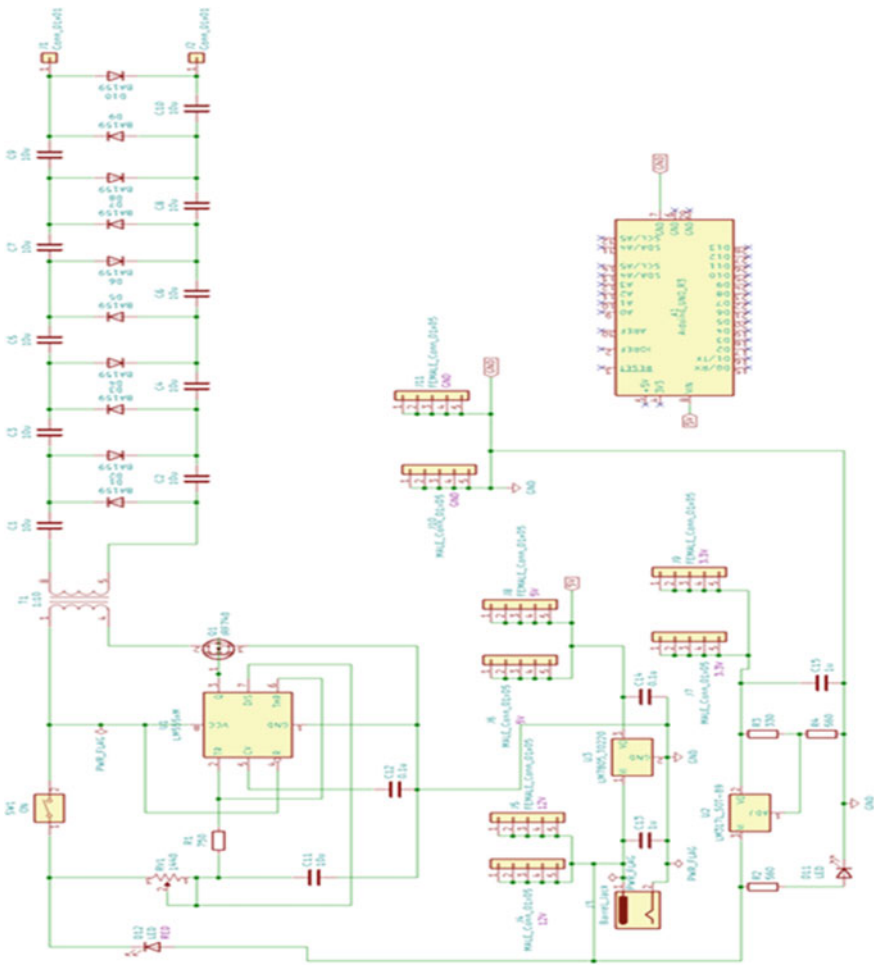


Fig. 1 Schematic of the circuit in Ki-CAD

robbers or rapists. They are handy and easy to use and also do not cause long-term harm.

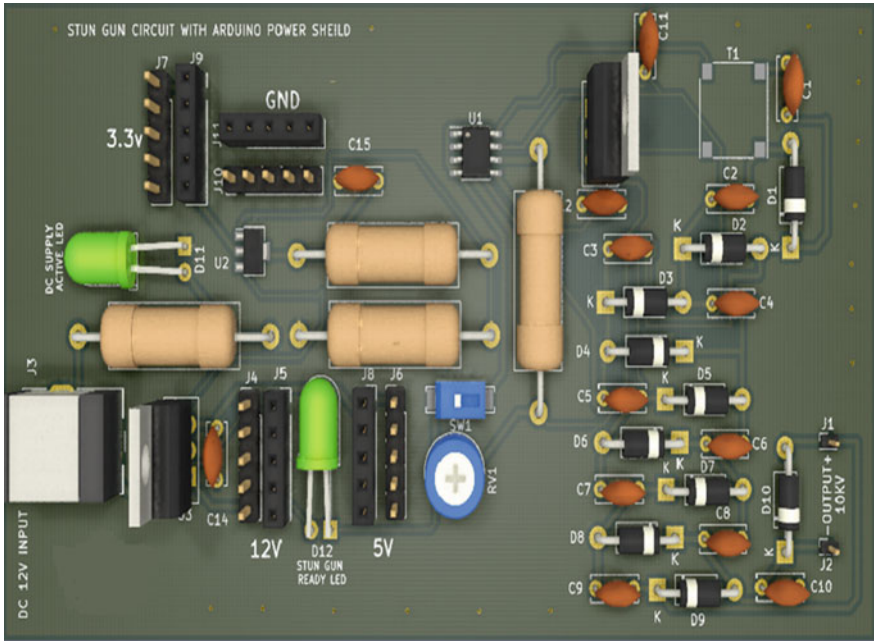


Fig. 2 PCB model of stun gun

References

1. Abdul Salam A, Rashid O, Rahul KR, Yashwanth, Sruthy K (2016) Stun Gun. *Int J Innov Res Manag Eng Technol* 1(4)
2. Harsh P, Megh P, Gupta S (2018) Design and analysis of electric shock gun with cheapest cost of manufacturing. In: *FIRST Int J Innov Res Sci Techno* 5(1). ISSN (online) 2349-6010
3. Priya M, Varshini M, Anita T (2019) Handheld electrical weapon with GPS and GSM for self defence system. *Int J Adv Res Educ Technol (IJARET)* 2(1)
4. Gowrishankar P, Logeswaran T, Surendar V, Shriram B, Sivamuruganandam J, Vasanth S (2019) Women safety dress using stun gun technology. *Int J Innov Technol Exploring Eng (IJITEE)* 9(2). ISSN 2278-3075
5. Mattei E, Censi F, Calcagnini G (2020) Electrical stun gun and modern implantable cardiac stimulators: update for a new stun gun model. *Health Phys.* Publish Ahead of Print. <https://doi.org/10.1097/HP.0000000000001332>
6. Choudhary R, Sabri I (2010) Taser technology: medical, legal, ethical & social implications of introduction of Taser gun in India. *J Indian Acad Forensic Med* 32(4)
7. Bleetman A, Lee C, Steyn R (2004) Introduction of the Taser into British policing. Implications for UK emergency departments: an overview of electronic weaponry. *Emerg Med J* 21:136–140

Comparative Study on Tuning PID and FOPID Using Genetic Algorithm for Heart Rate Control of a Pacemaker



Ritika Saini, Ravi Kumar, Rijhi Dey, Ujjwal Mondal,
and Rudra Sankar Dhar

Abstract In this paper the pacemaker is integrated with two different controllers, namely PID and FOPID, as they are widely used in today's scenario, followed with an optimization technique, called Genetic Algorithm, to obtain a more tuned result, when the whole system is integrated with the heart. The aim of the project, is to do a comparative study on tuning PID and FOPID using Genetic Algorithm for heart rate control of a pacemaker, by doing analysis of time domain criterions of the control system.

Keywords Pacemaker · Heart · Proportional integral derivative controller (PID) · Fractional order proportional integral derivative controller (FOPID) · Genetic algorithm (GA)

1 Introduction

Heart failure, also called congestive heart failure, occurs when the heart muscle is unable to pump blood as well as it should. Several treatments can improve its symptoms and help one to live longer. It can be prevented by controlling conditions that cause heart failure, such as obesity, diabetes, high blood pressure or coronary artery disease. The electrode device that is used to initiate the heartbeat when the heart's intrinsic electrical system cannot effectively generate a rate adequate

R. Saini (✉) · R. Kumar · R. Dey

Department of Electronics and Communication, Sikkim Manipal Institute of Technology, Sikkim Manipal University, East Sikkim, Majitar, Rangpo 737136, India
e-mail: ritika99saini@gmail.com

R. Dey

e-mail: rijhi.d@smit.smu.edu.in

U. Mondal

Instrumentation Engineering, Applied Physics, University of Calcutta, Kolkata, India

R. S. Dhar

Department of Electronics and Communication Engineering, National Institute of Technology Mizoram, Aizawl, India

to support cardiac output, is called a Pacemaker. It is implanted under the skin to help manage irregular heartbeats called arrhythmias, and also can be used to track and record the heartbeat. At present, several researches have been done, and are still ongoing on optimizing the output of Pacemaker. The Proportional Integral Derivative Controller (PID) is an immensely used feedback controller, and the most accurate-stable controller; the control process variables are obtained using a control loop feedback mechanism. The input of the PID controller is an error value, difference between measured process variable and desired setpoint, and thus, is used in regulating the heart rate control of the pacemaker. In this paper, the genetic algorithm for tuning PID controlling parameters has found the optimal solution with below 10% overshoot percentage with 100 generations. on 0.4 mutation effect and 0.8 crossover effect [1–3]. In this paper, the proposed controller has much better time domain characteristics compared with other conventional techniques. In this paper, an Adaptive Loop incorporating MIT rule with the Delta rule is implemented on the PID controller, the adaptive correction factor is used to modify all parameters [4, 5]. PID control stands for proportional–integral–derivative control. It is a closed-loop system controller which includes a feedback control system [6–8]. To generate an error signal, a fixed point is used by the system to evaluate the feedback variables. The system output is altered as per it. This procedure continues until ‘Zero’ is reached by the error else the value becomes same as the fixed point. Calculating/Controlling these three parameters – the proportional, integral and derivative of how much deviation is there from the desired set point value—different actions can be controlled for a specific work; thus, this type of control is also coined as three-term control. The control system family, considers it to be the best controller. The proportional, derivative and integral parameters can be expressed as— K_p , K_d and K_i . The closed loop control system, gets affected by all these three parameters. They also affect the settling time, rise time, steady state error and also the overshoot. The Fractional Order ($PI^\lambda D^\delta$) controller (FOPID) same as conventional PID controller but integral order (λ) and derivative order (δ) are fractional. The new and good performance extension for FOPID can be provided by Fractional calculus because of the flexibility order of fractional calculus. The parameters of FOPID comprise proportionality constant, integral constant, derivative constant, integral order (λ) and derivative order (δ). The design of ($PI^\lambda D^\delta$) controller needs to optimize five parameters while the design of conventional PID controller needs only three parameters to optimize. A cardiac pacemaker is a small, battery-operated medical device. It senses irregularity of the heart, and sends signal to the heart and makes it beat with the correct pace, thus the function of the electrical conduction system of the heart is replaced. The human heart consists of the sinoatrial node, a mass of tissue in the upper right chamber of the heart, responsible for sending a regular pulse 60 to 100 times per minute, which goes through the conduction pathways and causes the heart’s lower chambers contract and pump out blood. Sometimes, under some abnormal conditions any other heart’s muscle starts to play as the sinus node, this condition is called arrhythmia, this causes a faster or slower heartbeat. May occur when either the heart’s natural pacemaker (SA node) slows down or, due to an interrupted conduction pathway.

2 Methodology

When a PID controller is used along with the pacemaker, to stabilize and control the heart rate, we observe that the output can be more optimized, therefore the main objective is to improve the performance of a conventional PID controller to stabilize the operation of cardiac pacemaker (Fig. 1; Table 1).

$$\text{Transfer function of Heart : } H(s) = \frac{169}{s^2+20.8} \tag{1}$$

$$\text{Transfer function of Pacemaker: } P(s) = \frac{8}{s+8} \tag{2}$$

2.1 Using Simulink to Obtain Controller Parameters of Conventional Controllers

We have analyzed the performance of the controllers on the basis of these parameters:

- Overshoot Percentage.
- Settling Time.
- Rise Time
- Steady-State Error

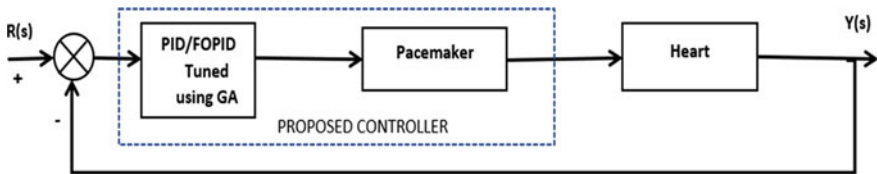


Fig. 1 Block diagram of the control system

Table 1 Comparison between control variables of conventional and tuned PID controller

Controller variable	Conventional PID controller	Tuned PID controller using genetic algorithm
K_p	1.683	1.99
K_i	0.9786	21.99
K_d	0.22	59.99

2.2 Genetic Algorithm

Genetic Algorithm (GA) is widely used for tuning different controllers. Any constrained and unconstrained maximization/minimization problem based on natural selection process which mimics natural evolution can be solved with help of it. The individuals are selected randomly from present population and are used as parents to generate the offspring for the next generation; the population “evolve” towards a superlative solution (Fig. 2).

Fig.2 Flowchart of genetic algorithm

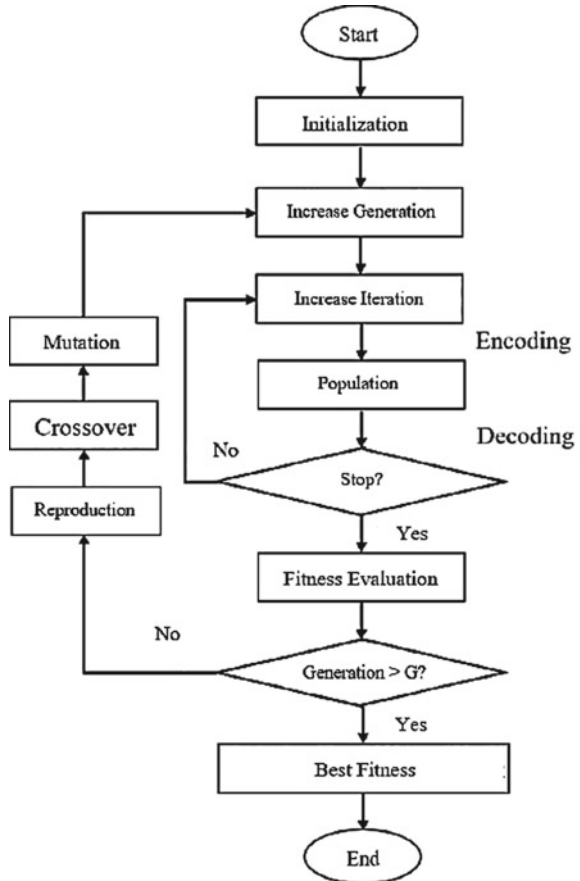


Table 2 Comparison between time domain criterions of conventional and tuned PID controller

Parameter	Conventional PID controller	Tuned PID controller using genetic algorithm
Steady state value (V)	50	1
Settling time (s)	3.8	0.30
Overshoot peak value (V)	76	1.85
Overshoot percentage (%)	52	85

2.3 Steps for Tuning PID/FOPID Using GA

Transfer function of PID Controller: $K_p + \frac{K_i}{s+K_d s}$. @ @ @

Transfer function of FOPID Controller: $K_p + \frac{K_i}{s^\lambda} + K_d s^\mu$.

First the initial values for the tuning process are obtained using auto tuning available in the Simulink. Using these values as the initial point and providing the upper limit to the GA program already available in the Simulink, the best values are obtained, by iterating upon these values to find best fit for control variables. We have chosen the upper limit as [2,22,60]/[3,1,1,0,1]. Now the obtained values are put in the above transfer functions to get an optimized graph (Table 2).

3 Result and Discussion

3.1 Comparing Error Tracking

The error signal generated by the system using PID and FOPID controllers respectively, is tracked as shown in the graph above (Fig. 3).

3.2 Comparing Output Plot

See Table 3.

We obtained the final output step response of the system as shown by the graph (Fig. 4.), and the time domain parameters can be inferred as Table 4.

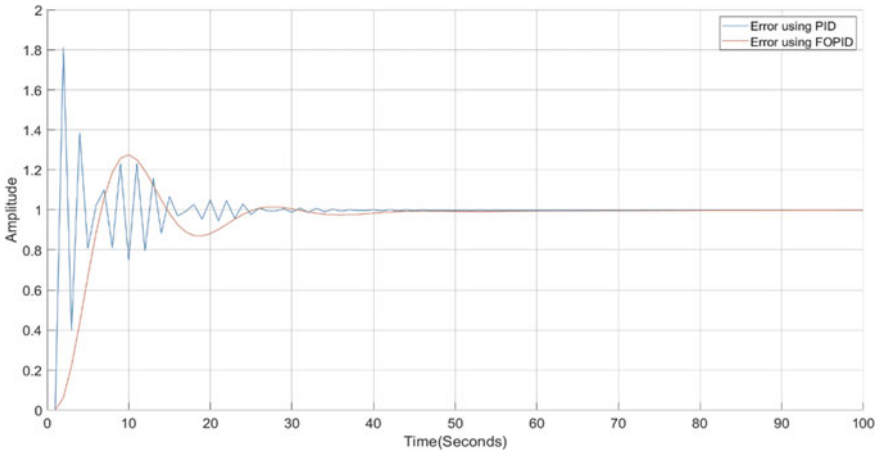


Fig. 3 Tracking feedback error using PID/FOPID Controllers

Table 3 Control variables obtained using genetic algorithm

Controller variable	Tuned PID controller using GA	Tuned FOPID controller using GA
K_p	1.99	3
K_i	21.99	1
K_d	59.99	1
λ	–	0
μ	–	1

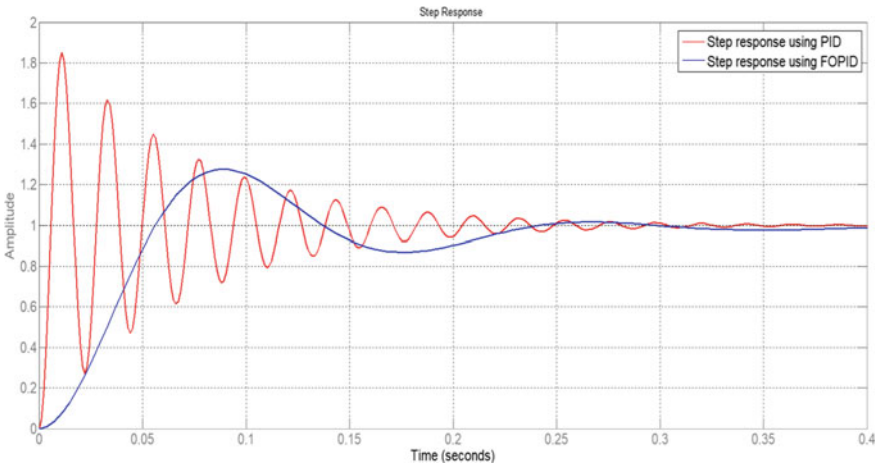


Fig. 4 Step response of PID/FOPID tuned with GA for pacemaker

Table 4 Comparison between time domain parameters

Parameter	Conventional PID controller	Tuned PID controller using GA	Tuned FOPID controller using GA
Steady state value (V)	20	1	1
Settling time (s)	3.8	0.26	0.23
Overshoot value (V)	76	1.85	1.30
Peak time (s)	1.5	0.01	0.09
Overshoot percentage (%)	52	85	30

4 Conclusion

The optimization technique Genetic Algorithm is used as a design method for determining the optimal parameters of PID and FOPID controllers accordingly to minimize time domain criterion. Comparison is done between the capabilities of PID and FOPID controllers for heart rate control of a Cardiac Pacemaker. We observe that the overshoot percentage has reduced from 52 to 30% and settling time has reduced to from 3.8 s to 0.23 s for the FOPID controller, and a more optimized output is obtained, for heart rate control of a Cardiac Pacemaker, tuned using Genetic Algorithm. Thus, we can conclude the FOPID is better over the conventional PID controller, and the optimization technique successfully tuned it.

References

1. Suseno EW, Ma'arif A (2021) Tuning of PID controller parameters with genetic algorithm method on DC motor. *Int J Rob Control Syst* 1(1):41–53
2. Acharya D, Das DK (2021) Swarm optimization approach to design PID controller for artificially ventilated human respiratory system. *Comput Methods Programs Biomed* 198:105776
3. Khan P, Khan Y, Kumar S (2020) Tracking and stabilization of heart-rate using pacemaker with FOF-PID controller in secured medical cyber-physical system. In: 2020 international conference on COMMunication Systems & NETworkS (COMSNETS). IEEE, pp 658–661
4. Mukhtar A, Mukhtar F (2020) Adaptive fractional order PID controller design of a pacemaker based on heart rate control strategy. *Int Res J Eng Technol (IRJET)* 1681–1685
5. Ibrahim MA, Mahmood AK, Sultan NS (2019) Optimal PID controller of a brushless DC motor using genetic algorithm. *Int J Pow Elec Dri Syst.* 2088(8694):8694. ISSN 2088-8694
6. Dey R, Sengupta A, Dey N, Mondal U (2019) Correction to: system identification and speed control of a DC motor using plug-in repetitive controller in the presence of periodic disturbance. In: International conference on communication, devices and networking. Springer, Singapore, pp C1–C1

7. Kumar H, Kumar R, Yadav J, Rani A, Singh V (2016) Genetic algorithm based PID controller for blood pressure and cardiac output regulation. In: 2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES). IEEE, pp 1–6
8. Govind KRA, Sekhar RA (2014) Design of a novel PID controller for cardiac pacemaker. In: 2014 international conference on advances in green energy (ICAGE). IEEE, pp 82–87

Finite Element Method Based Determination of Non-linear Inductances of Three Phase Induction Motor



Mohammed Nasir Ansari

Abstract Accurate parameter determination of any type of motor is an important aspect that further provides accurate motor performances. Equivalent circuits are widely used and are convenient means of computing the performances of three-phase induction motors. Amongst various equivalent circuit parameters, inductances are associated with the core material non-linear B-H curve, whose operating point shifts with the motor load. This results in motor inductances becoming non-linear with the variation of load on the motor. The method presented in this work determines the non-linear inductances of the three-phase induction motor implemented on finite element method (FEM) developing the two-dimensional motor model in FEM package ANSYS Maxwell Electronics Desktop. Mainly performing two tests on FEM, all the motor non-linear inductances are determined. For the sake of comparison of the results obtained, constant values of motor inductances are also obtained from the conventional tests on the induction motor i.e. stator winding D.C. resistance test, no-load test, and the blocked rotor test. These conventional tests for determining the constant values of motor inductances are also performed on the FEM package, ANSYS Maxwell. Finally, all the results are presented.

Keywords Finite element method · Equivalent circuits · Non-linear inductance measurement · Three-phase induction motors · Parameter estimation · Tests

1 Introduction

Electrical motors were invented more than a century ago, and it still finds widespread use in industries, as well as in day-to-day use in household appliances [1, 2]. There are several motor types such as D.C. motor, induction motor, synchronous motor, switched reluctance motor, permanent magnet motor, brushless D.C. motor, etc. available today. Amongst all these motors, induction motors have widespread usage in various electric drive applications [3]. Features of the induction motor such as

M. N. Ansari (✉)

Department of Electrical and Electronics Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, East Sikkim, P.O. Majhitar, Rangpo 737136, India
e-mail: nasir.ansari@smit.smu.edu.in

simple construction, low cost, reliability, ruggedness, and matured technology make it the popular choice in electric drive [4]. The use of the electrical equivalent circuit is a standard method of computing motor performances [5]. The accuracy of motor performances and characteristics obtained through the calculation of the equivalent circuit depends upon the preciseness with which its parameters are known. Therefore, there lies a scope in making the equivalent circuits accurate. The parameters of the motor equivalent circuits are generally obtained from the conventional tests, stator winding D.C. resistance tests, no-load tests, and the blocked rotor test. These parameters are constant in the entire range of motor operation [6–8].

Inductances of the motor are related to the non-linear core material B-H curve, and the point of operation on the curve shifts with the variation of load applied on the motor [9]. Therefore, the motor performance determined using non-linear parameters is more precise. Thus, the scope for including the non-linear B-H curve to the model of motor and thereby frame the non-linear model of induction motor and determine its parameters through some simplified tests. i.e. the synchronous speed test and the D.C. test as given in this work.

There has been a lot of work in the area of identifying the motor parameters accurately [10, 11]. Pedra et al. used a numerical method for finding the motor parameters out of the machine specification that the manufacturers generally provide [12]. Authors in [13] made use of an algorithm that determined the leakage inductance and resistance of the induction motor and again the magnetizing inductance in [14]. The finite element method (FEM) is also a widely used technique for finding the parameter [15, 16]. Authors in [17] have suggested a method and experimental tests that can be applied for the determination of three-phase induction motor non-linear inductances.

In this work by performing simulation tests on the two-dimensional motor model in the FEM package ANSYS Maxwell Electronics Desktop, the non-linear motor inductances are determined. Mainly two tests on FEM are performed i.e. synchronous speed test and the D.C. test, that result to determine all the motor non-linear inductances. For the sake of comparison of the results obtained from these two tests, constant values of motor inductances are also obtained from the conventional tests on the induction motor, i.e. stator winding D.C. resistance test, the no-load test, and the blocked rotor test. These conventional tests for determining the constant values of motor inductances are also performed on the FEM package, ANSYS Maxwell. It is finally seen that the motor inductances are significantly varying with the motor current and are thus considered to be non-linear.

Section 2 of the paper briefly presents the motor model of a three-phase induction motor. In Sect. 3 the FEM-based simulation for determining non-linear inductances using different tests is explained. Section 4 gives the results and discussions. In Sect. 5 the paper is concluded.

2 Motor Model of Three-Phase Induction Motor

Stators of three-phase induction motors are distributed windings, and the rotor has metal bars for the squirrel cage type or three-phase windings for the wound rotor type. Therefore, the induction motor of three-phase type can conceptually be shown as in Fig. 1. Both the stator windings and rotor windings have various inductances related with them as shown. Figure 2 shows the equivalent circuit of this motor [18].

The equivalent circuit has resistances and inductances as its parameters. The different inductances of this electrical equivalent circuit are dependent on the property of core material, i.e. B-H curve which is non-linear and varies with the amount of load on the motor or in other words, the motor current. In the next section, the simulation tests on the two-dimensional motor model in the FEM package ANSYS Maxwell Electronics Desktop for determining the non-linear motor inductances are undertaken.

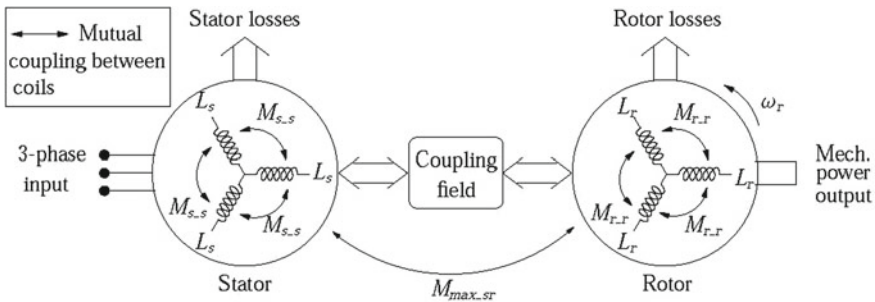


Fig. 1 Inductances in a three-phase induction motor

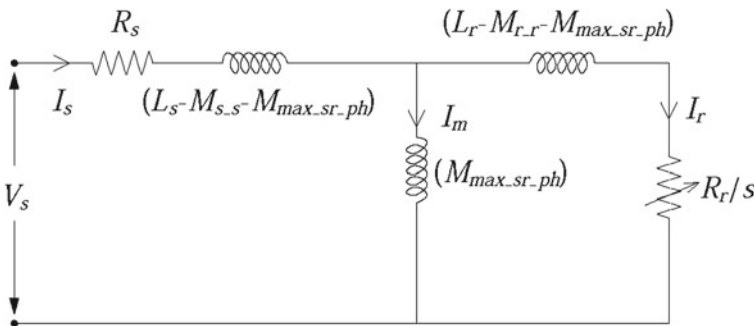


Fig. 2 Equivalent circuit of three-phase induction motor

3 FEM Based Simulation for Determining Non-linear Inductances

The property of the core material or its B-H curve is non-linear. The inductance is affected by the core material; thus the three-phase induction motor inductances display non-linearity. To obtain these non-linear inductances few types of simulation tests are performed on the three-phase induction motors that are presented here. The two-dimensional motor model in the FEM package ANSYS Maxwell Electronics Desktop of given specifications is used for simulation tests. The tests are mainly of two types, namely the synchronous speed test, and the D.C. test [17]. Conventional tests, namely the stator winding D.C. resistance test, no-load test, and the blocked rotor test for determining the constant values of motor inductances for comparison purposes are also performed on FEM package ANSYS Maxwell Electronics Desktop.

3.1 Synchronous Speed Test on FEM

The synchronous speed test captures magnetic property of the core i.e. its nature. Considering the stator and rotor to have the same core material used, the core material property can be captured simply through the data of stator voltage applied and its resulting stator current variation. The rotor effect of rotor voltage and current can be eliminated by keeping the rotor speed equal to its synchronous speed from settings in the simulation. The three-phase stator voltage at rated frequency applied to the motor model is increased in steps from zero voltage up to about 125% of rated voltage. The stator flux linkage curve is obtained as given by (1), from the data of the stator phase voltages and its corresponding values of stator phase currents.

$$\lambda_s = \frac{v_s - i_s r_s}{j\omega} \quad (1)$$

3.2 D.C. Test on FEM

D.C. test on the three-phase induction motor is performed to find the factors K_1 to K_4 linked with the several inductances. Any one stator phases is connected to the D.C. supply while the motor remains at standstill. The rise in stator D.C. current on switching to D.C. supply is recorded. The remaining phases of the stator are kept open-circuited. Stator D.C. current is given by (2) [17].

$$i_{s_dc} = \frac{V_{dc}(e^{ct}(dR_s - L_s R_r) \sinh(\frac{ta}{b}))}{aR_s} + \frac{V_{dc}((1 - e^{ct}) \cosh(\frac{ta}{b}))}{R_s} \quad (2)$$

where,

$$a = \sqrt{L_s^2 R_r^2 + R_s^2 d^2 - 2L_s R_s R_r d + 6R_s R_r M_{\max_sr_ph}}$$

$$b = 2L_r L_s - 3M_{\max_sr}^2 - 2L_s M_{r_r}$$

$$c = \frac{L_s R_r + dR_s}{2L_s M_{r_r} - 2L_r L_s + 3M_{\max_sr_ph}^2}$$

$$d = L_r - M_{r_r}$$

4 Result and Discussion

The results obtained from the simulation tests conducted on the three-phase induction motor implemented on FEM for determining its non-linear inductances are presented in this section. The specification of the motor used is given in Table 1. First, by conducting the conventional tests on the induction motor i.e. stator winding D.C. resistance test, no-load test, and the blocked rotor test the constant motor parameters are obtained and are given in Table 2.

The current rise in one of the stator phases, during the D.C. test, is shown in Fig. 3. The plot of current obtained from both FEM and calculated values is shown. The steady-state value of stator d.c. current for the phase winding switched to d.c. supply is seen to be about 16.5 A and is mainly limited by the winding resistance. However, the slope in the initial portion of the current is affected by the inductances of the winding.

Table 1 Specification of three-phase induction motor

Parameter	Value and unit	Parameter	Value and unit
Power	3.8 kW	Stator slot	36
Voltage	380 V	Rotor slot	26
Current	7.2 A	Motor length	150 mm
Frequency	50 Hz	Stator outside diameter	175 mm
Speed	1420 rpm	Stator inside diameter	121 mm
Poles	4	Rotor outside diameter	120.3 mm
Connection	Star	Rotor inside diameter	38 mm

Table 2 Constant parameters of three-phase induction motor

Parameter	Value and unit (Ω)	Parameter	Value and unit (Ω)
Stator resistance	0.8	Rotor resistance	0.2
Stator leakage inductance	1.5	Rotor leakage inductance	1.5
Magnetizing inductance	13.4		

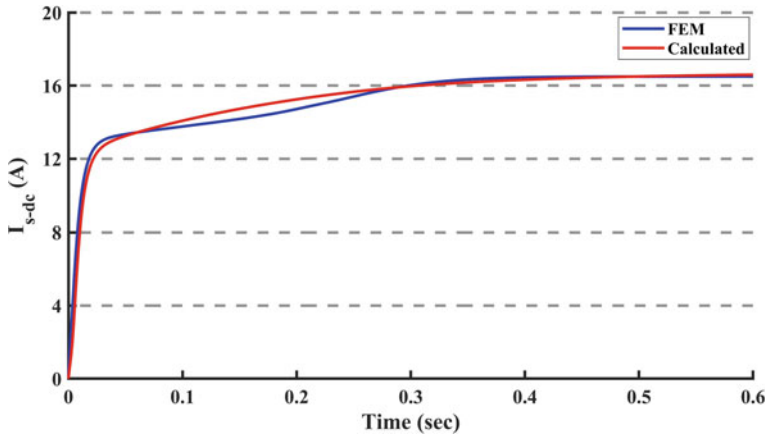


Fig. 3 Current rise in the stator phase during the D.C. test

The change in the magnetizing inductance, the stator leakage inductance, and the rotor leakage inductance with the motor current are shown in Figs. 4, 5, and 6 respectively. Essentially this variation in the inductances is due to the fact that the stator flux linkage of the motor itself is non-linear. It is seen from these figures that the inductances change widely with the variation of the motor current and are therefore non-linear. Three motor inductances shown have a similar variation and differ only in their magnitude.

5 Conclusion

Non-linear inductances of three-phase induction motor are determined using a two-dimensional motor model in the FEM package ANSYS Maxwell Electronics Desktop. Mainly two tests on FEM are performed. The non-linear nature of core material is determined from the synchronous speed test. This resulted to give L_{ss} in terms of I_m . All remaining motor inductances are represented in terms of L_{ss} using multiplication factors K_1 to K_4 . The multiplication factors are determined by

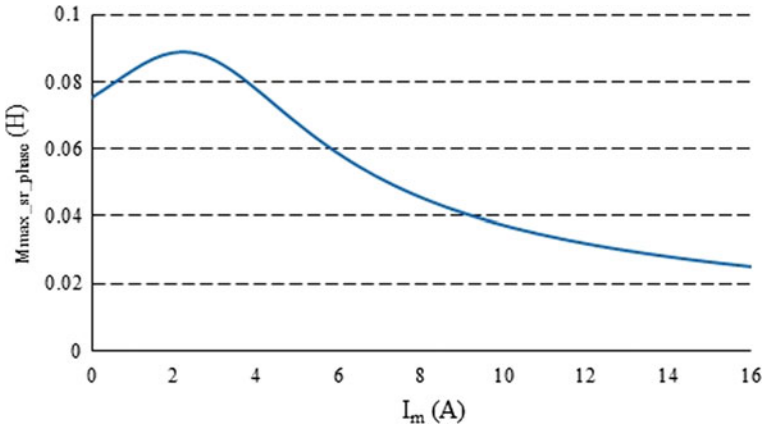


Fig. 4 Magnetising inductance variation

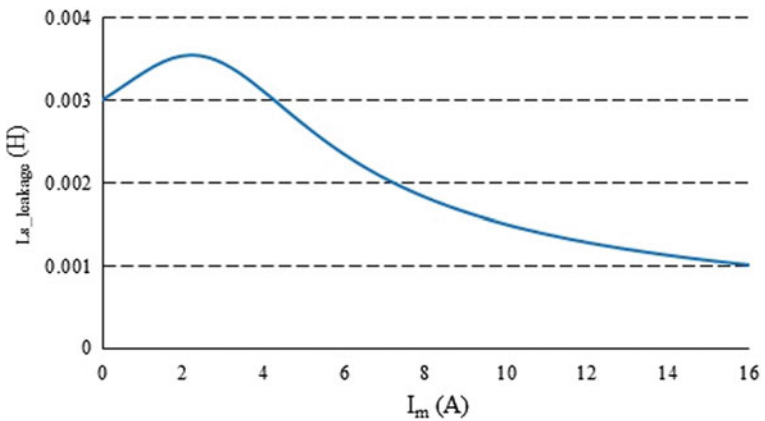


Fig. 5 Stator leakage inductance variation

comparing stator D.C. current obtained from FEM with that of the calculated stator D.C. current. The inductances are seen to be varying with the motor current and hence exhibit non-linear nature. These precise values of motor inductances when used in performance calculations or used in drive algorithms would give better results. For the sake of comparison of the results obtained, constant values of motor inductances are also obtained from the conventional tests on induction motor i.e. the stator winding D.C. resistance test, no-load test, and the blocked rotor test performed on FEM.

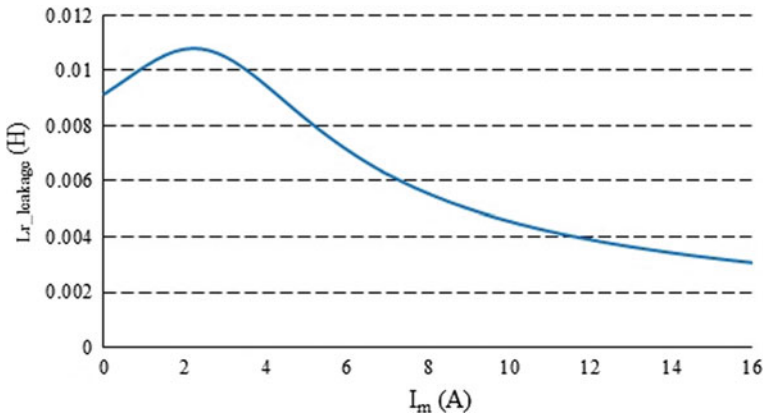


Fig. 6 Rotor leakage inductance variation

Acknowledgements The author thanks AICTE for supporting this work through the project grant under Research Promotion Scheme for the North East Region (RPS) with File No. 8-146/RIFD/RPS-NER/Policy-1/2018-19 and also Sikkim Manipal Institute of Technology.

References

1. Kirtley JL, Banerjee A, Englebretson S (2015) Motors for ship propulsion. *Proc IEEE* 103(12):2320–2332
2. Jiang X, Huang W, Cao R, Hao Z, Jiang W (2015) Electric drive system of dual-winding fault tolerant permanent-magnet motor for aerospace applications. *IEEE Trans Industr Electron* 62(12):7322–7330
3. Di Piazza M, Ragusa A, Vitale G (2011) An optimized feedback common mode active filter for vehicular induction motor drives. *IEEE Trans Power Electron* 26(11):3153–3162
4. de Lacerda Costa AS, Bastos RR, da Costa Paolinelli S, Nau SL, Valle RM, de Jesus Cardoso Filho B (2016) Characterization of electrical steels for high-speed induction motors applications: going beyond the common practices. *IEEE Trans Ind Appl* 52(2): 1350–1358
5. Ba X, Guo Y, Zhu J, Zhang C (2018) An equivalent circuit model for predicting the core loss in a clawpole permanent magnet motor with soft magnetic composite core. *IEEE Trans Magn* 54(11):1–6
6. Bentounsi A, Djeghloud H, Benalla H, Birem T, Amiar H (2011) Computer-aided teaching using matlab/simulink for enhancing an im course with laboratory tests. *IEEE Trans Educ* 54(3):479–491
7. Khambadkone A, Holtz J (1991) Vector-controlled induction motor drive with a self-commissioning scheme. *IEEE Trans Industr Electron* 38(5):322–327
8. Umans SD, Fitzgerald, Kingsley (2020) *Electric machinery*. Seventh Edition, McGraw Hill Education (India) Private Limited
9. Ansari M, Dalal A, Kumar P (2015) Analytical method for accurate determination of nonlinear magnetization curve of induction motor. In: XIIth IEEE INDICON 2015, New Delhi, India
10. Zaky M, Khater M, Shokralla S, Yasin H (2009) Wide-speed range estimation with online parameter identification schemes of sensorless induction motor drives. *IEEE Trans Ind Electron* 56(5):1699–1707

11. Kojooyan-Jafari H, Monjo L, Crocoles F, Pedra J (2014) Parameter estimation of wound-rotor induction motors from transient measurements. *IEEE Trans Energy Convers* 29(2):300–308
12. Pedra J, Corcoles F (2004) Estimation of induction motor double-cage model parameters from manufacturer data. *IEEE Trans Energy Convers* 19(2):310–317
13. Boglietti A, Cavagnino A, Lazzari M (2011) Computational algorithms for induction motor equivalent circuit parameter determination Part I: resistances and leakage reactances. *IEEE Trans Ind Electron* 58(9):3723–3733
14. Boglietti A, Cavagnino A, Lazzari M (2011) Computational algorithms for induction motor equivalent circuit parameter determination part ii: Skin effect and magnetizing characteristics. *IEEE Trans Ind Electron* 58(9):3734–3740
15. Bae D, Kim D, Jung H-K, Hahn SY, Koh CS (1997) Determination of induction motor parameters by using neural network based on fem results. *IEEE Trans Magnet* 33(2): 1924–1927
16. Zhang X, Ho S, Fu W (2014) Fast algorithm to obtain the torque characteristics with respect to load angle of synchronous machines using finite element method. *IEEE Trans Magnet* 50(11):1–4
17. Ansari MN, Dalal A, Kumar P (2018) A method for determining nonlinear inductances of electrical equivalent circuit for three-phase induction motor. *Electr Power Compon Syst* 46(4):379–390
18. Ansari M, Dalal A, Kumar P (2014, December) A generalized method for developing steady state models for different electric motors. In: 2014 IEEE international conference on power electronics, drives and energy systems (PEDES), Mumbai, India, pp 1–6

Design and Comparison of RC, RC-PD and RC-PID Controller Architectures in a Servo-Motor System



Litisha Mohapatra, Nishant Gupta, Rijhi Dey, Ujjwal Mondal, and Rudra Sankar Dhar

Abstract This work presents the different approaches by which repetitive control can be implemented to achieve better rejection of periodic disturbances present in the system. An overall analysis and comparison have been performed together with various controllers. The servo system experimental results are included to illustrate the impact of RC acting with a real-time plant. Plug-in RC (RC-PD—Repetitive Controller with Proportional and Derivative controller, RC-PID—Repetitive Controller with Proportional, Integral and Derivative controller) architecture combines the features of both repetitive controller and the controllers it is used in combination. Real-world applications need a stable output in result to the provided input when subjected to unwanted disturbances, thus by usage of RC-PD and RC-PID the intention is to minimize those fluctuations present in signal and get the desired output with reference to the desired application.

Keywords RC · RC-PD · RC-PID · Servo-motor · PID controller

1 Introduction

Tracking and rejection of periodic signals can be challenging in various engineering applications.

One of the most common approaches used for rejection of such disturbances is the internal model principle [1]. This model learns from the previous state's error

L. Mohapatra · N. Gupta (✉) · R. Dey

Department of Electronics and Communication, Sikkim Manipal Institute of Technology, Sikkim Manipal University, East Sikkim, Majitar, Rangpo 737136, India

R. Dey

e-mail: rijhi.d@smit.smu.edu.in

U. Mondal

Instrumentation Engineering, Applied Physics, University of Calcutta, Kolkata, India

R. S. Dhar

Department of Electronics and Communication Engineering, National Institute of Technology Mizoram, Aizawl, India

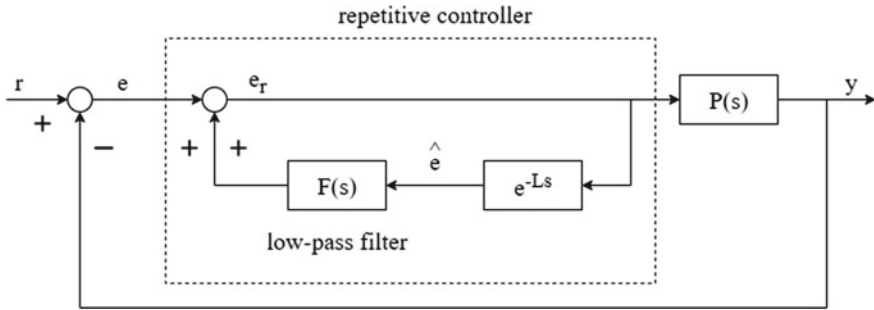


Fig. 1 Plug-in RC model

repeating in a cycle, from there it generates a control signal to refine the tracking output very close to the reference signal given as input to the system. PID has been the most popular controller for wide applications, but it also has some limitations like it is not flexible in terms of the parametric changes [2]. Many times PI controller is preferred instead of PID as value of D is not taken for ensuring system's stability and thus to overcome this problem, some adopt PID along with plug-in RC architectures (Fig. 1).

Internal model principle (IMP) addresses the rejection of periodic disturbances by implementation of plug-in RC architectures, helping users achieve the best possible output from combined usage of these controllers. RC has provided users with optimal results in case of robot control, accurate position control of piezoelectric actuators, optical disk of hard drives, underwater applications [3, 4], etc.

This paper investigates the rejection performance for periodic disturbances in RC and different plug-in RC architectures. The merits and demerits of each control strategy will be observed and reviewed for a DC servo system plant.

The drawbacks of PID controller are visible in its tuning complexity. If the number of harmonic frequencies is effectively considered, then a huge number of parameters will have to be designed. Repetitive control is a control scheme in which the objective is to remove errors due to the fundamental and harmonics of the periodic inputs [5]. The design of it is based on internal model principle which ensures a perfect tracking in steady state when the generator for various frequency modes of the periodic disturbance is included in our control loop. Here, as a controller a PD and PID along with a plug-in repetitive controller have been used. Combining design method of PID and RC system shows significantly high performance in several areas. PID control acts better where noise and a high degree of error is present but for it the tuning can get complicated whereas RC gets tuned relatively easily despite being highly sensitive to noise and high degree of error [6]. Therefore, merits of PD, PID and repetitive controller get implemented in this work.

2 Modeling of Plant

A servo-motor plant has been used to investigate the performance of RC, RC-PD and RC-PID. A motor is a machine which employs conversion of electrical energy into mechanical energy to operate. Servo systems are widely used in numerous position control industrial applications due to their advantages such as saving of energy, high power-to-weight ratio, less cost, simple structure and ease of maintenance [7].

The transfer function of stabilized servo-motor used in the simulation to observe the velocity response as output is [8]

$$G(s) = \frac{34,313,156.2}{s^2 + 742.32s + 202,662.4} \tag{1}$$

3 Methodology

Here, a set point is taken as input which subtracts the feedback signal and gets fed to the repetitive controller block as error signal, and here, the delayed version of the error signal is passed through our low pass filter where cutoff frequency of the periodic disturbance is matched and finally rejected repetitively. From there, it is fed to the respective controller, and from there, a control signal goes to the transfer function of our plant model which gets added with the periodic disturbance. This cycle repeats until the disturbance have been minimized (Fig. 2).

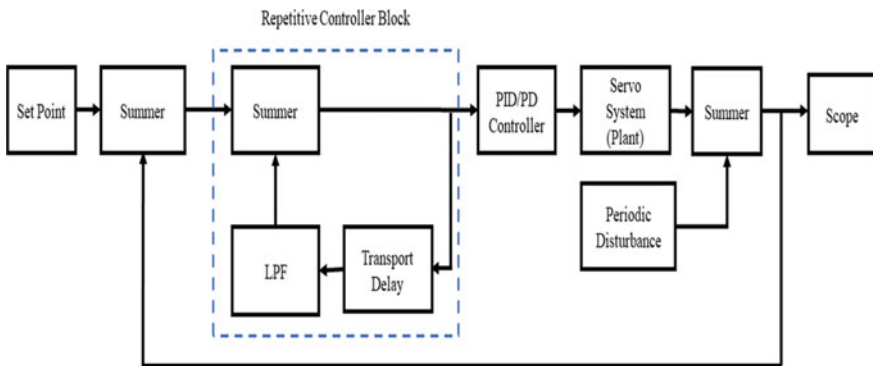


Fig. 2 Block diagram representation of methodology used

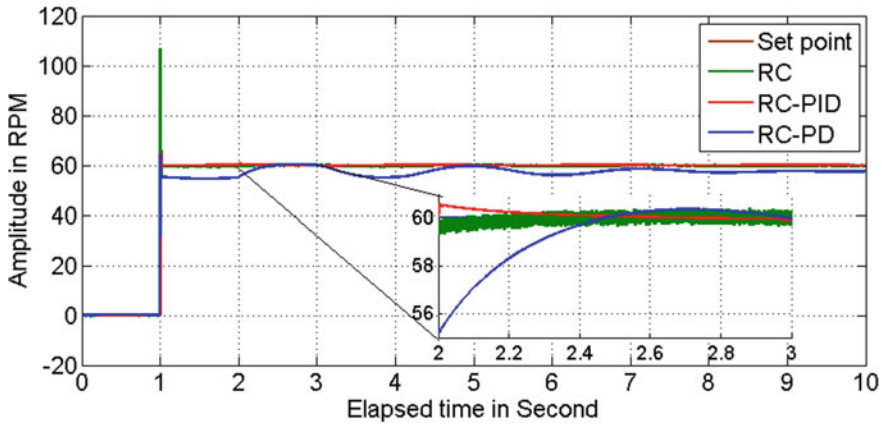


Fig. 3 Simulation results comparing RC, RC-PID and RC-PD

4 Results and Discussion

4.1 Comparing Output Plot

The steady-state error value which can be observed from the above graph shows clearly that both RC and RC-PD have significantly higher degree of error, especially at the beginning range of 0–1 s due to the presence of an overshoot in response. Comparatively, RC-PID has showed the least amount of error among all others (Fig. 3).

By comparison, it is found that RC-PID is best as it has increased gain due to P, less steady-state error due to I controller, faster response, less overshoot and settling time due to D. Whereas in RC alone the observation showed massive overshoot and in RC-PD, it showed a larger settling time and steady-state error due to lack of I controller. Also, in RC-PD due to lack of I controller, significant amount of steady-state error has been observed.

4.2 Comparing Error Tracking

The set point is at amplitude value 60. In time period 0 to 1 s, a huge overshoot is observed in the output waveform of RC from there onwards it stays around 60, whereas RC-PD has a less overshoot compared to RC but it comprises of few oscillations which would not be desirable for a real-time application. Finally, the response of RC-PID has come out to be the smoothest and most stable as the oscillation/disturbance is the least compared to all other configurations (Fig. 4).

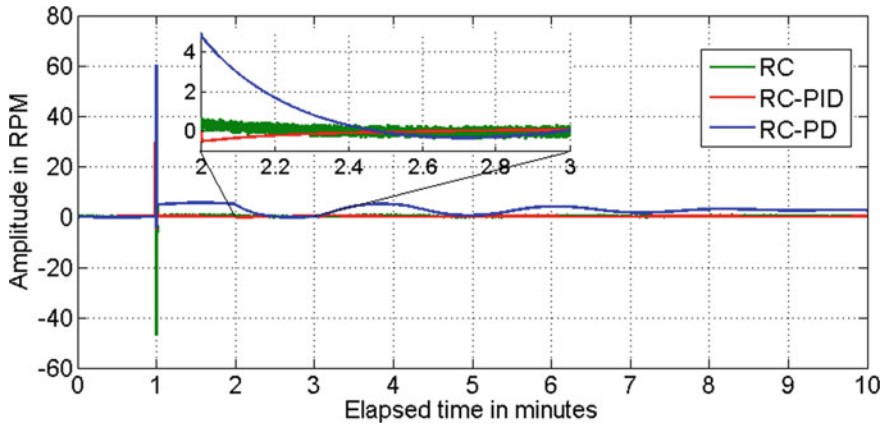


Fig. 4 Steady-state error value comparison of RC, RC-PD and RC-PID

5 Conclusion

This paper presents the rejection performance comparison of RC, RC-PD and RC-PID controllers. The transient response and difference in steady-state error have been discussed. Overall, RC-PID provided better rejection performance compared to the other plug-in modes. This happened due to presence of PID with RC which resulted in less rise time, faster response and minimum overshoot. The output of RC had a significantly high overshoot at the beginning and RC-PD had large settling time comparatively with few fluctuations at start.

With RC-PID, the desired output velocity response of the servo-motor used came out as the best compared to RC alone and RC-PD while being subjected to periodic disturbances.

References

1. Dey R, Sengupta A, Dey N, Mondal U (2019) System identification and speed control of a DC motor using plug-in repetitive controller in the presence of periodic disturbance. In: International conference on communication, devices and networking. Springer, Singapore, pp 433–438
2. Mondal U, Sengupta A, Sinha SM, Das B, Ghosh P (2012) Finite dimensional repetitive controller for identified model of a DC servo system. In: 2012 1st international conference on power and energy in NERIST (ICPEN). IEEE, pp 1–4
3. Kurniawan E, Afandi MI (2017) Repetitive control system for tracking and rejection of multiple periodic signals. In: 2017 international conference on robotics, automation and sciences (ICORAS). IEEE, pp 1–5
4. Costa Castelló R, Olm Miras JM, Vargas H, Ramos Fuentes GA (2012) An educational approach to the internal model principle for periodic signals. *Int J Innov Comput Inf Control* 8(8):5591–5606

5. Flores JV, Pereira LFA, Bonan G, Coutinho DF, da Silva JMG Jr (2016) A systematic approach for robust repetitive controller design. *Control Eng Pract* 54:214–222
6. Steinbuch M, Weiland S, Singh T (2007) Design of noise and period-time robust high-order repetitive control, with application to optical storage. *Automatica* 43(12):2086–2095
7. Thanigaivel G, Vijayakarhick M, Sathishbabu S (2018) New approach of repetitive control strategy for DC motor. *Int J Pure Appl Math* 119(16):3453–3465
8. Tsai M-C, Yao W-S (2002) Design of a plug-in type repetitive controller for periodic inputs. *IEEE Trans Control Syst Technol* 10(4):547–555

Effective Tone Control Circuit Using Proteus



R. B. Dharanidaran and Bappaditya Roy

Abstract Music production or signal processing is ignoring the hardware implementation and completely doing it on the software side. The main issue with this is the output speed and is reduced than the hardware, and it is dependent on the software processing speed. Also, we lose the original grunginess of the signal, an example of this is olden days retro songs in which you can hear and differentiate between all the different instruments used in the song. But in modern days, it is difficult to differentiate the instruments mainly because of the use of controlling and balancing tone on the software side and also the use of compression algorithms which completely reduces the quality of the original sound.

Keywords Bass · Treble · Volume · Preamplifier · Balance

1 Introduction

There is a well-known tone control circuit called Baxandall. It is a universal choice of audio amplifier manufacturers due to its simplicity [1–3]. It is difficult to control tone using only two controls, namely bass and treble control. It can be modified for flexibility for user desired output. A choice to expand the quantity of controls is made, and the field turns out to be totally open, the most clearest improvement being to have each control influencing the level of a restricted band of frequencies. Such circuits are clearly substantially more intricate than the essential Baxandall type and require a cautious plan to forestall over-the-top interaction between the controls. This project describes a modification to Baxandall circuit which increases versatility while maintaining its simplicity. The objective of this paper is to simulate the working of tone control circuit using Proteus, add basic controls: volume, bass, treble controls. Add preamplifier circuit, bass effect control, treble effect control, tone balance control circuit. Simulate and analyse the result.

R. B. Dharanidaran · B. Roy (✉)
VIT-AP University, Amaravati, India
e-mail: bappaditya.roy@vitap.ac.in

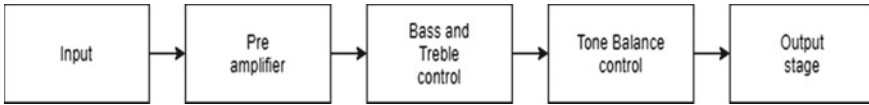


Fig. 1 Block diagram

2 Block Diagram

This circuit has been divided into five stages, namely input, preamplifier, bass and treble control, tone balance control and output stage. The preamplifier stage consists of preamplification and volume control knob. The bass and treble control stage consists of bass knob, treble knob, bass effect knob, treble effect knob and fade knob. Tone balance control stage consists of a balance knob and an effect fade knob (Fig. 1).

3 Input and Preamplifier Stage

Figure 2 shows the input stage and preamplifier stage. The amplifier used here is a high gain preamplifier circuit. The circuit of high gain preamplifier is work around a low power NPN semiconductor which is here designed in like common emitter mode utilizing self-bias or emitter bias. This biasing technique is utilized in the fact that it produces superb operation point stability. The volume knob is a potentiometer of 5 k ohm which is placed in such a way that if the potentiometer is high the output

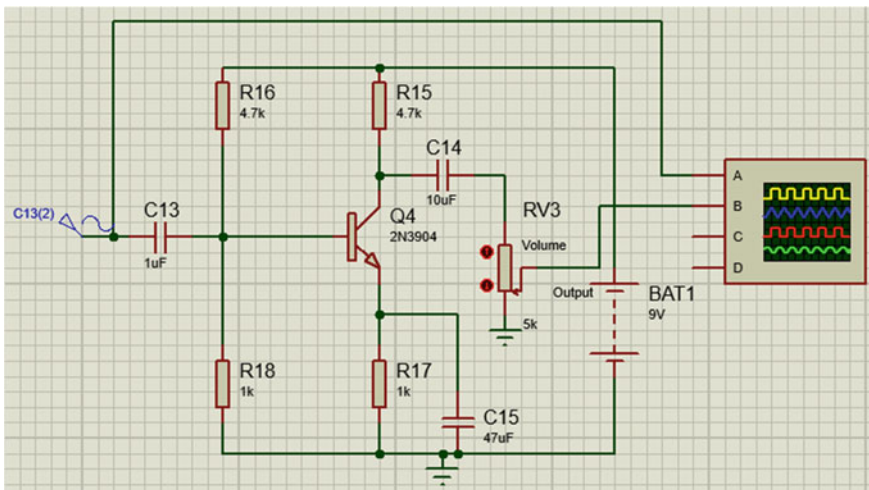
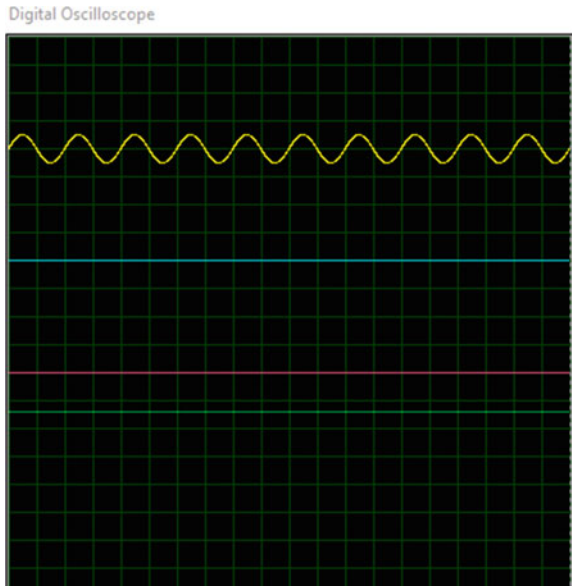


Fig. 2 Input and preamplifier circuit

Fig. 3 Volume at low



from the high gain amplifier will pass through and if the potentiometer is low the output from the high gain amplifier won't pass through since the output is directly grounded. C14 and C15 act as coupling capacitors to separate the AC signal from DC biasing voltage. This ensures that the bias condition by the additional amplifier stage is not affected by the DC voltage. $Gain = R_c/R_e$, $R_c = 4.7\text{ k}$, $R_e = 1\text{ k}$, $Gain = 4700/1000 = 4.7$ (Figs. 3 and 4).

4 Tone Control Stage

Tone control permits audience members to change sound; however, they would prefer. It additionally enables them to compensate for recording insufficiencies, hearing problems, room acoustics or inadequacies with playback gear. For instance, older people with hearing issues might need to increase the loudness of high pitched sounds which are difficult for them to hear. Tone control is additionally used to change a sound signal during recording. For example, if the acoustics of the recording site cause it to assimilate a few frequencies more than others, tone control can be utilized to enhance or "boost" the frequencies inside the room. In Fig. 5, both bass and treble control are set to maximum boost. Capacitors C2 and C4 become inactive. C3 and R4 act as low-pass filters. The capacitance of C3 is low. So the high frequency(241.14Ω) is passed through, and mid-frequency($4,823\Omega$) amplitude is decreased. And from the treble direction, C1 has reactance about 2411.4Ω at 20 kHz and $2,411,439\Omega$ at 20 Hz so it blocks low frequencies, and bypasses high frequency(2411.4Ω) and

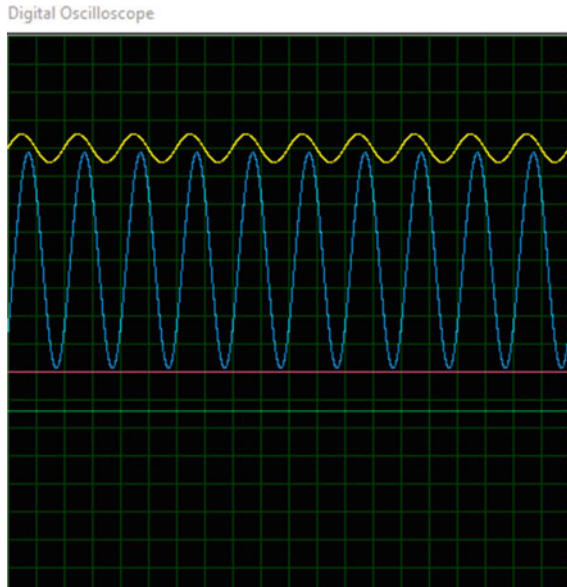


Fig. 4 Volume at high

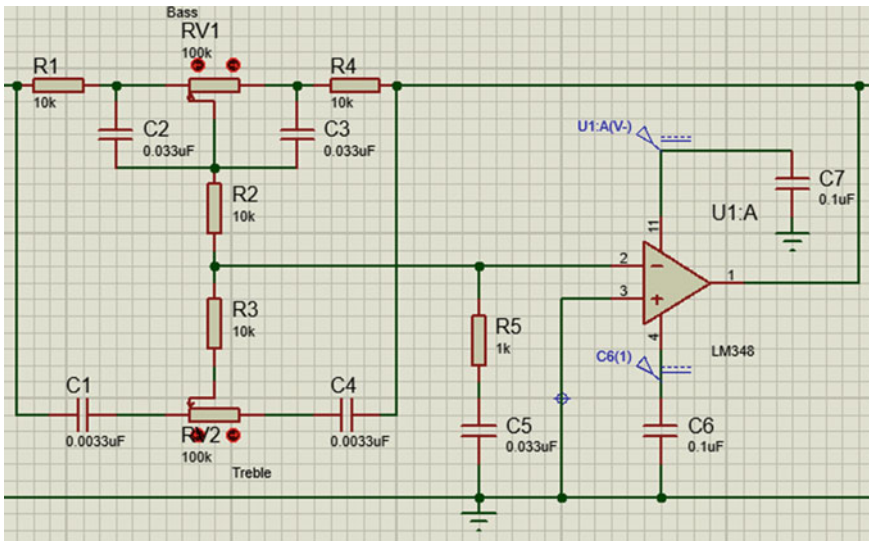


Fig. 5 Tone control with bass and treble boost

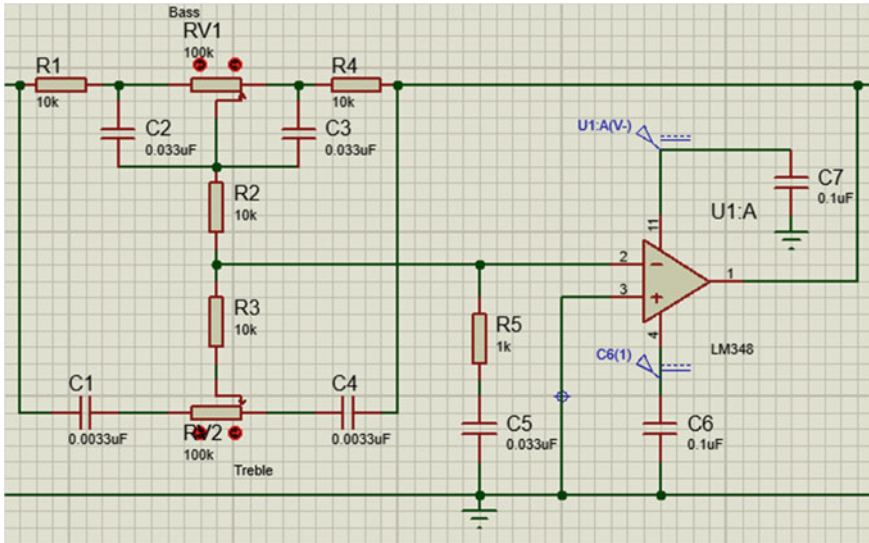


Fig. 6 Tone control with bass and treble cut

Table 1 Capacitive reactance for tone control

Capacitors	Frequencies			
	20 Hz	100 Hz	1 kHz	20 kHz
C1, C4 (0.0033uF)	2,411,439Ω	482,288Ω	48,229Ω	2411.4Ω
C2, C3 (0.033F)	241,144Ω	48,229Ω	4,823Ω	241.14Ω

reduced mid-frequency(48,229Ω). In whole, the output from this will be amplified high and low frequencies and very less mid-frequency is reduced. In Fig. 6, both bass and treble control are set to minimum cut. C3 and C1 are inactive because of a short circuit. The input goes to resistor R1. Capacitor C2 and resistor R4 form a high-pass filter. C2 has a reactance of about 4823Ω at 1 kHz so it passes high(241.14Ω) and mid-frequencies. Then the high frequency bypasses C4. We then left with mid-frequency which is then fed to the op amp for amplification. The output from this configuration would be mid-frequency. Capacitive reactance Table 1 can be seen.

5 Bass Effect Control Knob

Figures 7 and 9 represent the circuit working at extremes. When both bass and treble are set low, the output will be reduced low frequency and reduced high frequency; i.e. mid-frequency is unchanged which is then fed into the negative feedback op amp for

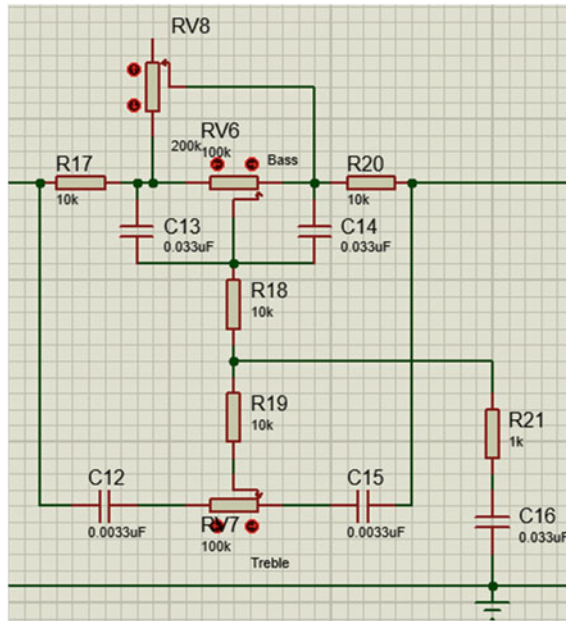


Fig. 7 Rv8 set low

boost. And also when the bass effect control is set low nothing modified in the low-frequency range other than the low frequency reduced by RV6 potentiometer. The changes when the bass effect control changed are shown below. Figure 9 represents the changes to the bass effect control RV8 potentiometer. The following results are shown below. Now the bass (low frequency) is unaltered; i.e. the original signal low frequency is reflected at the output. This happens because the bass effect control is set high. So basically what it does is how much bass you want to affect (Figs. 8 and 10).

6 Bass and Treble Effect Control in Tone Control Circuit

This circuit overcomes the drawback of Baxandall negative feedback tone control circuit [1]. And also it can fade between original signal and modified signal. Overall it performs better than M.V. Thomas’s addition to the Baxandall circuit with the widest range of controls [2]. My version solves the problem in addition to it you can fade the original signal frequency with the output frequency. Different users have different tastes in tone so if you want the original signal without modification you can do that also (Fig. 11).

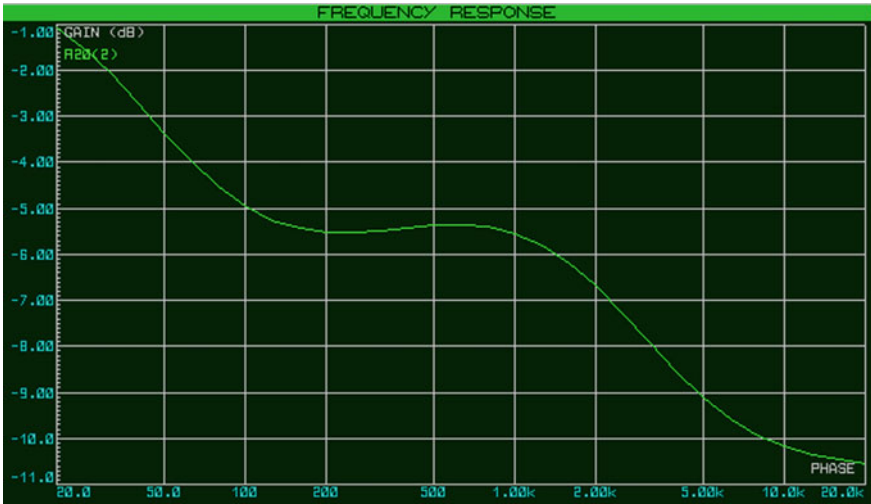


Fig. 8 Bass effect knob set low

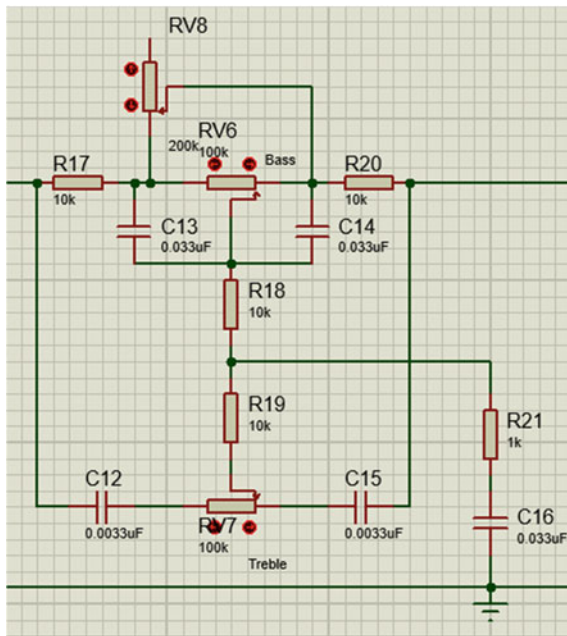


Fig. 9 Rv8 set high

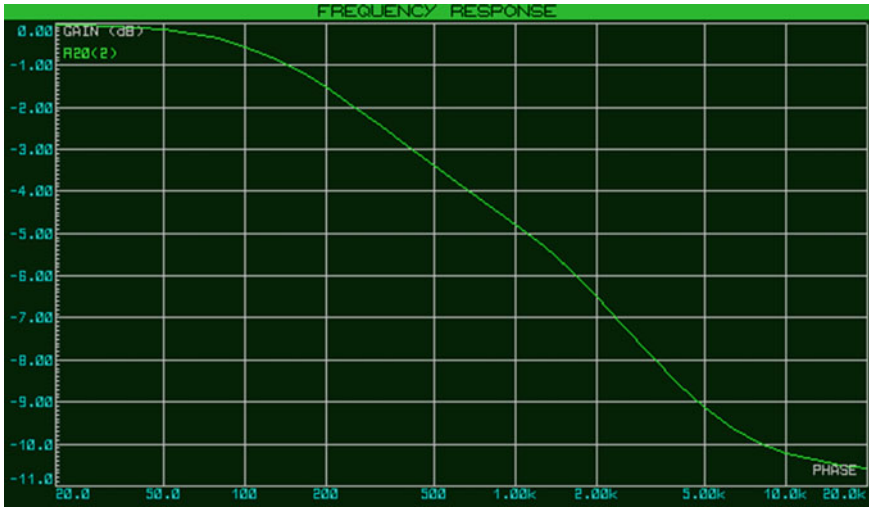
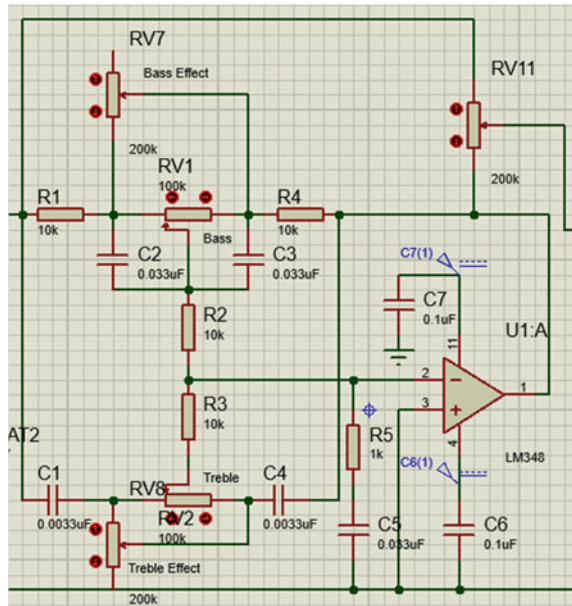


Fig. 10 Bass effect knob set high

Fig. 11 Bass effect knob set high



7 Tone Balance Control Circuit

Above is the circuit diagram for tone balance circuit. As you can see this acts as a variable capacitor circuit in which the capacitance can vary 3700pF to 7400pF.

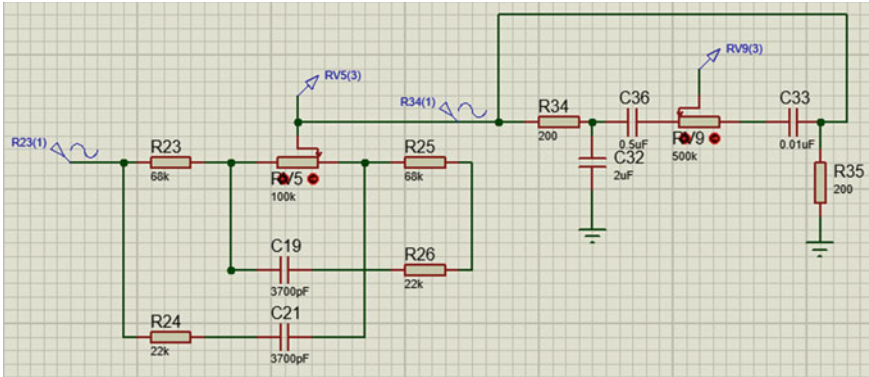


Fig. 12 Bass effect knob set high

The resistors R25 and R26 act as low resistance, respectively, to potentiometer RV5 (when it is greater than 90 k ohm). R25 and R26 values can be changed for the user desired output. The main reason to use this is in the event that the bass is initially excessively loud and the treble is excessively low, ordinary bass cut and treble boost might be applied anyway this eliminates a lot of the extreme bass, gives an excessive amount of extreme treble which is unuseful and still leaves the bass too strong and the overall treble too weak and vice versa, when the bass is initially extremely low and the treble is excessively low. These impacts are all the more frequent yet not perpetually discovered when the program source is on older or less expensive gramophone records, or a radio program. In R. Ambler’s tone balance control circuit if the bass is strong and treble is weak, normal bass cut and treble boost may apply and vice versa [3]. This circuit does better in high fidelity equipment. If the user wants the full control of the frequency range for example if he does not want the bass, he can completely remove it. This circuit can also be used in other applications like complete bass cut and treble cut. Main problem he faced is when using a tone control circuit either the bass or treble is generally too strong or too weak. So he produces a circuit which reduces this effect and smooths out the signal. My version does the same thing but in addition to that you have much more freedom to choose and can also act as a filter. For example, if you are in a concert hall and you don’t want the bass or treble you can cut it out separately instead of a combined circuit as in the Baxandall tone control circuit (Fig. 12).

8 Summary/Conclusion

The above results and analysis are true to my knowledge. As you saw above the controls, you can add are limitless. Some controls are better than previously existing work. You can create numerous controls to simulate the software side audio/signal

manipulation like echo, reverbs and so on. It is more like a modular tone control circuit where we can add any number of functionalities.

References:

1. Baxandall PJ (1952) Negative-feedback tone control. *Wirel World* pp 402–405
2. Thomas MV (1974) Baxandall tone control revisited. *Wirel World* pp 341–343
3. Ambler R (1970) Tone-balance control. *Wirel World* pp 124–126

Tetra Bot—A Multi-purpose Robot



B. Sricharan, K. Hima Bindu, A. Laasya Lata, and D. Ajitha

Abstract In our busy life, self-care can be one of the first things that are often neglected from our to-do list. Self-care is vital for our physical and mental health. Without adequate self-care, humans are less likely to be the best possible version of them which indeed affects their personal and professional life. A perfect way to recharge mind and body is to stay hydrated that safeguards our body from a minor health issue that is caused by not paying much attention to our self-care. In this paper, we have designed a multi-purpose robot which is a solution to all the small activities that are being ignored in our day-to-day life. The bot includes major domains like embedded systems, WeMoS D1 microcontroller, Internet of things (IOT) and image processing. It performs tasks such as cutting grass in the garden, acts as a dustbin using ultrasonic sensor, serving water when needed and it also reminds to carry umbrella using rain sensor, when it is about to rain. The bot uses Image processing through Haar cascade algorithms for face detection and sends information to the database and MIT app inverter. For the moment of robot, mechanical parts such as the center shaft, side shaft, spur and worm gear are used. It is a completely autonomous bot.

Keywords Rain sensor · Ultrasonic sensor · Embedded systems · WeMoS D1 · Haar cascade algorithms · Internet of things (IOT) and image processing

1 Introduction

Robotics technology effects on every area of life, both at work and home. Robotics can improve people's lives and job environments by increasing productivity and

B. Sricharan (✉) · K. Hima Bindu · A. Laasya Lata
Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India
e-mail: bachusricharan1@gmail.com

Present Address:

D. Ajitha
School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology (VIT),
Vellore 632014, Tamil Nadu, India

protection while also providing better service [1]. In large-scale industrial sectors, robotics is now a primary engine of productivity and versatility [2].

This paper focuses on a multi-purpose robot named tetra bot. Tetra bot performs tasks such as cutting grass in the garden, serving water when needed and it also reminds us to carry an umbrella when it is about to rain. The bot uses image processing for face detection and sends information to the database which serves water occasionally whenever it is needed. The robot has multiple input methods and output is designed based on inputs given. The input methods include audio visual input, temperature sensor, rain sensor, and ultrasonic sensor is used for intake of inputs. These inputs are given to the WeMoS D1 microcontroller board through IOT and through direct connections [3]. The inputs are sent to the cloud and then they are received by the microcontroller using Google firebase. Once the inputs are received by the microcontroller, it takes them as commands and gives the assigned output.

Any computer with any form of built-in sensor can capture and transmit data across a network without manual interference can be the “thing” in IoT. The embedded technology in the object enables them to communicate with internal states and the external environment, which in turn contributes to the process of decision-making [4]. The IoT is used to connect devices and systems which be made up of smart machines interacting and communicating with other machines, environments, objects and sensor network technologies. As a result, very large in size data are being generated, stored, and that data is being processed into useful actions.

In tetra bot, IOT and image processing are the major domains used. In image processing, Haar cascade algorithms are used for face detection [5]. It is a method based on machine learning where a cascade function is learned from a lot of positive and negative pictures. It is then used to identify the human face, Arduino IDE, Python and MIT app inverter make robot fully autonomous robot. This paper discusses the “tetra bot” of which we have built the prototype. Section 2 of this paper discusses the components used and the assembling of those components to form the tetra bot. Section 3 gives an idea about the working of the bot. The results are presented in Sect. 4, and the paper is concluded in Sect. 5.

2 Design of Tetra Bot

A bot is an integration of various hardware and software components. To design the tetra bot, the following hardware and software components are required.

2.1 *Hardware Components Required*

1. Side shaft motors
2. Li battery
3. Center shaft motor

4. Gears
5. L clamps
6. WeMoS D1
7. NodeMCU
8. Rain sensor
9. Ultrasonic sensor
10. Motor driver
11. Relay

2.2 *Software*

2.2.1 **Arduino IDE**

Arduino Integrated Development Environment (IDE) is an open-source community. This app is connected to the hardware of the Arduino, which requires the user to upload the application to the Arduino. Using special laws, the Arduino IDE supports C and C++ languages [6]. To dump the code into the Arduino, the Arduino board is connected to the device via USB.

2.2.2 **Python**

Python is a real-world programming that is suitable for fast learning. Guido van Rossum is the creator of Python, the high-level programming language [7]. In tetra bot, it features facial recognition that allows multiple individuals to be recognized based on their facial characteristics. The facial recognition code is written using Python. Python provides a wide variety of benefits over other programming languages with its powerful process automation functionality, such as increasing management capabilities that lead to improved speed for most applications and application usability.

2.2.3 **MIT App Inventor**

MIT app inventor, software which is an open-source online service, is used to create Android mobile applications. In this software, the user needs an email account to login and create apps. The software/website provides a graphical interface like the drag and drop option to create mobile applications. First, the user has to develop the layout of the design in provided screen, and then in the block section coding/function of the layout is developed according to the requirement of a user which is also a graphical interface. So, the user does not need to worry about coding knowledge. Because of this graphical interface, the user can develop apps very easily.

2.3 Construction

The body of the tetra bot is made by using cardboard. Wheels are attached at the bottom for movement. All the sensors are placed inside the cardboard. The front view robot is shown in Fig. 1. The details of the organization of various components in the tetra robot are as follows.

- Two side shaft motors with speed 200 rpm for locomotion as they have a good torque and produce high speed.
- The platform for water bottles is placed at some height for the comfortable placement of water bottles.
- Warm and spur mechanisms are used for opening the door of the places where a water bottle and an umbrella are present.

2.3.1 Base Wheels

Chassis is a hexagonal shape. The pointed corner is chosen as the front part of the bot and base. Wheels are arranged accordingly.

2.3.2 Disk Wheels

Two motors, four wheels and two dummy shafts (for better traction) are used in the robot locomotion.



Fig. 1 Front view of the tetra bot

2.3.3 Disk

Attached the walls along the circumference of the disk for support.

2.3.4 Top Doors

To lift top doors, a rack and pinion mechanism is used, and worm gear is kept to the motor and a wood piece is attached to the spur which helps the doors.

Open and spur are adjusted to worm and as worm rotates, spur rotates accordingly and the door opens.

3 Implementation

To build the bot, we require the integration of various kinds of sensors and boards. WeMoS D1 is the control unit of the tetra robot. All the sensors, motors are interfaced to WeMos D1 as shown in Fig. 2.

3.1 WeMos D1

WeMoS D1 is an ESP8266 12E-based WiFi development board shown in Fig. 3. The hardware looks identical to the Arduino Uno board. By using boards master, it can be used in Arduino programs. The D1 Mini is extremely powerful because the Arduino platform is cheap, WIFI-enabled, and completely compatible. In this, to have the D1 Mini functioning as an Arduino module, the ESP8266 library and board manager will be added [8]. The D1 Mini would operate as a web server, enabling the board to communicate with any WIFI-connected computer and to wirelessly control its pins. It plays a vital role, as it controls and receives the instructions to produce the result. All the sensors, motors, cloud and MIT app inventor were connected to the WeMoS D1.

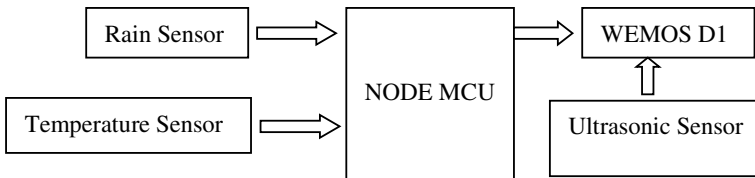


Fig. 2 Interfacing sensors to WeMoS D1

Fig. 3 WeMoS D1-ESP8266

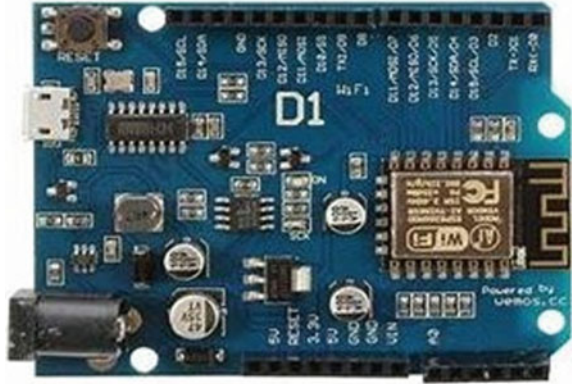


Fig. 4 NodeMCU



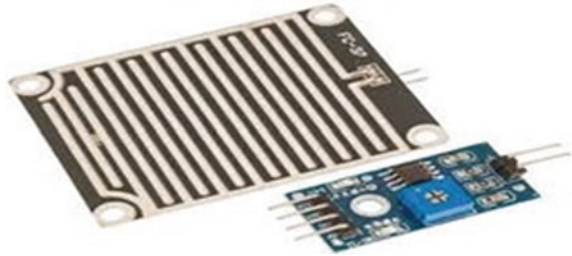
3.2 NodeMCU

NodeMCU is an open-source IoT platform that is low-cost. “Two words are combined into the name “NodeMCU.” The firmware was originally developed as a companion project to the popular NodeMCU development modules based on ESP8266 [9] and the firmware can now be run on any ESP modules shown in Fig. 4. NodeMCU is used as interface between the rain sensor and WeMoS D1. The rain sensor connected to NodeMCU, the NodeMCU and WeMoS D1 was connected through the internet. So, the values of the rain sensor go to the WeMoS D1, and it controls and gives the output.

3.3 Rain Sensor

The type of switching system that is used to track rainfall is a rain sensor. It works like a switch and the working principle of this sensor is that the switch will normally be closed whenever there is rain, [10]. This board includes lines coated with nickel and it works on the principle of resistance as shown in Fig. 5. This sensor module

Fig. 5 Rain sensor



allows moisture to be measured by analog output pins and provides a digital output when the moisture threshold is exceeded. In tetra bot, the rain sensor is used to predict the rainfall and to give the suggestion to carry the umbrella.

3.4 Ultrasonic Sensor

Ultrasonic sensor uses ultrasonic waves to measure the distance between the source and the target. It has two transmitter and receiver elements as shown in Fig. 6. The ultrasonic waves are emitted by the transmitter and the receiver receives the reflected waves from the target [11]. To measure the distance, the time difference between the transmission and reception of the processing used.

Tetra bot ultrasonic sensor is used to get the information about the quantity accumulated in the dustbin area of the bot. It measures the space available in the dustbin and indicates its user to empty the bin if necessary.

Fig. 6 Ultrasonic sensor

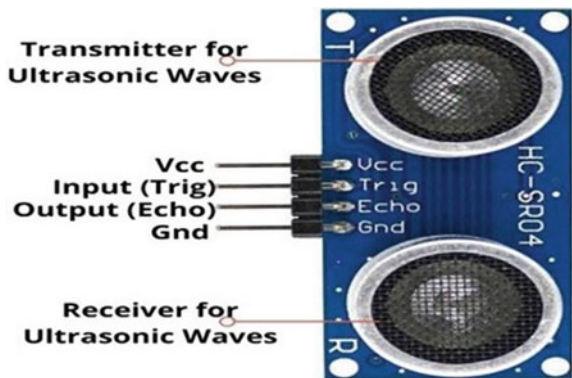


Fig. 7 Motor driver



3.5 Motor Driver

The motor driver acts as an interface between the WeMoS D1 (control circuit and motors. The control circuit operates by low current signals, whereas motors operate by high current. Thus, the motor driver shown in Fig. 7 takes the low- current control signal and turns it into a high-current signal appropriate for engine service.

3.6 4-Channel Relay Modules

An electrical switch that works on an electromagnet is a relay module shown in Fig. 8. The relay is the system that opens or closes the contacts to trigger the other electrical control to operate. For an allocated area, the relay module will define it is acceptable or unacceptable state and gives necessary circuit breaker commands to the affected area. This protects the system without any loss or damage of circuit.

Fig. 8 Relay



3.7 Temperature Sensor

A basic temperature sensor is a system that needs a thermocouple or RTDD to calculate the temperature using an electrical signal (resistance temperature detectors) [12]. The two metals in sensor prepare the thermocouple, producing the electrical voltage implicitly proportional to the change in temperature. A variable resistor RTD in sensor changes the resistance in a precise and linear manner; in this way, the temperature changes in indirect proportion. This sensor is used in our bot to know the information about the temperature and indicate the person to carry an umbrella when it's sunny outside.

3.8 Operation of Tetra Robot

Pictorial representation of tetra robot operation is shown in Fig. 9. When the robot detects the person, the camera placed inside it captures the picture of a particular

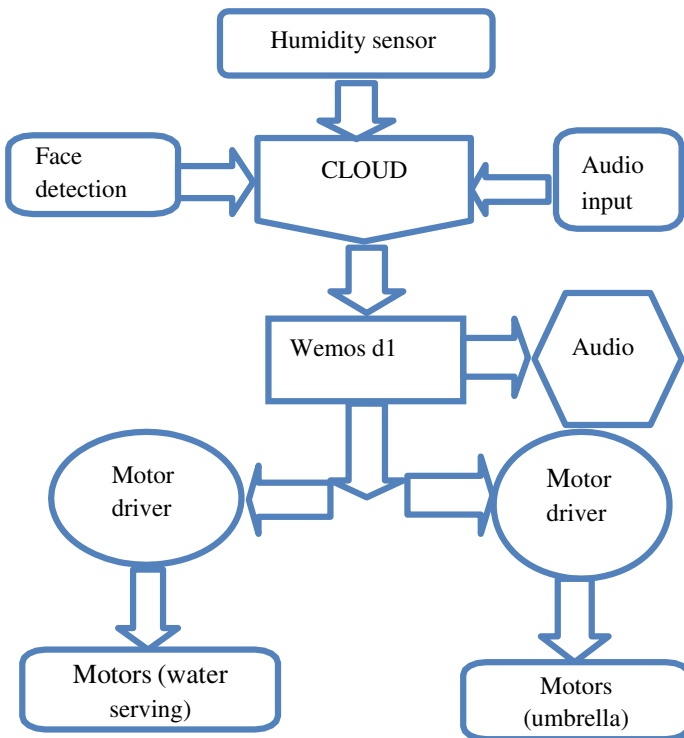


Fig. 9 Complete operation of tetra robot

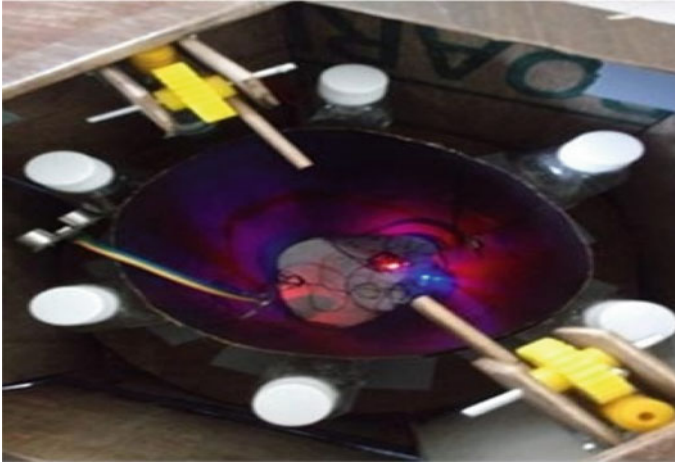


Fig. 10 Inside of the tetra bot

person. Then it matches with the database through the Haar cascade algorithm, and then the information of that person will be sent to the MIT app [13]. This app announces the name and asks “Do u need water?” If the person says yes, then this information will be transferred to the WeMoS D1 which is present in the robot.

WeMoS D1 processes the information to the spur and worm gear which opens the door of the robot. The person can now collect the water bottles placed in it. The bottles can be seen inside the robot Fig. 10.

After drinking the water, he can throw the water bottle into the bin which is present inside the bot; it can also work as a trash bin. The ultrasonic sensor is placed in that trash bin, if the bin is full; this data will be sent to the WeMoS D1. Then this information is transferred to MIT app through the cloud, which announces “that trash bin is filled, please empty it.” If the person does not feel thirsty at that moment, then processing will be stopped at the app itself. It does not send information to WeMoS D1.

3.9 Rain Sensor and Temperature Sensor Working

The rain sensor and temperature sensor are placed at the top of the house or roof; now, it checks the weather condition at that particular time and sends it to the NodeMCU which is placed along with sensors [14]. Now, this information in NodeMCU will be processed and send the data to the WeMoS D1 which is present in the robot. If there is a rainfall forecast on that day or high temperature (sunny day), then two operations will be done: The doors of the robot will be opened, in which the umbrella will be placed by default. Through WeMoS D1, information is transferred to the MIT app. Then it asks “Please carry an umbrella.” If a person requires it, then he can take

the umbrella. If the person does not need the umbrella, then that person can send a voice message that person does not need the umbrella to the MIT app. Then this information will be transferred to WeMoS D1 which closes the doors of the bot.

3.10 Haar Cascade Image Processing

Image processing is the process applied to images in order to get more information from the pictures which is useful for human interpretation; for easy transmission and storage; and for easy interpretation of pictures by autonomous machines [15]. Haar cascade image processing technique is one of the image processing techniques used to extract information from images.

Tetra bot uses the Haar cascade image processing system. It is the approach that Paul Viola and Michael Jones have suggested. The definition of machine learning is used here. By giving positive (images with faces) and negative images (images without faces), the cascade function is trained and other images can be identified later. The features of Haar are used in this method. Similar to the convolutionary kernel, each function is a single value obtained by subtracting the number of pixels from the sum of pixels under the black rectangle under the white rectangle [16].

The algorithm has four stages:

1. Selection
2. Creating images
3. Training
4. Classifiers.

First, the image is captured and store as pickle. After storing the image, a certain name is given to that image. The image is further trained through the algorithm and classified accordingly. When the camera captures the images, it checks with the images that are stored and identifies them from the database.

The bot uses library functions like:

```
import NumPy as np
import cv2
import pickle [17].
```

4 Result and Discussions

The rain sensor and temperature sensor connected to NodeMcu are taking up values and sending those values to WeMoS D1. Getting accurate values from those two sensors and respective task is perfect. Whenever it is raining, the value from NodeMcu is connected to wemosd1 through WiFi and now from WeMoS D1 to MIT app inventor, getting command to take umbrella. Whenever the box is filled the ultrasonic sensor detecting it accurately and sending to WeMoS D1 and the WeMoS D1 is sending instructions to MIT app inventor, that is box is filled, please clean it.

The image processing method we used here is not so accurate; in light condition, it is detecting, but during night and full light conditions, it is not working, and we can improve it by using advanced image processing methods for accurate results. The locomotion of the robot is good but still needs to improve, as the weight is more. The voice commands of the user are successfully transmitted via MIT app inventor to cloud and to WeMoS D1 and on reception; the desired operation successfully takes place. WeMoS D1 is successfully controlling and working efficiently.

The robot is a smart assistant which helps people in some real-life situations. It can help people stay hydrated which can fix most of the minor and sometimes even major health problems these days. We have successfully constructed a robot that can fix some of the basic everyday problems it also gives us an umbrella depending on the humidity levels. So, it is useful for old people, handicapped people who help them and it successfully generates voice commands, water, umbrella, and all sensors are working efficiently.

5 Conclusion and Future Scope

The construction and working of tetra bot is presented in detail. At present, the bot is being controlled by IoT but in later stages; this bot can be made into autonomous bot which is capable of checking water content of a person and calculates the time of recent water intake of a person and supplies water accordingly. It can also be linked to an app that reminds time to time weather report. MIT app works efficiently and it is an eco-friendly used app, and it is free of cost to download on Android phones it acts as a reminder, assistant for old people and handicapped.

Furthermore, this robot can be used as agriculture robot with few changes to make use for every farmer. The Haar cascade method for facial recognition is adaptable but it is not efficient. With the improvement of technology, changes in image processing methods can improve the efficiency. In future, all rain sensors can be removed in the advanced version of it, as it can be directly connected to the phone and then to WeMoS D1. In this way, the cost of construction will also be reduced. It is a very beneficial and useful robot and fully autonomous robot which works on the principle of IoT.

References

1. Qureshi MO, Syed RS (2014) The impact of robotics on employment and motivation of employees in the service sector, with special reference to health care. *Safety Health Work* 5(4):198–202
2. McLeay F et al (2021) Replaced by a robot: service implications in the age of the machine. *J Serv Res* 24(1):104–121
3. Dudhe PV et al (2017) Internet of things (IOT): an overview and its applications. In: 2017 International conference on energy, communication, data analytics and soft computing (ICECDS). IEEE
4. Ray S, Jin Y, Raychowdhury A (2016) The changing computing paradigm with internet of things: a tutorial introduction. *IEEE Des Test* 33(2):76–96
5. Berger W (2018) Deep learning Haar cascade explained. Will Berger
6. Plaza P et al (2018) Arduino as an educational tool to introduce robotics. In: 2018 IEEE International conference on teaching, assessment, and learning for engineering (TALE). IEEE
7. Sharma A, Khan F, Sharma D, Gupta S (2020) Python: the programming language of future. *IJIRT* 6(12). ISSN: 2349-6002
8. Patnaikuni P, Dinkar R (2017) A comparative study of Arduino, Raspberry Pi and ESP8266 as IoT development board. *Int J Adv Res Comput Sci* 8(5)
9. Chanthakit S, Rattanapoka C (2018) Mqtt based air quality monitoring system using node MCU and node-red. In 2018 Seventh ICT International student project conference (ICT-ISPC). IEEE, pp 1–5
10. <https://www.elprocus.com/rain-sensor-working-and-itsapplications/>
11. Dimitrov A, Minchev D (2016) Ultrasonic sensor explorer. In: 2016 19th International symposium on electrical apparatus and technologies (SIELA). IEEE
12. Girbau D et al (2012) Passive wireless temperature sensor based on time-coded UWB chipless RFID tags. *IEEE Trans Microw Theory Tech* 60.11, pp. 3623–3632, (2012).
13. Tang D, Utsumi Y, Lao N (2019) PIC: a personal image classification webtool for high school students. Proceedings of the 2019 IJCAI Edu AI Workshop. IJCAI
14. Srivastava M, Kumar R (2019) An IoT based weather monitoring system using node MCU and fuzzy logic. In: International conference on computer networks and inventive communication technologies. Springer, Cham
15. Hegadi RS (2010) Image processing: research opportunities and challenges. National seminar on research in computers. Bharathiar University, Coimbatore, India
16. Pai, Kamlesh VKN et al (2018) Face recognition using convolutional neural networks. In: 2018 2nd International conference on trends in electronics and informatics (ICOEI). IEEE
17. Team (2015) Open CV developers. Face detection using Haar cascades. Tersedia http://docs.opencv.org/master/d7/d8b/tutorial_py_face_detection.Html

Getting Started with LPWAN: LoRa, Sigfox and NB-IoT



Shridhar Sharma

Abstract This paper gets an understanding of the challenges faced in implementing LPWAN technologies and the resources to solve the most challenging problems, which will be related to increasing range, increasing bandwidth and decreasing power consumption. This paper presents the latest versions of the documents required to get there. Understanding the network and security requirements while developing an engineering solution for industrial or domestic use is paramount. Each provides its own unique set of challenges that the developer and research have to adapt to. This paper presents a brief of these two core concepts and provides important references to delve deeper into the subject and implement the cheapest and most power-efficient system possible. This paper compares these technologies and presents to you in a way which will highlight important concepts while giving you the resources to find solutions for specific problems.

Keywords LoRaWAN · Sigfox · NB-IoT · Network architecture · Security architecture

1 Introduction

The first industrial revolution saw the advent of steam-powered production, the second was able to mass produce using electrical energy, the third saw greater efficiency using programmable logic controllers, now the fourth promises to take these power savings and reduction in manufacturing costs to greater heights. This is backed by Forbes IoT survey which states that there are going to 75 billion IoT devices by 2025 [1]. With such a strong desire for IoT connectivity by humans, this is the latest exciting field to be in. Industry 4.0, big data analytics and the potential employment of artificial intelligence to ease manufacturing process go hand in hand. The most optimal solution for this would be a workerless production called computer-integrated manufacturing (CIM) [2] has its own pitfalls and deeper embedded system

S. Sharma (✉)
Sikkim Manipal Institute of Technology, Rangpo, Sikkim, India
e-mail: writetoshridharsharma@gmail.com

integration requirements which can only be realistically implemented after sensors devices and actuators have been connected to each other and the internet. Industrial applications are not limited to manufacturing on the assembly line many companies including AWS, Beewise, Zenseio and WeBee are utilizing LPWAN for poultry, pig, cattle farming and beekeeping [3]. Some minor differences do exist while implementing this technology over the globe, e.g., the China LoRa Application Alliance (CLAA), led by ZTE, generally recommends the use of the 470–518 MHz band but the use of the 470–510 MHz band for radiometers [4]. These minor differences make it just a bit more difficult to study, but not very much. IoT devices were first introduced for supply chain and logistical tracking only, and it makes sense that these LPWAN devices will certainly find applications there. Companies such as Amosense and Sensolus are providing hardware, software and cloud solutions for asset tracking via IoT devices. More than 68% of the world's population is projected to live in urban cities by 2050 [5]. This increase in urban population is definitely going to require accurate, precise and real-time gathering of data to accurately model human requirements. This can only be achieved by deeper integration of IoT in our lives. The drawback of this integration will be the consumption of batteries by the device, which will require periodic replacement. This is going to incur extra operation costs. Ambient energy harvesting is one method which is being looked into to avoid this problem [6]. This is very important because these LPWAN technologies can only provide coverage in short distance [7].

2 LoRaWAN

LoRa and LoRaWAN are one of the more sought-after communication technologies for low-power long-range communication protocols for Internet of things (IoT) and machine-to-machine applications (M2M). LoRaWAN is networks which are low on resources, require a long battery life, high capacity and low device and deployment costs [8]. As LoRAWAN is promising in said areas its applications are far and wide including but not limited to utilities, trucking and logistics, monitoring things in both a domestic and industrial environment like trash receptacles, plumbing, and electrical utilities and even domesticated and feral animals.

Some preliminaries are required to have a better understanding of LoRaWAN. LoRa is a RF modulation scheme called chirp spread spectrum (CSS) using which LoRaWAN devices communicate with each other. LoRaWAN is the set of technologies which together let sensor nodes transmit datum from point A to point B. LoRaWAN has a star of star topology and includes three main components to transmit datum which are network servers, gateways and end nodes [8]. LoRaWAN is a MAC layer communication protocol which supports device-to-device network layer parameters including MAC layer commands, frame content, communication classes, data rates, network frequency management, back-end interfaces which let the user dissolve the network into interoperable nodes and firmware updates over the air [9].

2.1 Network Architecture

LoRaWAN networks are assembled in a star of star topology, where gateways pass messages between end devices and a central network server. The network server routes each packet from every device on the network to an application server.

LoRaWAN offers a high ground coverage of a hundred square kilometre due to its very high link budget which is greater than any technology on the market [10]. LoRa calculates link budget in terms of received power,

$$P_{RX} = P_{TX} + G_{SYSTEM} - L_{SYSTEM} - L_{CHANNEL} - M \quad (1)$$

where P_{RX} (dBm) is received power, P_{TX} (dBm) is transmitted power, G_{SYSTEM} (dB) is system gains such as those associated with directional antennas, L_{SYSTEM} is losses associated with the system, $L_{CHANNEL}$ is losses due to propagation channel, and M is fading margin.

According to Shannon–Hartley theorem,

$$C = B \times \log_2(1 + S/N) \quad (2)$$

where C is the channel capacity and is measured in bits per second (bps). B is the channel bandwidth in Hz and S/N is the signal-to-noise ratio. For spread-spectrum technologies, S/N is low and sometimes power density is below the noise floor. With a noise level $S/N \ll 1$, Shannon's equation becomes,

$$C/B = 1.433 \times S/N \quad (3)$$

$$C/B \approx S/N \quad (4)$$

$$N/S \approx B/C \quad (5)$$

This means that in order to transmit an error-free signal into a channel for a given signal-to-noise ratio we only need to implement a spread-spectrum operation that will increase bandwidth for higher frequencies and then conversely perform a reducing operation on the receiver side. LoRa uses chirp spread spectrum (CSS) in its devices to receive and transmit datum. Chirp stands for ‘‘Compressed High-Intensity Pulse’’; this is a very common modulation scheme used in radar and sonar, and chirp signals have a constant amplitude and variable frequency [11]. Choosing the right location to deploy a gateway for low-power WAN networks is critical. These sites are few in number and are therefore shared by several providers and are not just suited for low-power WAN networks but also other cellular networks. RFID is one of those communication methodologies which operate in the same frequency band as LoRa and provides a lot of interference. Some of the most common issues for a LoRaWAN system due to co-located transmitters are out-of-band blocking,

in-band-blocking, third-order inter-modulation and generation of out-of-band noise. Mitigation methods include radio frequency filters embedded in the design of the gateways and external cavity filters [12].

2.2 Security

LoRaWAN uses advanced encryption standard (AES) cryptography algorithms. Each LoRaWAN device has its own unique 128-bit AES AppKey and a 64-bit globally unique identifier (EUI-64) called DevUI. LoRaWAN devices are end-to-end encrypted while the packets are being transmitted from the end device to the application server [13]. LoRaWAN devices are recognized with their tag which at the very least has to consist of these value and always in this order SchemaID, JoinEUI, DevEUI and ProfileID, and these mandatory values can be followed by one or multiple extensions which are CheckSum, OwnerToken, SerNum and some other proprietary extensions, prefixed by a parameter key. The full size of the tag can be 128 characters where 48 characters are used by the mandatory information which leaves 80 characters for optional security extensions. Each device is on-boarded where it is associated with an owner and given permission to use a network [14].

Data in LoRaWAN communication is protected by two session keys which are encrypted by AES-Control (CTR) and a Message Integrity Code (MIC) computed with an AES-CMAC [13].

3 Sigfox

Sigfox is a subset of the LPWAN set of technologies and was patented and developed by a French startup in 2010. It is also characterized by consuming low power between 19-49 mA, sending low packets per day, only 140 and providing cellular connectivity in tens of square kilometres in rural areas but allows a max payload of only 8 bytes. It uses differential binary phase-shift keying (D-BPSK) at a fixed bandwidth of 100 Hz transmitted at 100bps in Europe and 600bps in the USA in two unlicensed spectrum which is 868 MHz and 916 MHz, respectively, and allows 140 messages per object per day. Sigfox offers 100bps Up-link and 600bps down-link. Sigfox uses cellular base stations which transmit data to and from end devices over the cellular network and connect these devices to the cloud. Sigfox devices start communication by sending three up-link packages on three random carrier frequencies. The duty cycle is restricted to 1% which limits transmission time for a Sigfox device to 36 s per hour [15]. Sigfox devices are used where the datum to be transmitted is low but has to be transmitted over longer distances. This lets Sigfox be a good product to be used for agricultural, hydropower activities that have to take place over very long distances where frequent human traversal is not economically feasible. Sigfox

includes radio configuration (RC) zones which greatly increase ease of deployment. Main features of Sigfox application include [16].

- Application integration ready
- Multiple RC zones
- Sigfox Monarch
- Extremely low CPU load
- No latency requirements
- Small memory footprint
- Utilities service provided.

The up-link and down-link frequencies for Sigfox vary according to RC zones. The down-link frequency lies between 866.3 MHz to 922.3 MHz, and for up-link, it is between 865.2 MHz to 932.3 MHz.

3.1 Network Architecture

Sigfox lets its devices communicate to the base station over radio this base station then transmits data over IP or cellular which depends on the user. The datum is then transferred to a cloud bases server using RESTful APIs.

Sigfox discourages the use of SNR to measure quality of signal (QoS) and instead uses received signal strength indicator (RSSI), receiver redundancy and radio configuration (RC) zones [17] (Tables 1 and 2).

Table 1 For zones RC1, RC3 and RC5 [17]

RSSI	Number of Base Stations	Link Quality Indicator
$-122 \text{ dBm} < \text{RSSI}$	3	Excellent
$-135 \text{ dBm} < \text{RSSI} \leq -122 \text{ dBm}$	3	Good
$-122 \text{ dBm} < \text{RSSI}$	1 or 2	Good
$-135 \text{ dBm} < \text{RSSI} \leq -122 \text{ dBm}$	1 or 2	Average
$\text{RSSI} \leq -135 \text{ dBm}$	any	Limit

Table 2 For zones RC2 and RC4

RSSI	Number of Base Stations	Link Quality Indicator
$-114 \text{ dBm} < \text{RSSI}$	3	Excellent
$-127 \text{ dBm} < \text{RSSI} \leq -114 \text{ dBm}$	3	Good
$-114 \text{ dBm} < \text{RSSI}$	1 or 2	Good
$-127 \text{ dBm} < \text{RSSI} \leq -114 \text{ dBm}$	1 or 2	Average
$\text{RSSI} \leq -127 \text{ dBm}$	Any	Limit

3.2 Security

A lot of papers are available which discuss the most probable security architecture Sigfox could be using but not a lot of them go very deep this could be because Sigfox is a proprietary LPWAN technology and the manufacturers do not want to give out information which could be detrimental to their agenda which is backed by the fact that the organization only provides details on how to build the devices and gives out information only about standard implementation paradigms and APIs, what is known is that Sigfox uses AES in Cipher-based message authentication code (CMAC) mode with a 12-bit sequence number which can be replayed back to the user equipment which results in a denial-of-service (DoS) attack [18]. Sigfox uses a porting authorization code while manufacturing to secure the device (PAC) [19, 20].

Narrow-band Internet of things (NB-IoT) is another low-power wide area network communication system developed by 3GPP. This is built for indoor datum transmission with a bandwidth of 180 kHz and uses orthogonal frequency division multiplexing (OFDM) for down-link at a 15 kHz tone spacing and single carrier frequency division multiplexing (SC-FDMA) for up-link communication, single tone, 15 kHz and 3.75 kHz tone spacing. NB-IoT provides in-band and guard-band LTE, 164 dB of coverage and a peak rate of 50kbps at down-link and 50kbps at up-link for multi-tone and 20kbps at up-link for single tone. NB-IoT provides a lower cost than eMTC, 164 dB of maximum coupling loss, receiver sensitivity of -141 dB, 10 plus years of battery life and support at least 50,000 devices per cell. NB-IoT introduces a new physical layer which means that it is not backwards compatible with LTE.

NB-IoT user guidelines specify these minimum requirements to achieve the best power consumption and roaming service continuity,

- Power save mode
- Extended discontinuous reception
- Paging
- Cell reception
- Support for extended coverage
- Power class
- Rate control mechanisms
- Home subscriber server
- Multi-frequency band.

3.3 Network Architecture

The NB-IoT system is based on the long-term evolution (LTE) system, and the underlying architecture is called and evolved packet system (EPS). EPS is an IP-based network. The architecture shown in Fig. 3 is the architecture for cellular IoT (CIoT) devices.

The user equipment (UE) is an NB-IoT device and consists of a subscriber identification module (SIM) card. The UE connects to a eNodeB (eNB) on the LTE-Uu protocol. User plane data is transferred through the S1-U interface to a serving gateway (SGW), whereas the user plane datum routes through the packet data network gateway (PGW) to a CIoT application server; at the same time, the control plane datum moves to the mobility management entity (MME) through a S1-MME interface. 3GPP provides another option where if user data is transferred through a control plane the service capability exposure function (SCEF) gives access to a RESTful API for data transfer [21].

Machine-to-machine communication devices will need to work in an indoor environment where they will be in the presence of other data transmission protocols like Wi-Fi, LTE, GPRS; it is critical that these devices do not affect receiver sensitivity and are also averse to it. Maximizing coverage while taking care of previously mentioned variables is of prime importance. Maximum coupling loss (MCL) is used to measure coverage performance. (Add more science and maths references about MCL to increase reference count and paper size.) MCL is calculated on the basis of these parameters.

- Transmitter power (dBm)
- Thermal noise density (dBm/Hz)
- Receiver noise figure (dB)
- Interference margin (dB)
- Occupied channel bandwidth (Hz)
- Effective noise power (dBm)
- Required signal-to-interference plus noise ratio (SINR)(dB)
- Receiver sensitivity (dBm)
- Receiver processing gain

Effective noise power = Thermal noise density + Receiver noise figure + Interference margin + 10 log (Occupied Channel Bandwidth), Receiver sensitivity = Effective noise power + Required SINR and MCL = Transmitter power – Receiver sensitivity + Receiver processing gain [22].

3.4 Security

NB-IoT also uses AES and secure key generation. It also has the advantage of the knowledge developed through the years as it is a long-term evolution (LTE)-based IoT protocol. When NB-IoT devices are communicating over radio, they are encrypted at the user end and or the control plane. NB-IoT provides four choices of encryption algorithms defined by a four-bit identifier starting from “0000” and ending at “0011” and are christened EEA0, EEA1, EEA2 and EEA3 all of these have a 128-bit input key except EEA0 which is not encrypted and is called the null ciphering algorithm, EEA1 is a SNOW 3G-based algorithm, EEA2 is AES-based and EEA3 is ZUC-based [23].

4 Conclusion

We have gone through a wide range of things which have to be taken care of before we begin to build our LPWAN network. These things are very important and are contributors to major cash considerations throughout the project. Encryption will always keep proving to be a hurdle, especially as customer base increases. These customers will need more devices or there will occur a period of time where the rate of increase of devices is slower than the requirement for them; that is, supply is less than demand. This is a very favourable condition for business but a very unfavourable one for engineering. This does not mean that engineering should try to slow down business, this means that engineering is a skill and process-based job which requires meticulous arrangement of time, human labour, human capital, which may manifest itself as either organic or inorganic, which is that it may be a sociological requirement which has put engineering into action or may be a requirement for more hardware, software, skills. As requirement for more encryption increases, the requirement for more processing cores will also increase. This increase will require concentrated regions of hardware implementations which will process this oncoming data, benefiting off from economies of scale. This will benefit technologies which run off of an LTE network as they already have this concentrated region of mass logic execution. This will tilt the balance in favour of SIGFOX, which makes the implementer of these technologies to think about the range which these devices will offer its customers. This range is clearly in favour of LoRa. But these devices are able to deliver the lowest number of messages per day. That is indeed the cost of range. This is a problematic situation for the implementer of these technologies (Table 3).

Table 3 Comparison of LPWAN technologies

	LoRA	NB-IoT	SIGFOX
Adaptive data rate	No	Yes	No
Bandwidth	100 Hz	250kHz,125 kHz	200 kHz
Maximum data rate	100 bps	50 kbps	200 kbps
Bidirectional	Half Duplex	Half Duplex	Half Duplex
Max payload length	12/8 bytes (UL/DL)	243 bytes	1600 bytes
Range	10 km	5 km	1 km
Modulation	BPSK	CSS	QPSK
Interference immunity	Very high	Very high	Low
Encryption	None	AES128	LTE encryption
Localization	RSSI	TDOA	No
Allow private network	No	Yes	No
Maximum messages/day	140(UL),4(DL)	Unlimited	Unlimited

Notes and Comments. This paper presented a developer and or researcher implementing LPWAN technologies a starting point to find holistic information about them. The LPWAN set of technologies find wide applications. The most promising characteristic of LPWAN is that it is not suited for just one use case but is wholly adaptable to any challenge it encounters. All these technologies provide a researcher ample opportunity to implement cutting edge solutions to domestic and industrial problems. LPWAN offers tremendous opportunity for research, engineering and employment as it has an expected compounded annual growth rate (CAGR) of 109% till 2023 [24]. With the onslaught of 5G, it will be likely that a hybrid of these two technologies comes to the forefront

References

1. Chaudhari BS, Zennaro M, Borkar S (2020) LPWAN technologies: emerging application characteristics, requirements, and design considerations. *Future Internet* 12(3)
2. Vaidya S, Ambad P, Bhosle S (2018) Industry 4.0—A glimpse. *Procedia Manuf* 20:233–238
3. Whitepaper (2020) The farming of tomorrow is already here: how LoRaWAN® technology supports smart agriculture and precise animal production
4. Xu Z, Ni J (2021) *J Phys: Conf Ser* 1871:012011
5. Al Homssi B, Al-Hourani A, Magowe K, Delaney J, Tom N, Ying J, Wolf H, Maselli S, Kandeepan S, Wang K, Gomez KM (2021) A framework for the design and deployment of large-scale LPWAN networks for smart cities applications. *IEEE Internet Things Mag* 4(1):53–59
6. Yuksel ME, Fidan H (2021) Energy aware system design for battery less LPWAN devices in IoT applications. *Ad Hoc Netw* 122:102625
7. Wang H, Liu Y, Wei Y, He Y, Tsang K-F, Lai LL, Lai CS (2020) LP-Index: explore the best practice of LPWAN technologies in smart city. In: 2020 IEEE International smart cities conference (ISC2), pp 1–5
8. Ertürk MA, Aydın MA, Büyükakkaşlar MT, Evirgen H (2019) A survey on LoRaWAN architecture, protocol and technologies. *Future Internet* 11(10):216
9. <https://lora-alliance.org/lorawan-for-developers/>
10. Devalal S, Karthikeyan A (2018) LoRa technology-an overview. In: 2018 Second international conference on electronics, communication and aerospace technology (ICECA), p 284–290

11. Mroue H, Nasser A, Parrein B, Hamrioui S, Mona-Cruz E, Rouyer G (2018) Analytical and simulation study for LoRa modulation. In: 2018 25th International conference on telecommunications (ICT), pp 655–659
12. Whitepaper (2021) LoRaWAN® gateways: radio coexistence issues and solutions. <https://lora-alliance.org/wp-content/uploads/2021/04/LA-White-Paper-LoRaWAN-Gateways-1.pdf>
13. Whitepaper (2017) LoRaWAN security whitepaper. https://lora-alliance.org/wp-content/uploads/2020/11/lorawan_security_whitepaper.pdf
14. Technical Resource (2020) TR005 LoRaWAN® device identification QR codes
15. Vejlggaard B, Lauridsen M, Nguyen H, Kovacs IZ, Mogensen P, Sorensen M (2017) Mads coverage and capacity analysis of Sigfox, LoRa, GPRS, and NB-IoT. In: 2017 IEEE 85th Vehicular technology conference (VTC Spring). IEEE
16. Technical Document (2021) How to build a Sigfox™ application with STM32CubeWL https://www.st.com/content/ccc/resource/technical/document/application_note/group1/21/ff/e3/84/9b/bf/41/e3/DM00699238/files/DM00699238.pdf/jcr:content/translations/en.DM00699238.pdf
17. Link Quality: general knowledge. <https://support.sigfox.com/docs/link-quality:general-knowledge>
18. Coman FL, Malarski KM, Petersen MN, Ruepp S (2019) Security issues in internet of things: vulnerability analysis of LoRaWAN, Sigfox and NB-IoT. In: 2019 Global IoT summit (GIoTS). IEEE
19. Sigfox Technical Overview
20. Secure Sigfox Ready devices: Recommendation guide
21. Technical Specification (2021) General packet radio service (GPRS) enhancements for evolved universal terrestrial radio access network (E-UTRAN) access (Release 17)
22. Technical Specification (2015) Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)
23. Technical Specification (2020) 3GPP system architecture evolution (SAE); security architecture (Release 16)
24. LPWAN market report 2020–2025. <https://iot-analytics.com/product/lpwan-market-report-2020-2025/>

Enabling Cognitive Radio in NOMA-Assisted Reconfigurable Intelligent Surfaces: Outage Performance Analysis



Arjun Chakravarthi Pogaku, Nhan Duc Nguyen, Anh-Tu Le,
and Dinh-Thuan Do

Abstract Non-orthogonal multiple access (NOMA) proves to be a prominent solution for huge device connectivity and enhances spectral efficacy. NOMA can handle several users in a single time and frequency slot. The superimposed signal transmitted from the transmitter to the receiver will be applied to successive interference cancelation (SIC) technique at the user to extract the desired user's signal. However in cognitive radio (CR) networks, the users in the secondary network will only be assigned with the unused spectrum not being used by the primary network users. Though the secondary users utilize the idle spectrum, there is a minimum probability of primary users being affected by interference from secondary users. On the other hand, reconfigurable intelligent surfaces (RIS) can control the propagation of electromagnetic waves in the radio environment, according to the position of the user. In this paper, we suggest a simple CR-enabled NOMA-assisted RIS system to examine the performance of the users in the secondary network, consisting of a base station, RIS device, and two NOMA users. We have derived the closed-form expressions for outage probability (OP) at both the users and compared their performance in the presence of CR networks. The results show that, in the presence of CR, the number of meta-surfaces in RIS can significantly enhance the outage performance of the users. The comparison of both users' performance has manifested that near users will have better outage performance compared to the far users.

Keywords Cognitive radio networks · Reconfigurable intelligent surfaces · Non-orthogonal multiple access

A. C. Pogaku (✉) · D.-T. Do

Department of Computer Science and Information Engineering, College of Information and Electrical Engineering, Asia University, Taichung 41354, Taiwan
e-mail: arjun.chakravarthip@gmail.com

N. D. Nguyen

Innovation Center, Van Lang University, Ho Chi Minh City, Vietnam
e-mail: nhan.nd@vlu.edu.vn

A.-T. Le

International Cooperation and Scientific Research Department, Van Lang University, Ho Chi Minh City, Viet Nam

1 Introduction

Cognitive radio (CR) networks have proved to have immense potential to serve fifth-generation (5G) communication networks. Understanding its ability to enhance the spectrum utilization, it is widely suggested to be deployed along the side with 5G networks for massive communication and data transmission. CR consists of primary users and secondary users performing the data transmission in the same licensed spectrum band [1]. Therefore, the presence of secondary users causes intervention to the primary users, and it can be limited to a certain level. The possible approach to limit or eliminate the intervention caused is that CR can automatically identify and utilize the channel without interference or selecting the channel under certain intervention temperatures [2]. Another approach toward this scenario is the utilizing non-orthogonal multiple access (NOMA). NOMA allocates numerous subscribers in single cluster, in the same frequency and time block by varying their power coefficients. On the receiver side, SIC is performed to obtain the required user signal from the transmitted superimposed signal.

Reconfigurable intelligent surfaces can control the radio transmission environment, which encourages the integration of RIS with 5G and sixth-generation (6G) wireless networks. RIS is made of a low-cost, electromagnetic material that can be controlled electrically and have unique wireless communication capabilities. RIS has numerous advantages like enabling a wide range of communications, response to full band, easily deployed on any surfaces, and most importantly, less impact by the receiver's noise. As a result, a lot of studies have been done on enhancing secure data transmission, improving spectral efficiency, and eliminating interference in [3–6].

In [7], the authors have studied the channel quality of a poor channel user in NOMA and CR-NOMA with fixed power allocation, in which CR-NOMA proves to provide a better transmission than traditional NOMA. In [8], the CR-NOMA system was developed with power constraints to improve the spectral efficiency and evaluate the outage performance by computing closed-form expressions. In [9], to enhance the user fairness of secondary users in the CR-NOMA system, a cooperative protocol was designed by the authors. To examine the performance of two NOMA users at the secondary destination, the exact expression of OP was derived. To enhance the spectral efficacy of the network, in [10], the authors have considered RIS-assisted CR system, solved the resource allocation problem, and derived the OP expressions to examine the proposed system efficiency. In [11], the authors have investigated the RIS-CR system to maximize the throughput of secondary users due to the existence of heavy disturbance from the primary network.

In [12], the authors have investigated the integration of RIS system with NOMA communication networks and analyzed its effect in the presence of hardware impairments. The results manifested the importance of number of meta-surfaces in the system. A similar system was considered in the [13] to analyze the bit error rate of the system performance of the system. Authors in [14] have thoroughly demonstrated the theoretical performance of the RIS system using the mathematical techniques derived in [15]. In [16], the authors have considered RIS-assisted multi-user code-

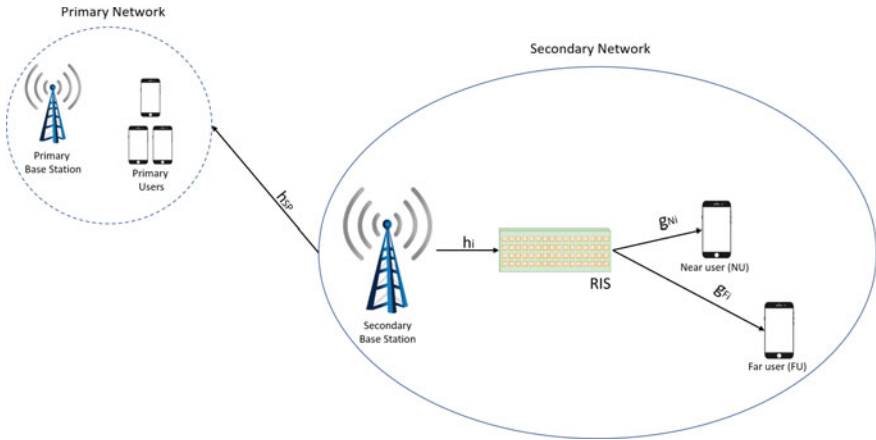


Fig. 1 System model of CR in NOMA-assisted RIS

domain NOMA system with multiple input multiple output antennas and studied the effect of phase-shift optimizations. In the presence of multiple clusters in NOMA network-aided RIS, the author in [17] identified that with sufficient increase in the number of RIS meta-surfaces, the interference caused due to multiple clusters can be minimized. In [18], the authors have considered a RIS-assisted NOMA network to investigate the system performance in the presence of perfect and imperfect SIC and derived the closed-form expressions for OP, ergodic capacity (EC). The authors have also compared the results with benchmark OMA scheme.

Motivated by the studies on CR, NOMA, and RIS, our paper proposed a CR-enabled NOMA network-aided RIS. We propose this system to examine and understand the performance of a NOMA system with two users aided by a RIS in the presence of cognitive radio. We analyze the performance of the system in terms of OP and throughput of the system.

2 System Model

This paper investigates the performance of CR-assisted NOMA networks with aid of the RIS system. The system consists of the primary network and secondary network, and there is no direct link between the primary base station (BS) and the receivers in the secondary network. Since NOMA serves multiple users in a single resource block, we consider two users, near user (NU) and far user (FU), with single antennas at both transmitter and receiver. N denotes the number of meta-surfaces present in the RIS, while h_{sp} denotes the channel coefficient from the primary BS to the secondary BS and assumed to follow Rayleigh distribution. The channel between the BS and RIS is assumed to be h_i , and channels between RIS to NU and FU are

assumed to be g_{Ni} and g_{Fi} , following baseband equivalent fading, respectively. The i in the channel denotes i -th element in the RIS. The channels are assumed to be independent and identical with flat and slow varying. The propagation of the signal is according to the NU and FU that follow the Rayleigh distribution with different scale parameters. The RIS can utilize the channel state information (CSI) to maximize the SNR receiving at the destinations. Therefore, considering the ideal case of CSI [19], the RIS is assumed to attain the CSI of these users (Fig. 1).

In the perspective of NOMA, the BS transmits superimposed signals, s_1 and s_2 , at NU and FU, by allocating higher power level, α_2 , to FU and lower power level, α_1 , to NU. The power allocation scheme must follow $\alpha_1 + \alpha_2 = 1$ and $\alpha_2 > \alpha_1$. Utilizing the principle of RIS, the received signals at NU and FU can be formulated by

$$Y_{NU} = \sum_{i=1}^N h_i e^{j\Phi_i} g_{Ni} \left(\sqrt{\alpha_1 P_s} s_1 + \sqrt{\alpha_2 P_s} s_2 \right) + n, \tag{1}$$

$$Y_{FU} = \sum_{i=1}^N h_i e^{j\Phi_i} g_{Fi} \left(\sqrt{\alpha_1 P_s} s_1 + \sqrt{\alpha_2 P_s} s_2 \right) + n, \tag{2}$$

where Φ_i denotes the adjustable phase induced by the i -th elements in RIS and n denotes the additive white Gaussian noise (AWGN) with variance N_0 . To make further computations easier, we denote $X_N = \sum_{i=1}^N h_i e^{j\Phi_i} g_{Ni}$ and $X_F = \sum_{i=1}^N h_i e^{j\Phi_i} g_{Fi}$.

In the perspective of CR networks, to limit the interference caused due to the secondary network, the power at the secondary network is limited as [20]

$$P_s \leq \left(\frac{P_P}{|h_{SP}|^2}, \overline{P_s} \right), \tag{3}$$

where P_P denotes the power of interference constraint at the primary network and ρ_P denotes the SNR of the interference constraint at the primary network. $\overline{P_s}$ denotes maximum average transmit power at the secondary network, and $\overline{\rho_s}$ denotes the SNR of maximum transmit power at the secondary network.

With consideration of FU signal as noise signal at NU according to the principle of NOMA, we write the signal to interference and noise ratio (SINR) to detect signal s_2 is given by Basar et al. [14]

$$\Psi_{NU,s_2} = \frac{|X_N|^2 \alpha_2 P_s}{N_0 + |X_N|^2 \alpha_1 P_s}, \tag{4}$$

where $\rho_s = \frac{P_s}{N_0}$ is the signal to noise (SNR) at the base station and $\rho_P = \frac{P_P}{N_0}$ is SNR at secondary network.

Therefore, we can rewrite SINR S_2 at NU as

$$\Psi_{NU,s_2} = \frac{|X_N|^2 \alpha_2 \rho_s}{1 + |X_N|^2 \alpha_1 \rho_s}. \tag{5}$$

After successfully performing successive interference cancellation at NU, the FU signal can be deleted, therefore, computing the SINR of NU as

$$\Psi_{\text{NU},s_1} = \frac{|X_N|^2 \alpha_2 P_s}{N_0} \quad (6)$$

$$\Psi_{\text{NU},s_1} = |X_N|^2 \alpha_2 \rho_s. \quad (7)$$

At FU, there is no requirement for performing SIC as it does not experience any interference from the users in the cluster. Therefore, the SINR of FU is computed as

$$\Psi_{\text{FU}} = \frac{|X_F|^2 \alpha_2 \rho_s}{1 + |X_F|^2 \alpha_1 \rho_s}. \quad (8)$$

3 Performance Analysis of the System

In this section, we derive the closed-form expressions for the OP and throughput, based on the expressions formed in the previous section. The outage probability, in our case, can be defined as the probability of SINR of the user that falls below the SINR threshold.

3.1 User NU Outage Probability

According to the principle of NOMA, at NU, the NU can detect the FU as a noise signal and NU detecting its signal. The outage will take place when NU is not able to detect both of these signals. Therefore, we can formulate the OP at NU as

$$P_{\text{NU}} = \Pr(\Psi_{\text{NU},s_2} \leq \rho_{th2}, \Psi_{\text{NU},s_1} \leq \rho_{th1}), \quad (9)$$

where the SINR threshold of each user is derived as $\rho_{th1} = 2^{2R_1} - 1$, $\rho_{th2} = 2^{2R_2} - 1$, with R_1 and R_2 as target rates.

Therefore, to make further computations easier, we compute OP at NU in two parts, and then we multiply to achieve the final expression of OP at NU.

$$P_{\text{NU},s_2} = \Pr(\Psi_{\text{NU},s_2} \leq \rho_{th2}). \quad (10)$$

Substituting (5) in (10), we can rewrite it as

$$P_{\text{NU},s_2} = \Pr\left(\frac{|X_N|^2 \alpha_2 \rho_s}{1 + |X_N|^2 \alpha_1 \rho_s} \leq \rho_{th2}\right). \quad (11)$$

Since the paper also follows cognitive radio networks, according to the principle of CR, we can rewrite it as [20]

$$\begin{aligned}
 P_{\text{NU},s2} &= \Pr \left(\frac{|X_N|^2 \alpha_2 \rho_s}{1 + |X_N|^2 \alpha_1 \rho_s} \leq \rho_{th2}, \bar{\rho}_s \leq \frac{\rho_P}{|h_{\text{SP}}|^2} \right) \\
 &+ \Pr \left(\frac{|X_N|^2 \alpha_2 \rho_P}{|h_{\text{SP}}|^2 + |X_N|^2 \alpha_1 \rho_P} \leq \rho_{th2}, \bar{\rho}_s > \frac{\rho_P}{|h_{\text{SP}}|^2} \right).
 \end{aligned} \tag{12}$$

Proposition 1 *The OP closed-form expression when NU cannot identify the FU signal, in the presence of CR, can be written as (13)*

$$\begin{aligned}
 P_{\text{NU},s2} &= \left(\left(\frac{1}{\Gamma(1+a)} \right) \gamma \left(1 + a, \sqrt{\frac{\rho_{th2}}{\bar{\rho}_s(\alpha_2 - \rho_{th2}\alpha_1)}} \right) \right) \left(1 - e^{-\frac{\rho_P}{\bar{\rho}_s \Omega_{\text{SP}}}} \right) \\
 &+ \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th2} \Omega_{\text{SP}}}{\rho_P(\alpha_2 - \rho_{th2}\alpha_1)}} \right)^{a+k+1} \gamma \left(\frac{a+k+3}{2}, \frac{\rho_P}{\bar{\rho}_s \Omega_{\text{SP}}} \right).
 \end{aligned} \tag{13}$$

Proof 1 Rewriting (12) as

$$\begin{aligned}
 P_{\text{NU},s2} &= \Pr \left(\underbrace{\frac{|X_N|^2 \alpha_2 \bar{\rho}_s}{1 + |X_N|^2 \alpha_1 \bar{\rho}_s} \leq \rho_{th2}, \bar{\rho}_s \leq \frac{\rho_P}{|h_{\text{SP}}|^2}}_{A_1} \right) \\
 &+ \Pr \left(\underbrace{\frac{|X_N|^2 \alpha_2 \rho_P}{|h_{\text{SP}}|^2 + |X_N|^2 \alpha_1 \rho_P} \leq \rho_{th2}, \bar{\rho}_s > \frac{\rho_P}{|h_{\text{SP}}|^2}}_{A_2} \right).
 \end{aligned} \tag{14}$$

To reduce the computational complexity, we compute the expressions separately as A_1 and A_2 and combine the result as shown in (13)

$$A_1 = \Pr \left(|X_N| \leq \sqrt{\frac{\rho_{th2}}{\bar{\rho}_s(\alpha_2 - \alpha_1 \rho_{th2})}} \leq \rho_{th2}, |h_{\text{SP}}|^2 \leq \frac{\rho_P}{\bar{\rho}_s} \right), \tag{15}$$

which can be written as

$$A_1 = F_{|X_N|} \left(\sqrt{\frac{\rho_{th2}}{\bar{\rho}_s(\alpha_2 - \alpha_1 \rho_{th2})}} \right) F_{|h_{\text{SP}}|^2} \left(\frac{\rho_P}{\bar{\rho}_s} \right). \tag{16}$$

We consider $F(x)$ as cumulative distributive function of variable x . Therefore, referring to [41, Eq. (8)], we can write it as

$$A_1 = \left(\left(\frac{1}{\Gamma(1+a)} \right) \gamma \left(1 + a, \sqrt{\frac{\rho_{th2}}{\bar{\rho}_s(\alpha_2 - \rho_{th2}\alpha_1)}} \right) \right) \left(1 - e^{-\frac{\rho_P}{\bar{\rho}_s \Omega_{\text{SP}}}} \right), \tag{17}$$

where $a = \frac{k_1^2}{k_2}$, $b = \frac{k_2}{k_1}$, $k_1 = \frac{N\pi}{2}$, $k_2 = 4N \left(1 - \frac{\pi^2}{16} \right)$,

$\Gamma(a, \frac{x}{b})$ denotes Gamma function, while $\gamma(a, \frac{x}{b})$ denotes upper and lower gamma functions, with respect to [[15], Eq. (8.350/2), (8.350/3)].

$$A_2 = \Pr \left(|X_N| \leq \sqrt{\frac{\rho_{th2} |h_{SP}|^2}{\rho_P (\alpha_2 - \alpha_1 \rho_{th2})}}, |h_{SP}|^2 > \frac{\rho_P}{\rho_s} \right) \tag{18}$$

$$A_2 = \int_{\frac{\rho_P}{\rho_s}}^{\infty} F_A \left(\sqrt{\frac{\rho_{th2}(x)}{\rho_s (\alpha_2 - \alpha_1 \rho_{th2})}} \right) f_{|h_{SP}|^2}(x) dx. \tag{19}$$

Applying [[15], 8.354.1], we can rewrite the above equation as

$$A_2 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th2} \Omega_{SP}}{\rho_P (\alpha_2 - \rho_{th2} \alpha_1)}} \right)^{a+k+1} \int_{\frac{\rho_P}{\rho_s}}^{\infty} x^{\frac{a+k+1}{2}} e^{-\frac{x}{\Omega_{SP}}} dx \tag{20}$$

$$A_2 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th2} \Omega_{SP}}{\rho_P (\alpha_2 - \rho_{th2} \alpha_1)}} \right)^{a+k+1} \Gamma \left(\frac{a+k+3}{2}, \frac{\rho_P}{\rho_s \Omega_{SP}} \right). \tag{21}$$

Substituting A_1 and A_2 in (14) will obtain the final expression at (13) This completes the proof.

$$P_{NU,s1} = \Pr(\Psi_{NU,s1} \leq \rho_{th1}). \tag{22}$$

Substituting (7) in (22), with respect to the CR, we can write the NU signal after SIC as

$$P_{NU,s1} = \Pr \left(|X_N|^2 \alpha_1 \rho_s \leq \rho_{th1}, \bar{\rho}_s \leq \frac{\rho_P}{|h_{SP}|^2} \right) + \Pr \left(\frac{|X_N|^2 \alpha_1 \rho_P}{|h_{SP}|^2} \leq \rho_{th2}, \bar{\rho}_s > \frac{\rho_P}{|h_{SP}|^2} \right). \tag{23}$$

Proposition 2 *The OP closed-form expression when NU cannot identify its own signal, in the presence of CR, can be written as (24)*

$$P_{NU,s1} = \left(\left(\frac{1}{\Gamma(1+a)} \right) \gamma \left(1 + a, \frac{\sqrt{\frac{\rho_{th1} \alpha_1}{\rho_s}}}{b} \right) \right) \left(1 - e^{-\frac{\rho_P}{\rho_s \Omega_{SP}}} \right) + \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th1} \alpha_1 \Omega_{SP}}{\rho_P}} \right)^{a+k+1} \Gamma \left(\frac{a+k+3}{2}, \frac{\rho_P}{\rho_s \Omega_{SP}} \right). \tag{24}$$

Proof 2 Rewriting (23) as

$$P_{NU,s1} = \underbrace{\Pr \left(|X_N|^2 \alpha_1 \rho_s \leq \rho_{th1}, \bar{\rho}_s \leq \frac{\rho_P}{|h_{SP}|^2} \right)}_{A_1} + \underbrace{\Pr \left(\frac{|X_N|^2 \alpha_1 \rho_P}{|h_{SP}|^2} \leq \rho_{th1}, \bar{\rho}_s > \frac{\rho_P}{|h_{SP}|^2} \right)}_{A_2}, \tag{25}$$

$$A_1 = \Pr \left(|X_N| \leq \sqrt{\frac{\rho_{th1}\alpha_1}{\rho_s}} \leq \rho_{th2}, |h_{SP}|^2 \leq \frac{\rho_P}{\rho_s} \right) \tag{26}$$

$$A_1 = F_A \left(\sqrt{\frac{\rho_{th1}\alpha_1}{\rho_s}} \right) F_{|h_{SP}|^2} \left(\frac{\rho_P}{\rho_s} \right) \tag{27}$$

$$A_1 = \left(\left(\frac{1}{\Gamma(1+a)} \right) \gamma \left(1 + a, \frac{\sqrt{\frac{\rho_{th1}\alpha_1}{\rho_s}}}{b} \right) \right) \left(1 - e^{-\frac{\rho_P}{\rho_s \Omega_{SP}}} \right), \tag{28}$$

$$A_2 = \Pr \left(|X_N|^2 \alpha_1 \rho_P \leq \rho_{th2}, \bar{\rho}_s > \frac{\rho_P}{|h_{SP}|^2} \right) \tag{29}$$

$$A_2 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th1}\alpha_1\Omega_{SP}}{\rho_P}} \right)^{a+k+1} \int_{\frac{\rho_P}{\rho_s}}^{\infty} x^{\frac{a+k+1}{2}} e^{-\frac{x}{\Omega_{SP}}} dx \tag{30}$$

$$A_2 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th1}\alpha_1\Omega_{SP}}{\rho_P}} \right)^{a+k+1} \Gamma \left(\frac{a+k+3}{2}, \frac{\rho_P}{\rho_s \Omega_{SP}} \right). \tag{31}$$

Substituting A_1 and A_2 in (25) will obtain the final expression at (24) This completes the proof.

Therefore, the final OP at the user NU is

$$P_{NU} = P_{NU,s2} \times P_{NU,s1}, \tag{32}$$

$$\begin{aligned} P_{NU} = & \left\{ \left(\left(\frac{1}{\Gamma(1+a)} \right) \gamma \left(1 + a, \frac{\sqrt{\frac{\rho_{th1}\alpha_1}{\rho_s}}}{b} \right) \right) \left(1 - e^{-\frac{\rho_P}{\rho_s \Omega_{SP}}} \right) \right. \\ & \left. + \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th1}\alpha_1\Omega_{SP}}{\rho_P}} \right)^{a+k+1} \Gamma \left(\frac{a+k+3}{2}, \frac{\rho_P}{\rho_s \Omega_{SP}} \right) \right\} \\ & \times \left\{ \left(\left(\frac{1}{\Gamma(1+a)} \right) \gamma \left(1 + a, \frac{\sqrt{\frac{\rho_{th2}}{\rho_s(\alpha_2 - \rho_{th2}\alpha_1)}}}{b} \right) \right) \left(1 - e^{-\frac{\rho_P}{\rho_s \Omega_{SP}}} \right) \right. \\ & \left. + \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_{th2}\Omega_{SP}}{\rho_P(\alpha_2 - \rho_{th2}\alpha_1)}} \right)^{a+k+1} \Gamma \left(\frac{a+k+3}{2}, \frac{\rho_P}{\rho_s \Omega_{SP}} \right) \right\}. \tag{33} \end{aligned}$$

3.2 User FU Outage Probability

Similar to the s_1 signal in NU , we compute the FU signal as

$$P_{FU} = \Pr (\Psi_{FU} \leq \rho_{th2}). \tag{34}$$

Substituting (8) in (34), with respect to the CR, we can write the FU signal as

$$P_{FU} = \Pr \left(\frac{|X_F|^2 \alpha_2 \bar{\rho}_s}{1 + |X_F|^2 \alpha_1 \bar{\rho}_s} \leq \rho_t h_2, \bar{\rho}_s \leq \frac{\rho_P}{|h_{SP}|^2} \right) + \Pr \left(\frac{|X_F|^2 \alpha_2 \rho_P}{1 + |X_F|^2 \alpha_1 \rho_P} \leq \rho_t h_2, \bar{\rho}_s > \frac{\rho_P}{|h_{SP}|^2} \right). \quad (35)$$

Proposition 3 *The OP closed-form expression at user FU, in the presence of CR, can be written as (36)*

$$P_{FU} = \left(\left(\frac{1}{\Gamma(1+a)} \right) \gamma \left(1 + a, \sqrt{\frac{\rho_t h_2}{\bar{\rho}_s (\alpha_2 - \rho_t h_2 \alpha_1)}} \right) \right) \left(1 - e^{-\frac{\rho_P}{\bar{\rho}_s \Omega_{SP}}} \right) + \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+a+1)\Gamma(1+a)} \left(\frac{1}{b} \sqrt{\frac{\rho_t h_2 \Omega_{SP}}{\rho_P (\alpha_2 - \rho_t h_2 \alpha_1)}} \right)^{a+k+1} \Gamma \left(\frac{a+k+3}{2}, \frac{\rho_P}{\bar{\rho}_s \Omega_{SP}} \right). \quad (36)$$

Proof 3 Since the derivation of this expression is similar to Proposition 1, we do not write it here.

3.3 Throughput of the System

In our case, we consider a delay-limited data transmission between the RIS-aided NOMA CR system. To calculate further performance metrics of the system, we analyze the throughput of the whole system with target rates as R_1 and R_2 at NU and FU, respectively. Therefore, the throughput of the system can be computed as

$$T = R_1(1 - P_{NU}) + R_2(1 - P_{FU}). \quad (37)$$

4 Numerical Results and Discussions

In this section, we investigate the performance of the proposed system by simulating the obtained expressions in the previous section. We set the target rates $R_1 = R_2 = 0.5$. Since we set similar target rates, we simulate for same SINR threshold to obtain a fair comparison between the users. The SNR of interference constraint is set to $\bar{\rho}_P = 20$ dB. The power level coefficients are set as $\alpha_1 = 0.3$ and $\alpha_2 = 0.7$. Monte Carlo simulations are performed to analyze the correctness of the obtained OP expressions.

In Fig. 2, the simulations are performed to OP versus ρ_s of user NU by varying the number of meta-surfaces N . We can observe that the analytical results match exactly with the simulations, verifying the authenticity of theoretical analysis. As we can observe from the simulation, the OP performance enhances significantly for the rise in the number of meta-surfaces. But for the increase in the SNR, the OP increases, decreasing the performance of the user.

Fig. 2 Outage probability versus ρ_s varying number of meta-surfaces at user NU

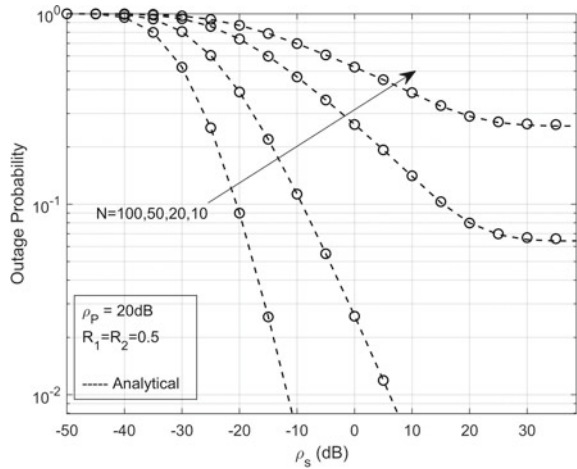
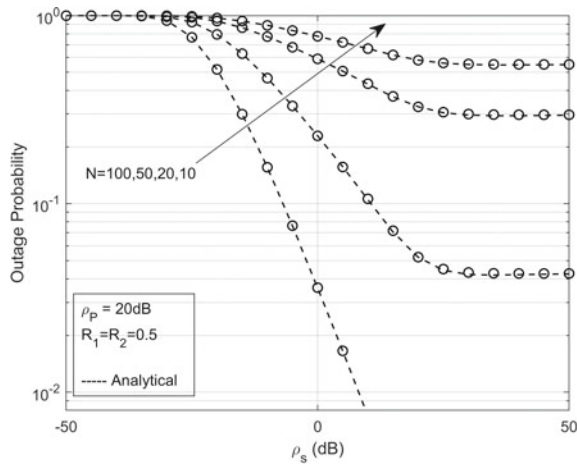


Fig. 3 Outage probability versus ρ_s varying number of meta-surfaces at user FU



In Fig. 3, the simulations are performed to OP versus ρ_s at user FU, by varying the number of meta-surfaces. We can observe that analytical and simulations are matched, verifying the theoretical analysis. The performance of user FU is enhanced efficiently with the increase in the number of meta-surfaces. Figure 4 shows the comparison of OP versus ρ_s for the different number of meta-surfaces of user NU and FU. As we can observe from the simulation, the performance of NU is far better than the performance of FU when the RIS-assisted NOMA system is enabled by CR.

Figure 5 shows the throughput versus ρ_s of the entire system by varying the number of meta-surfaces. As we can observe, with the increase in the meta-surfaces, the performance of the CR-enabled NOMA-aided RIS system increases significantly. Figure 6 shows the comparison of the proposed NOMA system with the benchmark OMA system by similarly varying the number of meta-surfaces.

Fig. 4 Comparison of outage probability versus ρ_s at users NU and FU by varying the number of meta-surfaces

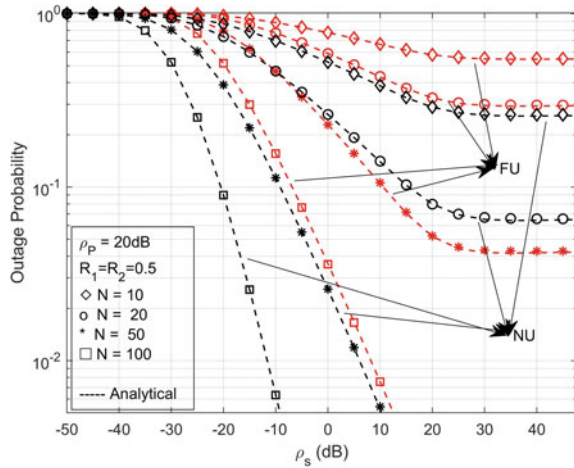
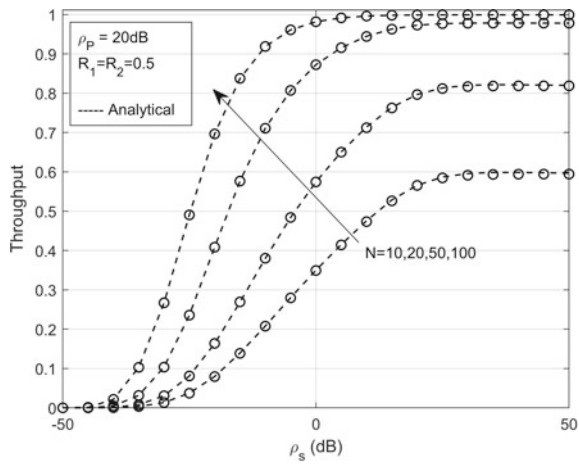


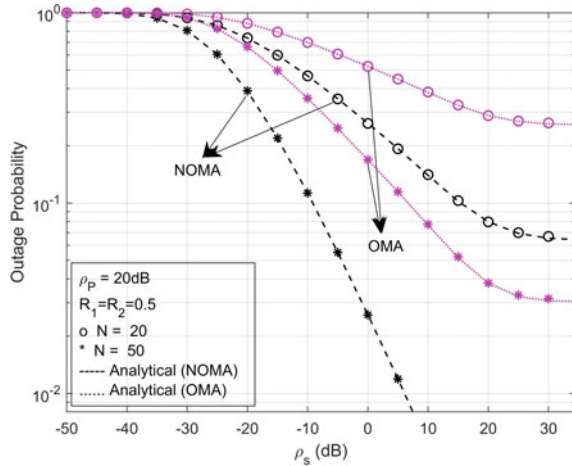
Fig. 5 Throughput of the system



5 Conclusion

With the introduction of cognitive radio networks and NOMA, the effective utilization of spectrum has begun to multiple users. NOMA can serve multiple users in a single resource block, and CR can assign unused spectrum to the secondary users, by maximum limiting the disruptive interference to primary users. However, RIS can direct the received signal according to the user requirement and position. In our paper, we proposed a NOMA-assisted RIS system enabled by cognitive radio technology to increase the efficiency of data transmission between the base station and users. The proposed system consists of a primary network and a secondary network with two NOMA users being served at the destination. We have derived the closed-form expressions for outage performance of both users NU and FU. Monte

Fig. 6 Comparison of NOMA and OMA system aided by CR-enabled RIS system



Carlo simulations are performed to verify the authenticity of the derived expression. Throughout the simulations, the number of meta-surfaces and transmit SNR has been identified as the primary factors to affect the outage performance of the system. The throughput of the system is analyzed to understand the overall performance of the proposed system. We conclude that for a sufficiently large number of meta-surfaces, the proposed system outage performance is significantly enhanced.

The future scope of this work can be followed by implementing imperfect SIC, imperfect CSI, and hardware impairments in the secondary network of the system. Implementation of imperfect scenarios can lead to understand the performance of the system in real-time cases.

References

1. Lv L, Chen J, Ni Q (2016) Cooperative non-orthogonal multiple access in cognitive radio. *IEEE Commun Lett* 20(10):2059–62
2. Kang X, Liang YC, Nallanathan A, Garg HK, Zhang R (2009) Optimal power allocation for fading channels in cognitive radio networks: ergodic capacity and outage capacity. *IEEE Trans Wirel Commun* 8(2):940–50
3. Gong S, Lu X, Hoang DT, Niyato D, Shu L, Kim DI, Liang YC (2012) Towards smart radio environment for wireless communications via intelligent reflecting surface: a comprehensive survey. *IEEE Commun Surv Tutor* (to appear)
4. Holloway CL, Kuester EF, Gordon JA, O’Hara J, Booth J, Smith DR (2012) An overview of the theory and applications of metasurfaces: the two-dimensional equivalents of metamaterials. *IEEE Antennas Propag Mag* 54(2):10–35 Jul 3
5. Huang C, Alexandropoulos GC, Zappone A, Debbah M, Yuen C (2018) Energy efficient multi-user MISO communication using low resolution large intelligent surfaces. In: 2018 IEEE globecom workshops (GC Wkshps). IEEE, pp 1–6
6. Subrt L, Pechac P (2012) Controlling propagation environments using intelligent walls. In: 2012 6th European conference on antennas and propagation (EUCAP). IEEE, pp 1–5

7. Arzykulov S, Nauryzbayev G, Tsiftsis TA, Maham B (2019) Performance analysis of underlay cognitive radio nonorthogonal multiple access networks. *IEEE Trans Veh Technol* 68(9):9318–22
8. Liu Y, Ding Z, Elkashlan M, Yuan J (2016) Nonorthogonal multiple access in large-scale underlay cognitive radio networks. *IEEE Trans Veh Technol* 65(12):10152–7
9. Chen Y, Wang L, Jiao B (2017) Cooperative multicast non-orthogonal multiple access in cognitive radio. In: 2017 IEEE international conference on communications (ICC). IEEE, pp 1–6
10. Xu D, Yu X, Sun Y, Ng DW, Schober R (2020) Resource allocation for IRS-assisted full-duplex cognitive radio systems. *IEEE Trans Commun* 68(12):7376–94
11. Guan X, Wu Q, Zhang R (2020) Joint power control and passive beamforming in IRS-assisted spectrum sharing. *IEEE Commun Lett* 24(7):1553–7
12. Hemanth A, Umamaheswari K, Pogaku AC, Do DT, Lee BM (2020) Outage performance analysis of reconfigurable intelligent surfaces-aided NOMA under presence of hardware impairment. *IEEE Access* 24(8):212156–65
13. Thirumavalavan VC, Jayaraman TS (2020) BER analysis of reconfigurable intelligent surface assisted downlink power domain NOMA system. In: 2020 international conference on COMmunication systems NETworkS (COMSNETS). IEEE, pp 519–522
14. Basar E, Di Renzo M, De Rosny J, Debbah M, Alouini MS, Zhang R (2019) Wireless communications through reconfigurable intelligent surfaces. *IEEE Access* 13(7):116753–73
15. Gradshteyn IS, Ryzhik IM (2000) Table of integrals, series, and products, 6th edn. Academic, New York
16. Tahir B, Schwarz S, Rupp M (2021) RIS-assisted code-domain MIMO-NOMA. arXiv preprint [arXiv:2103.06985](https://arxiv.org/abs/2103.06985)
17. Hou T, Liu Y, Song Z, Sun X, Chen Y (2020) MIMO-NOMA networks relying on reconfigurable intelligent surface: a signal cancellation-based design. *IEEE Trans Commun* 68(11):6932–44
18. Yue X, Liu Y (2020) Performance analysis of intelligent reflecting surface assisted NOMA networks. arXiv preprint [arXiv:2002.09907](https://arxiv.org/abs/2002.09907)
19. Wu Q, Zhang R (2019) Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts. *IEEE Trans Commun* 68(3):1838–51
20. Do DT, Le AT, Lee BM (2020) NOMA in cooperative underlay cognitive radio networks under imperfect SIC. *IEEE Access* 6(8):86180–95

An Intelligent Vehicular Communication-Based Framework to Provide Seamless Connectivity in WBAN



Koushik Karmakar, Sohail Saif, Suparna Biswas, and Sarmistha Neogy

Abstract Providing data connectivity seamlessly without any kind of interruption is one of the compulsory requirements of sensor-based remote health monitoring system. It is therefore an important topic for research in modern times. It should support mobility handling as well as patients' movement equally as patient may move to different places at different times. Different cases are required to be considered to provide this service. A health monitoring framework should support all such movements and should ensure that proper connectivity is maintained even when the patient is moving in a car, and there may not be proper infrastructure along the road side for providing uninterrupted data connectivity. Ideally, a health monitoring framework should support such connectivity at all times. In our present work, we considered different types of patients' mobility and worked on how to provide seamless connectivity using intelligent vehicular communication-based architecture. This work is motivated based on the concept of Internet of vehicles and vehicular movement. In this paper, we have designed a mobility handling protocol which will handle patients' movement and provide seamless data connectivity at all times.

Keywords Health monitoring system · WBAN · Internet of vehicles · Mobility handling protocol

1 Introduction

Researchers work on different fields like the smart devices, healthcare monitoring system, sensor networks and many other similar things are carried out for last few

K. Karmakar (✉)

Narula Institute of Technology, Kolkata, India

e-mail: k.karmakar100@gmail.com

S. Saif · S. Biswas

Maulana Abul Kalam Azad University of Technology, Haringhata, Nadia, West Bengal, India

S. Neogy

Jadavpur University, Kolkata, India

years. Health care is one of the prominent areas, and sensor-based healthcare monitoring system is getting popularity. Chronic diseases and related expenditure are growing. Healthcare facility is also not available at all the places equally specially in remote rural areas. Sensor-based virtual health monitoring method can be a substitute of the traditional system [1, 2]. As per this technique, small sensor nodes are used to capture and send patients' health data to the doctor's chamber. To get the accurate result, proper connectivity in WBAN is required. Proper fault detection method in WBAN helps us maintain uninterrupted data connectivity. In this work, we made a survey of different state-of-the-art fault detection techniques [3–5] and later proposed a method to maintain seamless connectivity when the patient is moving in the vehicle.

This paper is written according to the following sequences. In Sect. 2, a literature survey and state-of-the-art works on fault detection in WBAN is shown. Problem description is given in Sect. 3. Proposed solution is described in Sect. 4. Experimental results have been described in Sect. 5. Performance analysis has been shown in Sect. 6. Conclusion is given in Sect. 7.

2 Literature Survey

Fault detection is a vital factor because faults in sensors can hampers the seamless data connectivity in WBAN. Fault detection in routing of WBAN is therefore need to be discussed properly. A lot of fault detection works are carried out based on machine learning-based techniques. In this work [6], such an approach is followed where different algorithms like K-NN, SVM and other algorithms are used. Some works are done based on coordination between the neighbors [7]. Various types of methods are used like data received technique based on multiple-hop, periodic timeout mechanism and many other methods. In this paper [8], co-operative coding approach is followed. This technique is based on establishing a co-operation among all the bio-sensors. Various types of fault detection techniques are described in this paper [9]. Some works are also done based on machine learning and artificial neural network (ANN) [10]. In this survey paper [11], several such fault detection techniques have been described. A few such fault detection technique [12] is designed based on the concept of assigning priority to different types of collected sensor data. Later, these sensor data are compared with the threshold values. Fault is detected based on this comparison results. Another fault detection technique is described in this work [13]. This model for fault detection is based on trapezoidal fuzzy numbers and other factors. Within the wireless sensor network, different fault finding as well as fault tolerance methods is given [14]. In another fault detection technique, one algorithm is designed based on the unsupervised learning method [15]. According to this method, condition of each sensor node is decided based on different factors. Later using a distributed algorithm, status of the sensor nodes is finally corrected. This method is especially applicable for transient faults (Table 1).

Table 1 Comparison of different fault detection techniques

Serial No.	Authors	Technology used and application	Simulation environment	Results
1	Kaur et al. [6]	Different machine learning methods are used like random forest, K-NN, MLP and support vector machine	Used Waikato Environment for Knowledge Analysis (WEKA) open source tool	Highest accuracy is 96.42%
2	Swain et al. [7]	Based on coordination based approach between neighbors	NS-2.35 based simulation is done	Accuracy is up to 95%
3	Mehmoodi et al. [8]	Based on network coding technique as well as cooperative communication method. It reduces channel related and other problems	Simulation is done in MATLAB	Better performance achieved
4	Sajan et al. [9]	Many state-of-the-art works related to wireless sensor network are discussed	Based on NS-2.35 based simulation	Accuracy is 93.18%
5	Nagdeo and Mahapatro [10]	Based on back-propagation neural network and linear regression method	Physionet database is used in the experiment	Very good results obtained
6	Meena Abarna and Venkat-chalapathy [11]	Proposed a sensor-based fault detection and recovery method applicable for WBAN	Based on NS-2-based simulation	Accurate result obtained
7	Jeong et al. [12]	Proposed a cloud-based fault monitoring and detection system used for fault tolerance system	Implemented in a cloud service area	Gives better results in terms of its use in a cloud-based environment

(continued)

Table 1 (continued)

Serial No.	Authors	Technology used and application	Simulation environment	Results
8	Rezvani and Ghorasi [13]	Discussed channel-level fault detection method in WBAN	Simulation was done while considering IEEE values as the standard	Better output with respect to different parameters like power consumption, average delay, packet loss

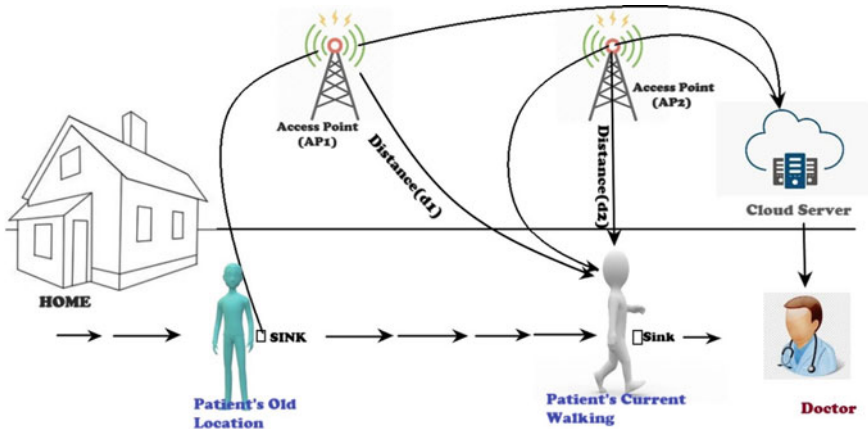


Fig. 1 Movement of the patient through walking

3 Problem Description

Heterogeneous sensor nodes can be used for collecting physiological data. They collect information which will be send to the remote server for details analysis. However, for proper diagnosis seamless connectivity is absolutely required. But, we need to consider mobility of the patient as well. We have considered two different cases. In the very first case, we consider the patient movement when the patient is moving from one place to another place (Fig. 1). In the second case, we consider a situation when the patient is moving from one place to another place in a car (Fig. 2). We have to consider both the cases. We have to maintain connectivity in all the cases.

4 Proposed Solution

Providing seamless connectivity in WBAN setup for real-time health data transmission at all times at a negligible delay and without any significant data loss is the

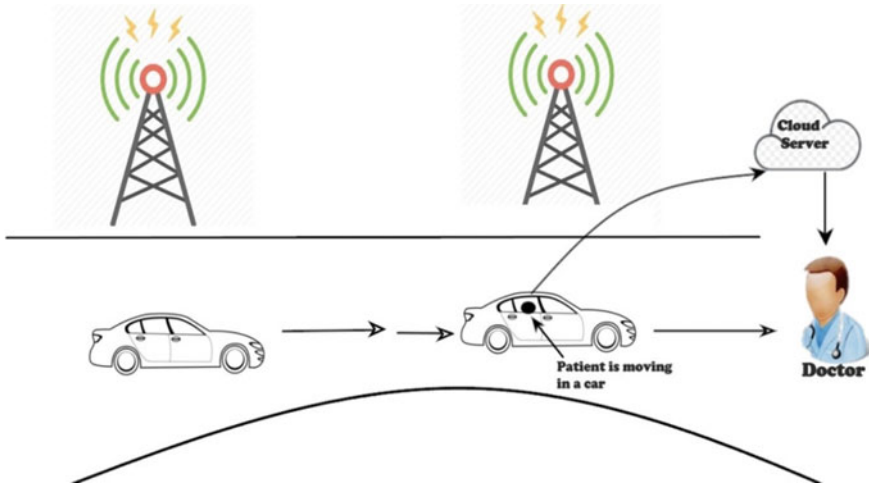


Fig. 2 Movement of the patient in a car

prime objective of our current research. It should be able to transfer data at different times when the patient will be at different positions. In brief, the WBAN-based health monitoring system must handle mobility properly for smooth data transfer. In this paper, we have considered following different cases for patients movement and discuss how seamless connectivity can be provided there.

Case 1: When the patient is staying at home. Connectivity is through the fixed network setup and will be as follows:

Sensor nodes => Sink node => Fixed Access Point (AP) => Cloud server => Doctor's chamber.

Case 2: When the patient is walking outside his house. Here the connectivity will be like case1 but for data transfer AP may be changed. New connectivity will be as follows:

```
Calculate,  
Sink_to_AP_distance (d1) = (Old AP - Sink node) distance;  
Sink_to_AP_distance (d2) = (New AP - Sink node) distance;  
If (d2 > d1).  
{  
  Sensor nodes => Sink node => Old Access Point (AP) => Cloud server =>  
  Doctor's chamber.  
}  
Else  
{  
  Sensor nodes => Sink node => New Access Point (AP) => Cloud server =>  
  Doctor's chamber.  
}
```



Fig. 3 WBAN vehicle-to-vehicle connectivity

Case 3: One of the major challenges in this research is how to provide seamless connectivity while moving in a car especially on a highway. Intelligent vehicular communication-based framework can be used to solve this problem. Vehicle running on the road must have network internet connectivity. Those vehicles will share their information between one another. Information includes road conditions, current traffic conditions and total number of vehicles running on the road at that time and other related thing. This process is continued until next AP is found. This entire process is shown in Fig. 3.

1. Driver's smart phone captures outside information and sent to the local cloud.
2. After thorough analysis in the vehicular cloud, an idea will be available regarding current traffic condition and road condition.
3. Patients' vehicle will go through this path.
4. Similarly a request from patients' vehicle will go to the next vehicle and carry forward information to the subsequent destination.
5. That vehicle will forward the information to the next vehicle.
6. It will be followed till the next AP is available.
7. AP to the cloud communication will be done normally.

5 Proposed Scenario

Consider the following source to destination map (Fig. 4). T1 is the source, and T10 is the destination according to this graph. There are several paths between T1 and T10. Different paths along with their distances in kilometer are shown below. T2 to

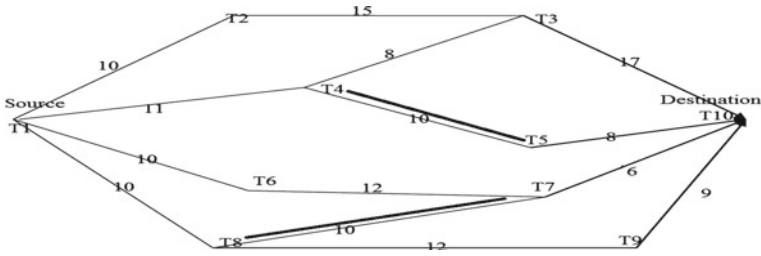


Fig. 4 Source to destination path map

T9 denote different crossing or turning point on the road. There are several ways to go the destination from the source, and they are as follows:

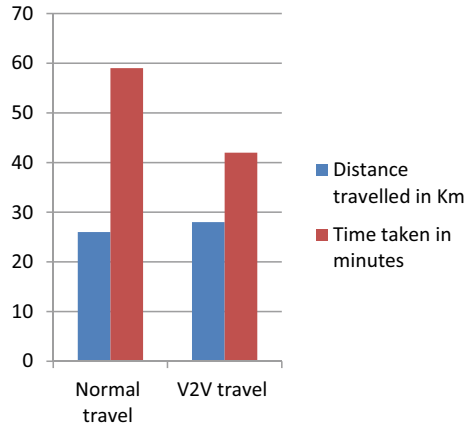
1. T1- > T2- > T3- > T10 = > [10 + 15 + 17] = 42;
2. T1- > T4- > T5- > T10 = > [11 + 10 + 08] = 29;
3. T1- > T6- > T7- > T10 = > [10 + 12 + 06] = 28;
4. T1- > T8- > T9- > T10 = > [10 + 12 + 09] = 31;
5. T1- > T8- > T7- > T10 = > [10 + 10 + 06] = 26;

A patient wearing WBAN will start from source T1 for the destination T10. Car will follow any path, but normally it should follow the shortest path. From the above graph, we can calculate the shortest path, and it should be T1- > T8- > T7- > T10. The car carrying the patient will suppose to follow that path but found that the road connecting T8 and T7 is congested. This traffic condition will also be available from Google Maps. However, if Access Point (AP) is not available in the nearby location, then connectivity may be disrupted. In that case, intelligent vehicle-to-vehicle communication will help maintain connectivity seamlessly. In that case, it will follow the remaining shortest path, i.e., T1- > T6- > T7- > T10. It will travel only 28 kms. With an average speed of 40 km/hr, this car will reach its destination in 42 min only.

6 Performance Analysis

We divide our journey into two parts. 1st part of the journey covers that part when the patient will walk from his home to car (Fig. 1). We consider this distance as D1 when proper infrastructure is there. Remaining distance (D2) will be covered in a car (Fig. 2). When proper infrastructure may not be available there in the highway, vehicle-to-vehicle connectivity can be used to provide real-time information regarding the shortest distance, traffic condition and above all maintaining uninterrupted connectivity through the WBAN. In normal case, it will follow the normal shortest distance path T1- > T8- > T7- > T10 and will cover a distance of 26 kms and car will reach its destination in 39 min. However, considering huge traffic jam between T8 and T7 a waiting time is required which we assume to be 20 min. So,

Fig. 5 Distance traveled versus time taken



total time required for the journey will be $[39 + 20] = 59$ min. Distance traveled vs. time taken is shown graphically in Fig. 5.

7 Conclusion

Our current work ensures seamless data connectivity in WBAN. This protocol will work even when the patient is moving in a car from one place to another. It provides full connectivity when any fixed infrastructure is not available nearby. We have intelligently used vehicle-to-vehicle connectivity in health data monitoring system. This is the first time such vehicular connectivity concept has been used in health care for patient monitoring. We have considered different cases for patient movement, and the protocol was designed accordingly. It allows remote access for the patients and also the outdoor movement.

References

1. Qu Y, Zheng G, Ma H, Wang X, Ji B, Wu H (2019) A survey of routing protocols in WBAN for healthcare applications. *Sensors* 19(7):1638
2. Pandey I, Dutta HS, Banerjee JS (2019) WBAN: a smart approach to next generation e-healthcare system. In: 2019 3rd International conference on computing methodologies and communication (ICCMC), 27–29 March 2019, Erode, India
3. Masdari M (2020) Towards effective fault detection in heterogeneous wireless sensor networks. *Int J Ad Hoc Ubiquit Comput* 33(4):216–228
4. Salayma M, Al-Dubai A, Romdhani I (2017) Wireless body area network (WBAN): a survey on reliability, fault tolerance and technologies co-existence. *ACM Comput Surv* 50(1). Article 3
5. Karmakar K, Saif S, Biswas S, Neogy S (2021) A WBAN-based framework for health condition monitoring and faulty sensor node detection applying ANN. *Int J Biomed Clin Eng* 10(2):44–65

6. Kaur P, Ravinder K, Munish K (2019) A healthcare monitoring system using random forest and internet of things (IoT). In: Springer multimedia tools and applications, pp 1–12
7. Swain RR, Khilar PM, Bhoi SK (2020) Underlying and persistence fault diagnosis in wireless sensor networks using majority neighbors co-ordination approach. *Wireless Pers Commun* 111(2):763–798
8. Mehmoodi G, Khani MZ, Abbas S, Faisal M, Rahman HU (2020) An energy-efficient and cooperative fault-tolerant communication approach for wireless body area network. *IEEE Access* 8:69134–69147
9. Sajan S, Chacko SJ, Pai V, Pai BHK (2020) Performance evaluation of various algorithms that affect fault detection in wireless sensor network. In: 2020 Fourth International conference on inventive systems and control (ICISC), 2020 Jan 8. IEEE, pp 540–545
10. Nagdeo SK, Mahapatro J (2019) Wireless body area network sensor faults and anomalous data detection and classification using machine learning. IEEE Bombay section signature conference (IBSSC) at Mumbai, India, on 26–28 July 2019. <https://doi.org/10.1109/IBSSC47189.2019.8973004>
11. Meena Abarna KT, Venkatchalopathy K (2014) Cluster based failure detection and recovery technique in wireless body area network. *Res J Appl Sci Eng Technol* 7(17):3458–3465
12. Jeong Y-S, Kim H-W, Park JH (2014) Visual scheme monitoring of sensors for fault tolerance on wireless body area networks with cloud service infrastructure. *Int J Distrib Sensor Netw*. <https://doi.org/10.1155/2014/154180>
13. Rezvani S, Ghorashi SA (2013) Context aware and channel-based resource allocation for wireless body area networks. *IET Wireless Sensor Systems* 3(1):16–25. <https://doi.org/10.1049/iet-wss.2012.0100>
14. Yang Y, Liu Q, Gao Z, Qiu X, Meng L (2015) Data fault detection in medical sensor networks. *MDPI Sensors* 15(3):6066–6090. <https://doi.org/10.3390/s150306066>
15. Sheth H, Jani R (2020) Fault tolerance and detection in wireless sensor networks. In: Data science and intelligent applications. Springer, Singapore, pp 431–437

A Novel RAW Slot Allocation Scheme for Improving the Performance of IEEE 802.11ah Multi-rate IoT Networks



Badarla Sri Pavan and V. P. Harigovindan

Abstract IEEE 802.11ah is implemented for Internet of things (IoT) applications. The important feature in IEEE 802.11ah MAC layer is restricted access window (RAW) mechanism, which divides stations (STAs) into groups and every group designate with a RAW slot. In IEEE 802.11ah, uniform grouping (UG) is the default scheme having distinct rate STAs in each group is the root cause of anomaly. Due to distinct rate STAs, the high rate STAs are penalized by low rate STAs which consume more channel time for transmission. In this paper, we present a RAW slot allocation (RSA) scheme based on achievable data rates to overcome the anomaly. We also present an analytical model to compute the throughput and energy efficiency of IEEE 802.11ah network. From results, the RSA scheme resolves the anomaly and upgrades the IEEE 802.11ah network's performance. The analytical results validated with simulations.

Keywords Internet of things · IEEE 802.11ah · RSA scheme

1 Introduction

The rapid changes in communication technology invoke exponential increase in stations (STAs) connected to Internet. The Internet of things (IoT) is a booming technology where different objects are connected with advanced capabilities for several applications known for smart grid, e-health, smart home, etc. [1]. IoT serves different STAs by sharing information through Internet at anytime. The ever-increasing applications for different requirements provide an opportunity for developers with daunting tasks. Since three decades, IEEE 802.11 WLANs proved their superiority to offer Internet services and draw the researchers eye [2]. It affords physical and

B. S. Pavan (✉) · V. P. Harigovindan
Department of E.C.E., National Institute of Technology Puducherry, Karaikal 609609, India
e-mail: sripavan.rvce@gmail.com

V. P. Harigovindan
e-mail: hari@nitpy.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
S. Dhar et al. (eds.), *Advances in Communication, Devices and Networking*, Lecture Notes
in Electrical Engineering 902, https://doi.org/10.1007/978-981-19-2004-2_54

593

medium access control (MAC) layer specifications. The STAs can select multiple rates using modulation and coding schemes (MCSs) with respect to their position from access point (AP). The STAs near AP have higher rate, whereas at fringes have lower rate [3].

IEEE 802.11ah is developed by IEEE 802 task group to overcome the contention among STAs in dense networks [4]. It features low power, scalability, and high throughput. It works in sub 1 GHz bands with bandwidths of 1, 2, 4, 8, and 16 MHz. The STAs utilize several MCSs support for multiple rate selection. In MAC layer, an important restricted access window (RAW) mechanism is promoted to mitigate contention [5].

1.1 Performance Anomaly

In multi-rate networks, the performance anomaly is a severe issue where low rate STAs penalizes high rate STAs. The low rate STAs at fringes of AP occupy the channel more time because of bad channel conditions. The anomaly issue arises because of distributed coordination function (DCF) where it ignores data rates of STAs [6]. Hence, lower rate STAs use more channel time lead to penalization of high rate STAs [7]. This anomaly issue also existed in IEEE 802.11ah multi-rate networks due to DCF. This affects the aggregate performance of network. The paper structured as follows: Sect. 2 provides related research work. Section 3 describes the network architecture of IEEE 802.11ah. The throughput analysis of IEEE 802.11ah and resolving of anomaly is discussed in Sect. 4. Section 5 gives the results and discussion. Section 6 provides conclusion.

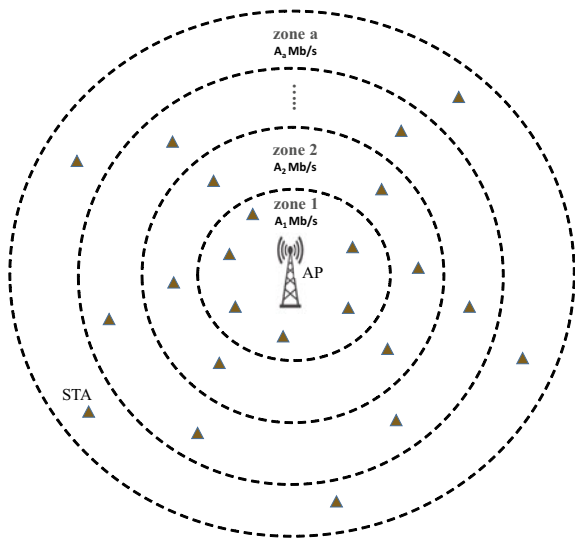
2 Related Work

From the related works of IEEE 802.11ah, the anomaly issue was discussed in [8–11]. A mathematical model developed in [11] and proposed a scheme to overcome the anomaly. The authors of [10] reported the performance anomaly in IEEE 802.11ah. In [9], the authors used a machine leaning approach to solve the anomaly for IEEE 802.11ah network. The authors of [8] proposed a scheme to enhance the performance of the multi-rate IEEE 802.11ah network.

3 Network Architecture

We consider an AP and N STAs around the AP for IEEE 802.11ah network as shown in Fig. 1. It supports a distinct data rates depending on distance $S_d^{(A,d)} | d \in [1, a]$ from AP. The transmission range of AP is classified into a zones. Here, $A_d | d \in [1, a]$ are

Fig. 1 Network architecture



supported rates. The zone 1 is closer the AP with high rate STAs, whereas zone a is at the fringes with lower rate STAs. The STAs in particular zone transmit based on MCS using rate adaptation. We consider no hidden STAs with network in saturation. The STAs are divided into a groups with g_d ($\sum_{d=1}^a g_d = N$). Each group has equal number of STAs. δ_R is RAW period, and $\delta_{t,q}$ is the q^{th} RAW slot length [12], and the RAW slots are of equal length. The $\delta_{t,q}$ is classified into effective duration ($\delta_{t',q}$) and holding period (δ_h). The achievable data rate of a STA with rate $A_d | d \in [1, a]$ is given by[10],

$$A(S_d^{(A_d)}) = P_t + G_t + G_r - PL(S_d^{(A_d)}) - N_0 - F - \left(\frac{E_b}{N_0} \right). \quad (1)$$

where G_t denotes transmitter (Tx) gain, G_r is receiver (Rx) gain, P_t is Tx power, N_0 is noise spectral density, fade margin is F , $\left(\frac{E_b}{N_0} \right)$ is energy per bit, and path loss $PL(S_d^{(A_d)}) = 8 + 37.6 \log_{10}(S_d^{(A_d)})$.

4 Throughput Analysis

In IEEE 802.11ah, STAs compete for channel using EDCA relies on DCF. Initially, STAs sense the channel and waits for DIFS (δ_f). In idle state, the counter decrements in the RAW slot. In busy state, counter freezes and starts back-off in the range $[0, CW_0 - 1]$, with CW_0 of minimum contention window (CW). For every collision, the back-off stage increases till maximum CW of $W_m = 2^m CW_0$, where m is equal

to retry limit (R) [6]. The channel is classified into mini-slots of δ_m length. The transmission probability [13] in zone- a of rate $A_d|d \in [1, a]$ is expressed as

$$\tau_q^{(A_d)} = \frac{2(1 - 2p_{c,q}^{(A_d)})(1 - (p_{c,q}^{(A_d)})^{R+1})}{\left(CW_0(1 - (2p_{c,q}^{(A_d)})^{m+1})(1 - p_{c,q}^{(A_d)}) + (1 - 2p_{c,q}^{(A_d)})(1 - (p_{c,q}^{(A_d)})^{R+1}) \right. \\ \left. + 2^m CW_0(1 - 2p_{c,q}^{(A_d)})(p_{c,q}^{(A_d)})^{m+1} \times (1 - (p_{c,q}^{(A_d)})^{R-m}) \right)}, \quad (2)$$

where $p_{c,q}^{(A_d)}$ is the collision probability given by [8]

$$p_{c,q}^{(A_d)} = \frac{(1 - (1 - \tau_q^{(A_d)})^{g_q^{(A_d)} - 1}) \prod_{y=1, y \neq d}^a (1 - \tau_q^{(A_y)})^{g_q^{(A_y)}}}{(1 - E_r)^{-1}}, \quad (3)$$

Here, $E_r = 0.1$ is packet error rate. The success probability of k^{th} group STA of rate $A_d|d \in [1, a]$ in q^{th} RAW slot is expressed as

$$p_{s,q}^{(A_d)} = \frac{g_q^{(A_d)} \tau_q^{(A_d)} (1 - \tau_q^{(A_d)})^{g_q^{(A_d)} - 1}}{\prod_{y=1, y \neq d}^a (1 - \tau_q^{(A_y)})^{g_q^{(A_y)}}} p_{\text{tr},q}^{(A_d)} (1 - E_r)^{-1}, \quad (4)$$

where $p_{\text{tr},q}$ is given by

$$p_{\text{tr},q}^{(A_d)} = 1 - \prod_{y=1}^a (1 - \tau_y^{(A_y)})^{g_q^{(A_y)}}. \quad (5)$$

The sum of a $\delta^{(A_d)}$, δ_f , and back-off slots spent gives a transaction. The probability of Δ_Ω with Ω -transactions in the RAW slot which use Λ mini-slots [12] is given by

$$P_{\Delta_\Omega}(\Lambda) = \binom{\Lambda - \Omega t' - 1}{\Lambda - \Omega t' - \Omega} (p_{\text{tr},q}^{(A_d)})^\Omega (1 - p_{\text{tr},q}^{(A_d)})^{\Lambda - \Omega t' - \Omega}, \quad (6)$$

where $t' = t^{(A_d)} + \delta_f$ and $t^{(A_d)}$ gives transmission duration. Here, H is random variable gives transactions during $\delta_{t',q}$. The probability of Ω -transactions in q^{th} RAW slot is given by

$$P_{H,q}(\Omega) = \sum_{y=\Omega}^{\delta_{t',q} - (\Omega-1)t' - \delta_f - \delta_m} \text{Prob} \left\{ \sum_{q=1}^{\Omega} \Lambda_{b,q} = y \right\}. \quad (7)$$

where $\Lambda_{b,\Omega}$ is mini-slots spent and $\mathbb{N}_t = \lfloor \frac{\delta_{t',q}}{t' + \delta_m} \rfloor$. Therefore, the mean of H 'in q^{th} RAW slot is given by,

$$E_q[H] = \sum_{\Omega=1}^{\mathbb{N}_t(\delta_{t',q} - \delta_m)} \Omega P_{H,q}(\Omega). \quad (8)$$

The average throughput in q^{th} RAW slot with rate $A_d|d \in [1, a]$ is expressed as [12]

$$U_q^{(A_d)} = \frac{a E_y[H] p_{s,q}^{(A_d)} E[P]}{\delta_R}. \quad (9)$$

The aggregate throughput with UG scheme of rate $A_d|d \in [1, a]$ in the q^{th} RAW slot is given by

$$U_q = \sum_{d=1}^a U_q^{(A_d)}. \quad (10)$$

The data transferred per zone in q^{th} RAW slot is

$$UD_q^{(A_d)} = \frac{a E_q H p_{s,q}^{(A_d)} E[P]}{\delta_R} (\delta_{t',q} + \delta_h). \quad (11)$$

The energy efficiency (EE) is given by [8].

$$E_t^{(A_d)} = \frac{E_{\text{id}} + E_{s,q}^{(A_d)} + E_{c,q}^{(A_d)}}{E[P]}. \quad (12)$$

where E_{id} is the energy spent in idle state, $E_{s,q}^{(A_d)}$ gives energy in successful, and $E_{c,q}^{(A_d)}$ is energy during collision.

4.1 Performance Anomaly-RSA Scheme

This section discusses the mitigation of performance anomaly and further enhance the throughput of IEEE 802.11ah network. We propose a RSA scheme to enhance the throughput of high rate STAs. The RSA scheme is shown in Algorithm 1. Initially, the achievable rates are computed by each STA and communicate to the AP prior the initialization of network. After that, AP forms the groups depending on rate of each STA. The high rate STAs are assigned with more RAW slots and low rate STAs with minimum RAW slots because low rate STAs consume more channel time [8]. Suppose, if a RAW period is classified into 10 RAW slots from slot 0 to 9. RAW slots 0, 4, 7, 9 assigned with high rate STAs, and RAW slot 3 is assigned with low rate STAs. Then, aggregate throughput is computed with the RAW slots assigned to the group.

Algorithm 1 RSA Scheme

- 1: Initialize the network parameters T_{RAW} , N , number of zones;
 - 2: Compute the achievable data rates using Eq. (1) and share with the AP;
 - 3: AP groups the STAs with same data rates;
 - 4: Assign the high rate STAs with more number of RAW slots and low rate STAs with minimum number of RAW slots;
 - 5: Compute the throughput of high and low rate STAs;
-

5 Results and Discussion

The analytical (A) works executed with MATLAB 2020a and simulations (S) using ns-3 are discussed in this section [14]. The system parameters are shown in Table 1. The IEEE 802.11ah multi-rate network is considered with 1000 STAs. For result analysis, the transmission range is categorized into four zones with data rates shown in Table 2. The zone 1 has high rate STAs, while zone 4 has low rate STAs. The simulations are executed with a confidence interval of 95%.

Figure 2 shows the analysis of anomaly reduction using RSA scheme and UG scheme (distinct rate STAs). The anomaly issue is resolved by the RSA scheme. With the Algorithm 1, the zone-1 STAs reserve with more RAW slots than the zone-

Table 1 System parameters

Parameter	Value
$E[P]$	64 bytes
Header	268 bytes
ACK	112 bits
δ_f	264 μ s
δ_s	160 μ s
δ_m	52 μ s
δ_R	500 ms
Fade margin, F	9.77 dB
Tx radio frequency	900 MHz
Tx gain, G_t	3 dB
Rx gain, G_r	0 dB
P_{id}	3 mW
Tx power P_t	10 mW

Table 2 Achievable data rates with zone lengths

Zones	1	2	3	4
A_d (Mbps)	78	8.77	1.95	0.6
$S_d^{(A_d)}$ (m)	115	210	307	420

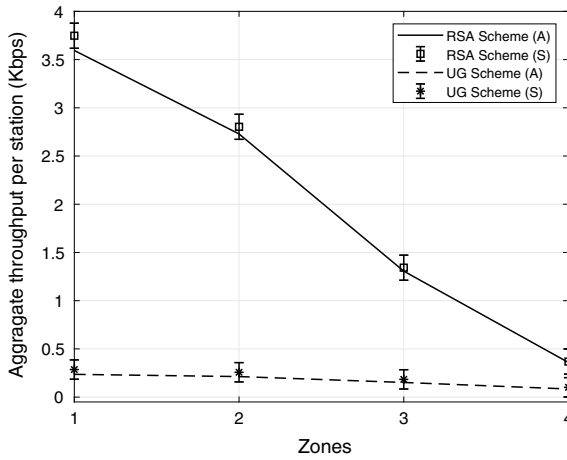


Fig. 2 Aggregate throughput

Table 3 Data transferred per zone

Zone	Data transferred (Kb)			
	RSA scheme		UG scheme	
	A	S	A	S
1	112.3351	129.6190	7.3882	8.9389
2	85.2842	87.6023	6.6574	8.0548
3	40.8239	41.9549	4.7517	5.7539
4	11.2117	11.5243	2.6304	3.1850

4 STAs (low rate) depending on transmission rates. Now, the zone-1 (high rate) STAs can effectively use the channel. From Fig. 2, the anomaly is resolved using the RSA scheme compared to UG scheme.

Table 3 gives the data transferred with respect to zones. Because of the more RAW slots assigned to zone-1 STAs, the data transfer also enhanced in IEEE 802.11ah network. In UG scheme, the data transfer is comparatively lesser than RSA scheme. From Fig. 2 and Table 3, the throughput and data transfer of IEEE 802.11ah using RSA scheme outperform the UG scheme. Table 4 shows EE with respect to zones. From Table 3, we can notice the RSA scheme gives better EE over UG scheme because the zone-4 STAs consume more energy due to lower rate transmissions.

Table 4 Energy efficiency

Zone	EE (PJ/bit)			
	RSA scheme		UG scheme	
	A	S	A	S
1	8.4522	8.4523	135.2343	135.2496
2	13.3358	13.3360	160.0299	160.0447
3	32.2169	32.2215	257.7348	257.7496
4	135.1068	135.1074	540.4271	540.4415

6 Conclusion

IEEE 802.11ah standard is developed to intended to support IoT applications. In IEEE 802.11ah, UG scheme is default scheme having distinct rate STAs in each group. The performance anomaly exists in UG scheme of IEEE 802.11ah, and it is observed that the lower rate STAs penalize the high rate STAs. We have presented a RSA scheme to resolve the anomaly and enhance the IEEE 802.11ah network's performance. Using RSA scheme, the RAW slots are assigned with respect to data rates of STAs. From results, the performance of IEEE 802.11ah is enhanced using RSA scheme compared to UG scheme. We used ns-3 for simulations to corroborate the analytical results.

References

1. Zhou I, Makhdoom I, Shariati N, Raza MA, Keshavarz R, Lipman J, Abolhasan M, Jamalipour A (2021) Internet of Things 2.0: concepts, applications, and future directions. *IEEE Access* 9:70,961–71,012. <https://doi.org/10.1109/ACCESS.2021.3078549>
2. Khorov E, Kiryanov A, Lyakhov A, Bianchi G (2019) A tutorial on IEEE 802.11ax high efficiency WLANs. *IEEE Commun Surveys Tuts* 21(1):197–216. <https://doi.org/10.1109/COMST.2018.2871099>
3. Harigovindan VP, Babu AV, Jacob L (2014) Proportional fair resource allocation in vehicle-to-infrastructure networks for drive-thru Internet applications. *Comput Commun* 40:33–50
4. IEEE Standard for Information Technology (2017) Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements—Part 11: wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications amendment 2: sub 1 GHz license exempt operation. *IEEE Std 802.11ah—2016*, pp 1–594. <https://doi.org/10.1109/IEEESTD.2017.7920364>
5. Baños-Gonzalez V, Afaqui MS, Lopez-Aguilera E, Garcia-Villegas E (2016) IEEE 802.11ah: a technology to face the IoT challenge. *Sensors* 16(11). <https://doi.org/10.3390/s16111960>
6. Chatzimisios P, Boucouvalas AC, Vitsas V (2003) IEEE 802.11 packet delay—a finite retry limit analysis. In: *GLOBECOM'03*, vol 2, pp 950–954
7. Heusse M, Rousseau F, Berger-Sabbatel G, Duda A (2003) Performance anomaly of 802.11b. In: *IEEE INFOCOM 2003*, vol 2, pp 836–843. <https://doi.org/10.1109/INFCOM.2003.1208921>

8. Badarla SP, Harigovindan VP (2021) Restricted access window-based resource allocation scheme for performance enhancement of IEEE 802.11ah multi-rate IoT networks. *IEEE Access* 9:136, 507–136, 519 (2021). <https://doi.org/10.1109/ACCESS.2021.3117836>
9. Mahesh M, Pavan BS, Harigovindan V (2020) Data rate-based grouping using machine learning to improve the aggregate throughput of IEEE 802.11ah multi-rate IoT networks. In: 2020 IEEE international conference on advanced networks and telecommunications systems (ANTS), pp 1–5. <https://doi.org/10.1109/ANTS50601.2020.9342758>
10. Pavan BS, Mahesh M, Harigovindan V (2020) Performance anomaly of group-synchronized distributed coordination function in IEEE 802.11ah based multi-rate IoT networks. In: 2020 5th International conference on computing, communication and security (ICCCS), pp 1–5. <https://doi.org/10.1109/ICCCS49678.2020.9277378>
11. Sangeetha U, Babu A (2020) Fair and efficient resource allocation in IEEE 802.11ah WLAN with heterogeneous data rates. *Comput Commun* 151:154–164. <https://doi.org/10.1016/j.comcom.2019.12.043>
12. Zheng L, Ni M, Cai L, Pan J, Ghosh C, Doppler K (2014) Performance analysis of group-synchronized DCF for dense IEEE 802.11 networks. *IEEE Trans Wirel Commun* 13(11):6180–6192. <https://doi.org/10.1109/TWC.2014.2337315>
13. Pavan BS, Harigovindan VP (2021) A novel channel access scheme for NOMA based IEEE 802.11 WLAN. *Sādhanā* 46(144)
14. Tian L, Šljivo A, Santi S, De Poorter E, Hoebeke J, Famaey J (2018) Extension of the IEEE 802.11ah Ns-3 simulation module. In: Proceedings of 10th workshop on Ns-3, WNS3'18, pp 53–60. ACM, New York, NY, USA. <https://doi.org/10.1145/3199902.3199906>

Optimal Allocation of Micro-phasor Measurement Units in Distribution Network Considering Security Constraints



Manam Ravindra, Donepudi Tata Rao, Rayapudi Srinivasa Rao, Adireddy Ramesh, and Karri Manoz Kumar Reddy

Abstract This paper presents optimal allocation of Micro-Phasor Measurement Units (M-PMU) in distribution network using Binary Integer Linear Programming (BILP) method considering security constraints. Placement of PMUs at every node of feeder network is infeasible as it is highly economical. So, to minimize the number of M-PMU's and to locate at optimal places, observability constraints are considered in one case, and allocation of M-PMUs at critical nodes in priority is considered in another case as security constraints. M-PMU allocation in case of single line outage or PMU loss is formulated. A node observability index (NOI) is proposed to check the observability of every node in the network. Complete feeder network observability (CFNO) is proposed to check performance of observability of complete network. MATLAB simulations are considered for IEEE-13, 33, 37, 69 feeder network, and obtained results are compared with standard methods to show its efficacy.

Keywords Critical nodes · M-PMU · Observability · Security constraints

1 Introduction

Today, Phasor Measurement Units (PMUs) are almost used for complete observability of transmission system. When transmission system and distribution system are compared, voltage angle variation between buses on a distribution system will be lesser in magnitude than that on transmission system [1]. Errors, like instrumental errors and white noise, affect the distribution measurements compared to transmission system. These Micro-Phasor Measurement Units (M-PMU) considered in this paper can be produced at cheaper cost, than current commercial PMUs used transmission network [2]. As these M-PMUs are cheaper compared to commercial PMUs, these could be deployed and bring a much higher resolution to the distribution grid

M. Ravindra (✉) · D. Tata Rao · A. Ramesh · K. Manoz Kumar Reddy
Aditya College of Engineering, Surampalem, India
e-mail: ravieejntu@gmail.com

R. Srinivasa Rao
Jawaharlal Nehru Technological University, Kakinada, India

[1, 2]. Therefore, a M-PMU-based sensor is proposed to provide a real-time monitoring for the distribution system. The M-PMU technology provides fast and accurate, network-wide synchronized measurements of voltage phasors and currents up to 512 samples per nominal 50/60 Hz cycle [2, 3]. For optimal deployment of PMU devices in power system network, two methods are followed: (1) classical method and (2) evolutionary method. When evolutionary methods are considered for solution, different solutions are obtained each time when the program was run. But the classical method gives only one solution every time. The author in Ref. [4] proposed BGWO method for optimal PMU placement considering redundancy in distribution network. The author in Ref. [5] proposed upgraded BGWO for optimal allocation of PMUs in transmission network to attain complete observability. The author in Ref. [6] proposed the ILP method for optimal placement PMUs in distribution network considering topology changes in network. The author in Ref. [7] proposed hybrid ILP method for optimal location of PMU devices and attained maximum redundancy. The author in Ref. [8] proposed ILP method for deployment of PMUs and branch links in distribution network. The author in Ref. [9] proposed binary harmony search algorithm for optimal deployment PMUs in transmission network considering security constraints. In this paper, M-PMUs are considered for optimal deployment in distribution network considering security constraints to achieve complete observability.

2 Problem Formulation

2.1 Formulation for Micro-PMU Placement in Distribution Network

The problem is defined to optimize the number of M-PMUs to observe the state of the network in order to provide real-time protection to the network. Economically, it is not feasible to install M-PMUs at every node of the network. So in order to minimize the cost, the M-PMUs that are to be located have to be minimized. The problem is designed to deploy M-PMUs in feeder network in a process to attain complete observability of the network. The optimal M-PMU deployment is formulated as follows:

$$\min \sum_{p=1}^N W_p N_p \quad (1)$$

$$\text{Subject to } N(p) = \begin{cases} N_1 = n_1 \geq 1 \\ N_2 = n_2 + n_3 \geq 1 \\ N_3 = n_2 + n_3 + n_4 \geq 1 \\ N_4 = n_3 + n_4 + n_5 \geq 1 \\ N_5 = n_4 + n_5 + n_6 \geq 1 \\ N_6 = n_5 + n_6 \geq 1 \\ N_7 = n_7 + n_8 \geq 1 \\ N_8 = n_7 + n_8 + n_9 + n_{10} \geq 1 \\ N_9 = n_8 + n_9 + n_{13} + n_{10} + n_4 \geq 1 \\ N_{10} = n_9 + n_{10} + n_{11} \geq 1 \\ N_{11} = n_{10} + n_{11} \geq 1 \\ N_{12} = n_8 + n_{12} \geq 1 \\ N_{13} = n_9 + n_{13} \geq 1 \end{cases} \quad (2)$$

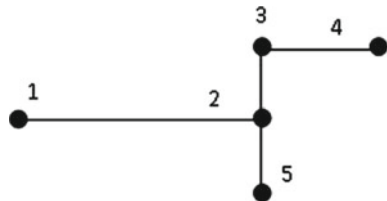
where N is number of buses and W_p presents cost coefficient for M-PMU deployed at bus p .

$$N_p = \begin{cases} 1 & \text{if micro-PMU is deployed at bus } p \\ 0 & \text{otherwise} \end{cases}$$

2.2 Optimal Deployment Considering Redundancy in Network

For instance, from Fig. 1 it can be observed that node 2 is connected with three branches, i.e., (2,1), (2,5) and (2,3), and (2,2) the node itself. Totally, the redundancies of measurements considered for node 2 are 4. Here the node can be considered as the most critical node for deployment of M-PMU. If the M-PMU is installed at node 2, the nodes 2, 1, 5 and 3 are observable. Therefore, deployment of M-PMUs at maximum redundant node minimizes the number of locations for placement of M-PMUs. The OPP is formulated as follows:

Fig. 1 Five-node feeder network



$$\text{Min} \sum_{p=1}^N W_p N_p \quad (3)$$

$$\text{Subject to } \begin{matrix} NP \geq B, \\ NP = B \end{matrix} \quad (4)$$

Here the constraint $NP = B$ allocates the PMU at maximum redundant node, where $B = [1, 1, 1, 1, \dots]^T$.

Where N is number of buses, W_p presents cost coefficient of M-PMU located at bus p .

$$N_p = \begin{cases} 1 & \text{if M - PMU is allocated at bus } p \\ 0 & \text{otherwise} \end{cases}$$

2.3 Optimal Deployment Considering Single Line Outage/Single M-PMU Loss

Single line outage problem is important contingency case to be considered for deployment of M-PMU in network. To minimize such problem in distribution network, every node should be observed twice by at least two numbers of M-PMUs. The problem is formulated as follows:

$$\text{min} \sum_{p=1}^N W_p N_p \quad (5)$$

$$\text{Subject to } NP \geq B \quad \text{where } B = [2, 2, 2, 2, \dots]^T \quad (6)$$

2.4 Performance Check Through Complete Feeder Network Observability

Measurement redundancy is to be calculated for every node in the network. In this paper, CFNO and NOI are proposed to check the measurement redundancy of the network to estimate complete network observability. Node observability index (NOI) presents the number of times the node is observable by PMU device. CFNO gives the summation result of all the nodes NOI.

$$\text{NOI}_i = N^T P_i \quad (7)$$

$$CFNO = \sum_{p=1}^N NOI_p \tag{8}$$

CFNO shows the performance of observability of the network with M-PMU locations.

3 Optimal Allocation of Micro-PMUs with Binary Integer Linear Programming Considering Redundancy in Measurements

Complete observability of network is defined as if all nodes are directly or indirectly observed by the located M-PMU devices in network. The deployment of M-PMUs in network is considered with BILP method.

The procedure is shown in flowchart. Figure 2 shows the procedure for optimal allocation of M-PMUs in the distribution network considering security constraints to obtain.

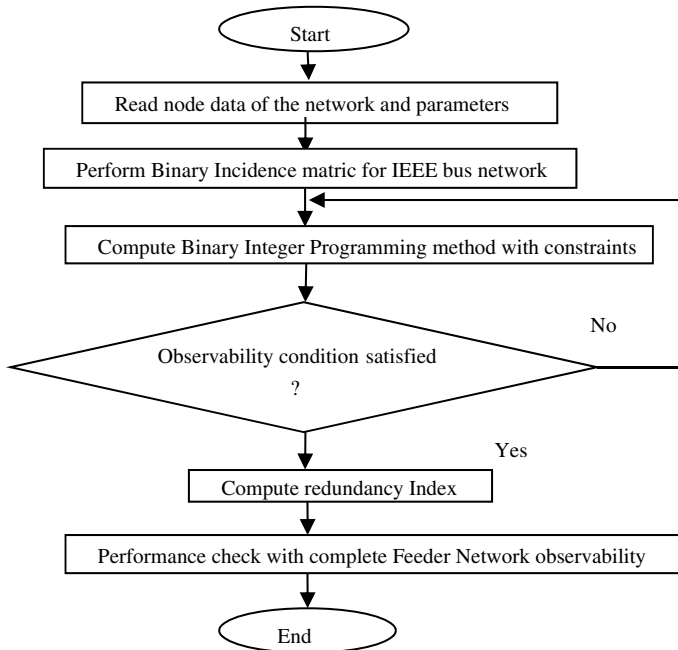


Fig. 2 Flowchart for optimal Micro-PMU placement using BILP

4 Results and Analysis

The optimal deployment of M-PMU devices in distribution network for different test case systems is formulated using BILP approach and simulated using MATLAB programming. Test cases 13, 33, 37 and 69 bus systems are run on Intel(R) core(TM), i3 processor at 2.20 GHz, 4 GB of RAM.

Table 1 shows the feeder nodes that have highest redundancy. The nodes that are with highest redundancy are considered as critical nodes for prior deployment of Micro-PMU. Table 2 shows the optimal deployment of Micro-PMUs with BILP method. Table 3 shows optimal deployment of Micro-PMUs with BILP method considering redundancy constraints along with security constraints, i.e., complete observability. Table 4 shows node observability index for 13 feeder network.

Table 1 Feeder nodes having highest redundancy

IEEE feeder test case system	Buses having highest redundancy	Number of nodes in network
13-node test feeder	4, 9, 8	3
33-node test feeder	2, 3, 6	3
37-node test feeder	3, 6, 4, 9, 35, 26, 30, 22, 12, 19, 15	11
69-node test feeder	3, 4, 7, 8, 10, 11	6

Table 2 Optimal deployment of M-PMUs with BILP method

IEEE feeder test case systems	Optimal deployment of M-PMUs	Number of nodes in network
13-node test feeder	3, 4, 5, 8, 9, 10	6
33-node test feeder	2, 5, 8, 11, 14, 17, 21, 24, 27, 30, 32	11
37-node test feeder	2, 3, 6, 9, 14, 15, 19, 22, 27, 29, 30, 33, 34, 35	14
69-node test feeder	2, 6, 9, 14, 17, 20, 23, 26, 29, 32, 34, 37, 40, 43, 45, 47, 49, 51, 55, 58, 61, 64, 66, 68	24

Table 3 Optimal deployment of M-PMUs with BILP method considering redundancy constraints

IEEE feeder test case systems	Optimal deployment of M-PMUs	Number of nodes in network
13-node test feeder	3, 4, 5, 8, 9, 10	6
33-node test feeder	2, 3, 6, 9, 12, 15, 17, 21, 24, 28, 31, 32	12
37-node test feeder	2, 3, 4, 6, 9, 12, 15, 19, 22, 26, 29, 30, 33, 34, 35	15
69-node test feeder	1, 3, 4, 7, 8, 10, 11, 14, 17, 20, 23, 26, 30, 33, 35, 38, 41, 44, 46, 49, 52, 54, 57, 60, 63, 65, 67, 69	28

Table 4 Node observability index (NOI) for 13 feeder network

N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13
1	1	2	4	2	1	1	2	3	2	1	1	1

Table 5 presents the optimal deployment of Micro-PMUs considering single line contingency, and Table 6 presents the complete feeder network observability of the system. Table 7 shows the comparative analysis of the proposed method with other standard methods.

Table 5 Optimal deployment of Micro-PMUs considering single line contingency

IEEE feeder test case systems	Optimal deployment of M-PMUs	Number of nodes in network
13-node test feeder	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13
33-node test feeder	1, 2, 3, 5, 6, 9, 11, 12, 14, 15, 17, 18, 20, 21, 22, 24, 25, 26, 27, 29, 30, 32, 33	24
37-node test feeder	1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37	31
69-node test feeder	1, 2, 4, 6, 7, 9, 10, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 27, 28, 29, 31, 32, 34, 35, 36, 37, 39, 40, 42, 43, 45, 46, 47, 49, 55, 57, 58, 60, 61, 63, 64, 65, 66, 67, 68, 69	50

Table 6 Complete feeder network observability

Test case network	CFNO	Redundancy
13 feeder	22	9
33 feeder	39	6
37 feeder network	52	15
69 feeder network	82	13

Table 7 Comparative analysis of the results

Methods	Test feeders			
	IEEE-13	IEEE-33	IEEE-37	IEEE-69
Proposed method ILP with security constraints	6	11	12	24
BGWO[5]	6	11	12	–
Binary ILP [6]	–	11	–	
Hybrid ILP [7]	–	11	–	NA

5 Conclusion

This paper presented the optimal deployment of Micro-PMUs using Binary Integer Linear Programming considering security constraints. The node which requires highest number of measurements is considered as critical node at which the M-PMU is allocated. Single line contingency case is formulated for allocation of M-PMUs. The observability performance of complete network is computed using proposed complete feeder network observability index (CFNOI). MATLAB simulations on IEEE-13, 33, 37 and 69 feeder networks are presented, and the results are compared with standard methods to show its efficacy.

References

1. Jain A, Bhullar S (2018) Micro-phasor measurement units (μ PMUs) and its applications in smart distribution systems. In: ISGW 2017: Compendium of technical papers. Springer, Singapore, pp 81–92
2. Hojabri M et al (2019) A comprehensive survey on phasor measurement unit applications in distribution systems. *Energies* 12(23):4552
3. Sun Y et al (2017) Micro PMU based monitoring system for active distribution networks. In: 2017 IEEE 12th International conference on power electronics and drive systems (PEDS). IEEE
4. Priyadarshini S, Kumar Panigrahi C (2020) Optimal allocation of synchrophasor units in the distribution network considering maximum redundancy. *Eng Technol Appl Sci Res* 10(6):6494–6499
5. Ravindra M (2020) Complete observability constrained deployment of PMUs in power system network: using an upgraded binary Grey Wol. *Int J Future Gener Commun Netw* 13(1):1095–1105
6. Su H et al (2019) Optimal placement of phasor measurement unit in distribution networks considering the changes in topology. *Appl Energy* 250:313–322
7. Chen X et al (2019) Full coverage of optimal phasor measurement unit placement solutions in distribution systems using integer linear programming. *Energies* 12(8):1552
8. Zhao Z et al (2019) Optimal placement of PMUs and communication links for distributed state estimation in distribution networks. *Appl Energy* 256:113963
9. Ravindra M, Srinivasa Rao R, Shanmukha NagaRaju V (2020) A binary harmony search algorithm approach for security constrained allocation of PMUs to obtain complete observability. In: *Intelligent communication, control and devices*. Springer, Singapore, pp 25–35

Home Automation Using Packet Tracer



Pooja Reddy Bathula, Snigdha Pv, Laasya Lata Anumakonda,
Mohammed Mahaboob Basha, and Pandya Vyomal Naishadhkumar

Abstract With growing technological advancements, every process is becoming easier. Everything is being automated. One of the most popular automation processes include home automation. A smart home not only results in a comfortable living, but also reduces a lot of manual tedious tasks. In this paper, a smart home system is simulated with the help of Cisco packet tracer software. Cisco packet tracer allows various network elements, setup options, wireless and smart components and a testing environment. Since this project is based on Internet of Things (IoT), the IoT command is used in the Cisco packet tracer. The software provides a means for us to create a network virtually and configure the devices and later implement it in real world. Besides smart devices, Cisco packet tracer also provides sensors, actuators and microcontrollers which can be used to create the smart home automatic system.

Keywords Internet of Things · Packet tracer · Smart home · RIP

1 Introduction

With the advancement in technology in the recent years, comfort and safety have become an integral part in today's human life. Automated systems are being preferred over manual systems for the past few decades due to the comfort and ease they provide. With the increase in the number of internet users, Internet has become an irreplaceable technology in the modern-day world. IoT is the newest form of Internet technology. Internet of Things is an automated system, where computers

P. R. Bathula (✉) · S. Pv · L. L. Anumakonda · M. M. Basha · P. V. Naishadhkumar
Sreenidhi Institute of Science and Technology, Hyderabad, India
e-mail: poojabathula04@gmail.com

S. Pv
e-mail: 18311a04an@sreenidhi.edu.in

M. M. Basha
e-mail: mmahaboobbasha@sreenidhi.edu.in

P. V. Naishadhkumar
e-mail: vyomalpandya@sreenidhi.edu.in

or other mobile devices are used to operate and control simple home functions and features using the Internet from anywhere and at any time [1]. IoT refers to a system of physical devices that receive and transfer data over wireless networks with limited human intervention [2]. A home which is automated using smart objects with specific functions assigned to them is called a smart home. A smart home aims to improve the comfort, safety and efficiency of a modern-day human being. It does so by conserving the physical energy used by people to do mundane tasks and the resources needed to do them [3, 4]. This includes centralized control, status monitoring and automation of home activities within a house, using devices which are automatic and controllable [5]. For example, A smart security system can provide different features to ensure automatic security by using an alarm system, an LCD and a siren sound, in case the sensor detects any security issues [3]. The devices used in the IoT system control and monitor, electronic, mechanical and electrical systems used in different types of infrastructures. These devices are connected to the cloud and are controlled by an individual user. In this manner, many electronic devices are remotely managed through different infrastructures [6]. A smart home uses microcontroller or computer technology to handle and monitor home appliances. All the devices in the home automation system are registered at a Home Gateway and are operated by an individual user or admin. IoT is an adept technology which uses smart devices and sensors in order to optimize life in order to make it simpler and more efficient [3].

2 Literature Survey

Tabeidi et al. [7] develop the Smart College Network using a simulation concept based on Cisco packet tracer. The Home Gateway was used, which serves as a transmission media path and provides automatic addressing to multiple devices connected via wireless networks, the IoT server and smart phone, which serve as interfaces in controlling and monitoring electronic devices, and the microcontroller (MCU), which is used to register IoT devices on it to control them.

Smart home is implemented using the 2018 version of Cisco packet tracer by Shemsi [8]. Which includes different features such as sensors, IoE devices, and networking devices? To control the smart objects and sensors used in this simulation, microcontroller (MCU-PT) and Home Gateway are used. They provide a programming environment to control the connected smart objects in the system. JavaScript and Python are the programming languages used here.

Soliman et al. [9] implemented smart home by integrating IoT with web services and cloud computing. They approached this system by embedding intelligence into sensors and actuators using Arduino platform, networking smart things using ZigBee technology, facilitating interactions with smart things using Cloud services for easy access in different locations, and improving data exchange efficiency using JSON notation. This approach was used in the demonstration of services for measuring home conditions, monitoring home appliances.

Chattoraj [10] explored the architecture for an economic smart home and implemented. Arduino Uno and MATLAB GUI are used in the implementation of this project. In this project, affordable and easily available components are used. It is based on free open-source software such as Visual Basic and Arduino platform. This overall reduces the cost of implementation. This paper aims at designing a low-cost system meant to improve a standard home.

Ashok [3] utilized the home portal for home automation and monitoring smart devices, as well as a microcontroller (MCU-PT) to connect various sensors and IOE devices. MCU, on the other hand, provides a computing environment for a variety of devices and programming languages.

3 Methodology

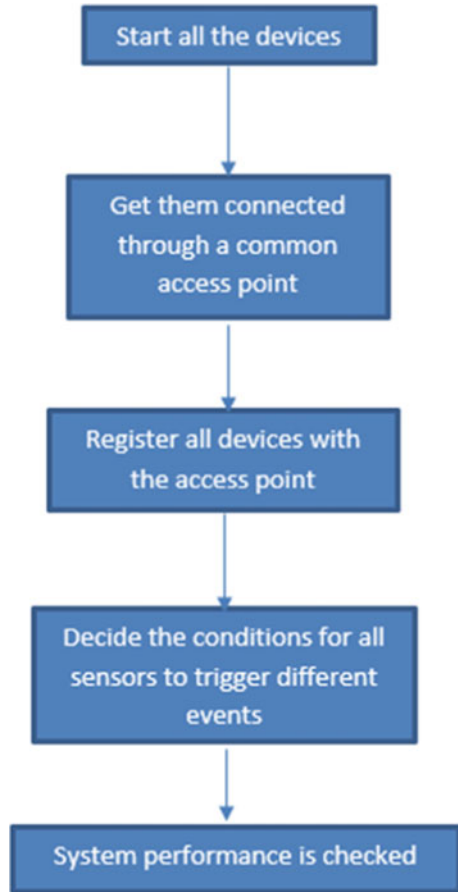
Various smart devices for the implementation of home automation are introduced. Home Gateway is used for connecting all the end devices, and smart phone is used to monitor all the connected devices. All devices are registered to Home Gateway. The Home Gateway offers four different ports and a wireless contact with a SSID. The Home Gateway is connected through WAN Ethernet port. The IP address of Home Gateway can be found in the LAN section [3]. In the flowchart of Fig. 1, the process of designing of scenario also shows the steps to perform the simulation.

In Fig. 2, the step to provide an IP address to Gateway is displayed. As shown in Fig. 3 to connect the smart devices, here it is displayed for smart light, go to the configuration of that device and connect the IoT server to Home Gateway and make sure that the network adapter is wireless in the advanced I/O settings which is shown in Fig. 4.

After registering the smart devices to Home Gateway, the devices can be accessed and remotely monitored via smart phone through web or special IoT monitor [3, 11] which can be observed in Fig. 5. As shown in Fig. 6, to access it via the web, go to smart phone and go to desktop, select web browser and enter the IP address of your Home Gateway in the URL and enter the default username and password both as “admin” and to access them via IoT monitor, go to IoT monitor in the desktop page of smart phone enter the IP address, username and password and access the linked smart devices, this is shown in Fig. 7.

To fully automate and control the devices, we need to write a few conditions in the IoT monitor page/web server page under conditions using various sensors and smart devices. After writing, the conditions and the devices become fully automated. Figure 8 shows the status of all connected devices. This way connected devices can be checked for their statuses and can be accessed as per requirements. Some conditions must be written to operate all these devices. Like, the lower and upper boundary of temperature must be set to operate the room’s window if temperature in room is crossing the upper temperature limit, we must open the window. These conditions are displayed in Fig. 9. After setting of all these parameters, final scenario of our

Fig. 1 Flowchart



work model is shown in Fig. 10. It displays total home assembly including different parts of home and parking area.

4 Simulation Results

Cisco packet tracer tool is used to simulate the home automation. In this, Fig. 11 shows the bedroom layout of the home. Motion detector is acting as a trigger for door operations. Once the motion is detected, door should be opened and if no motion is there it must be locked. In bedroom, we have 1 window which is being operated based on the input from wind detector. Along with that light is also being operated based on the inputs from wind detector. If, wind is flowing outside, window is kept in closed mode and for illumination in room lights are turned ON.

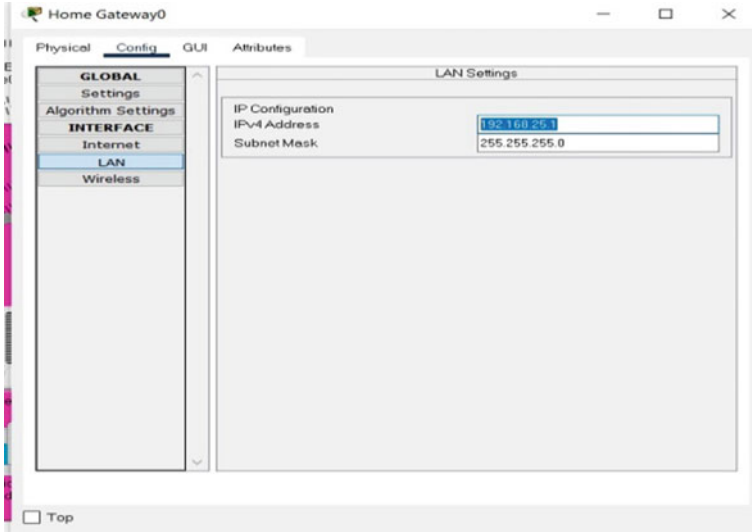


Fig. 2 Home Gateway IP address

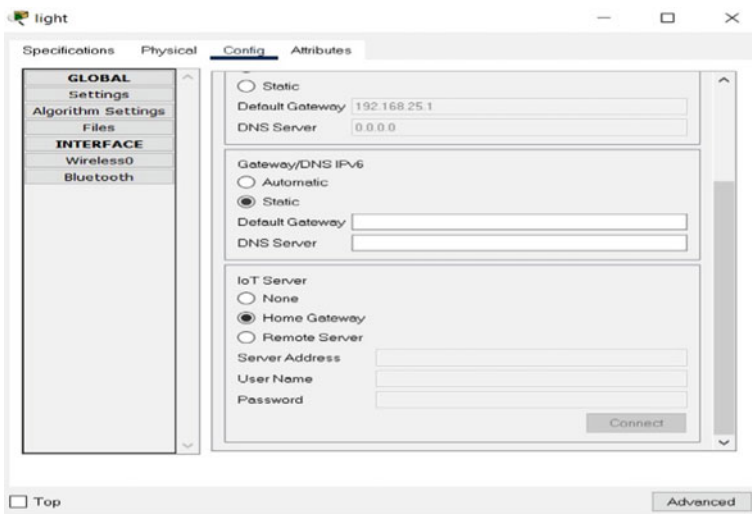


Fig. 3 Connecting the smart devices to Home Gateway through IoT server

In Fig. 12, living room appliances are shown. First we have a RFID detector, which allows only a specific person after receiving exact RFID tag readings. Face detection mechanism is also included in this project for second order of security. Temperature sensor in room checks always temperature to operate fan and AC. Portable music

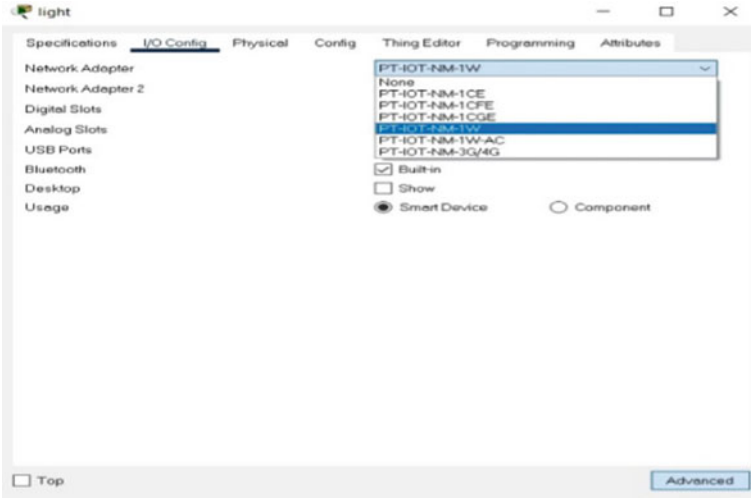


Fig. 4 Selecting the wireless network option in I/O configuration settings of the smart devices

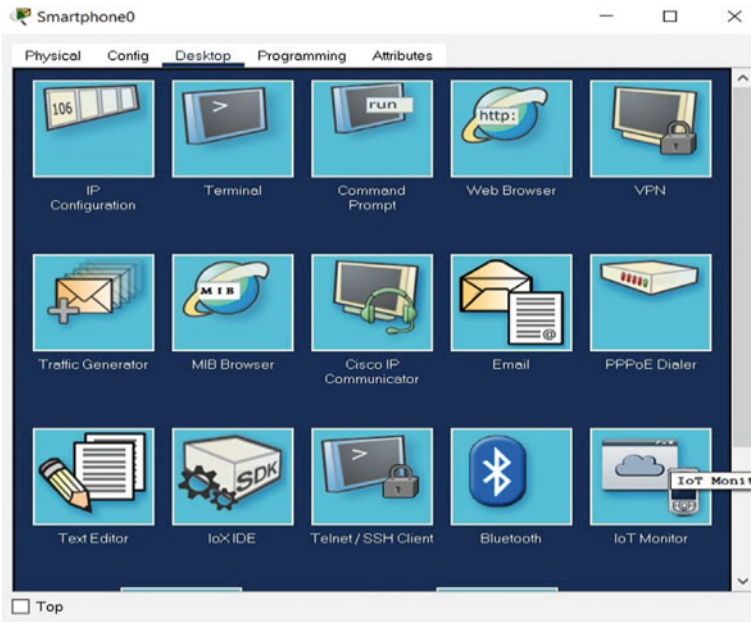


Fig. 5 Desktop window of the smart phone containing web browser and IoT monitor

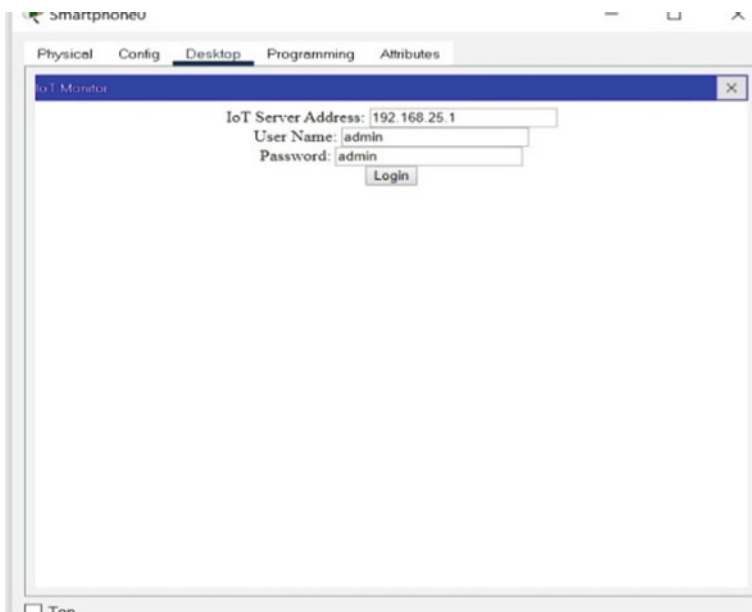


Fig. 6 IoT monitor login

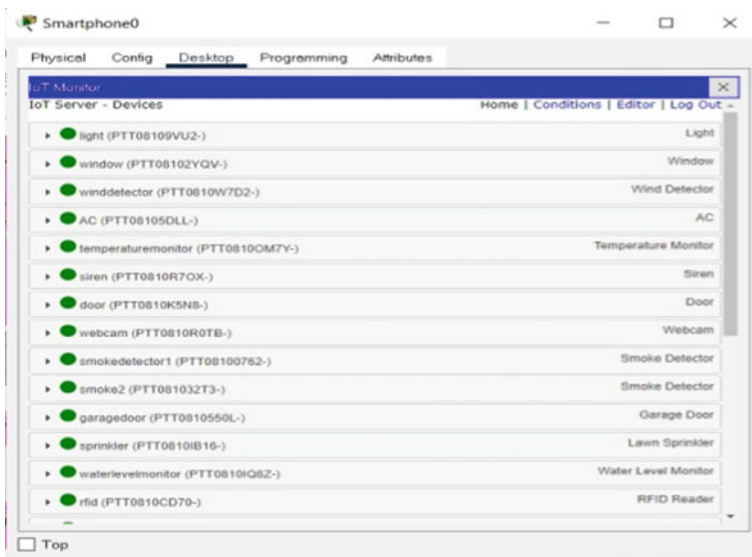


Fig. 7 Home section of IoT monitor

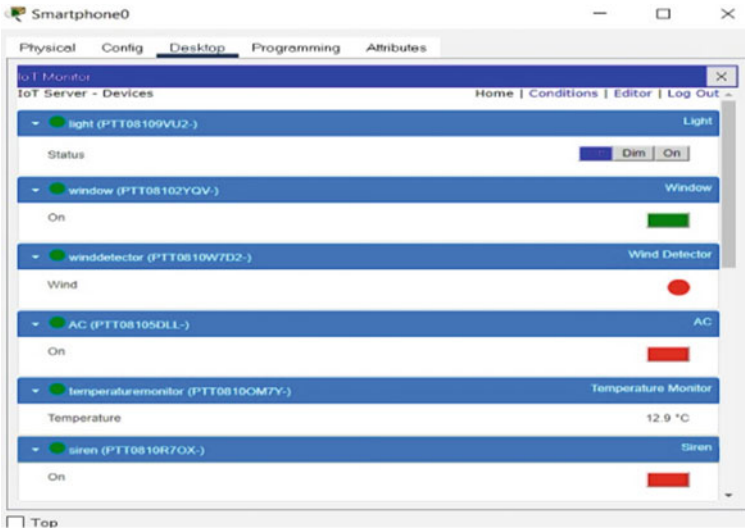


Fig. 8 Statuses of different IoT-registered devices

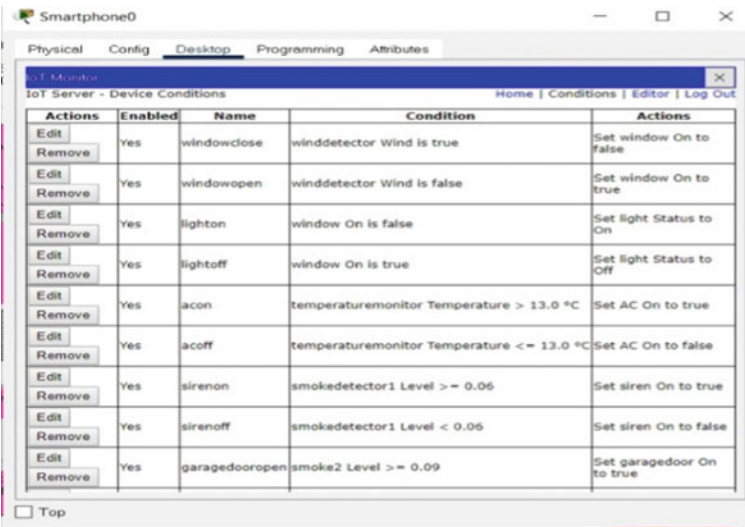


Fig. 9 Conditions written for the controlling and automation of end devices

player is used for entertainment purpose. Once the player is in operations, it will automatically turns the Bluetooth speakers on. For refreshment, coffee making machine is also kept in this scenario. It is displayed in Fig. 13, which is a motion operated machine. Motion detector will check for the motions to operate the machine. If there is no motion near the machine, it stops the coffee preparation process.

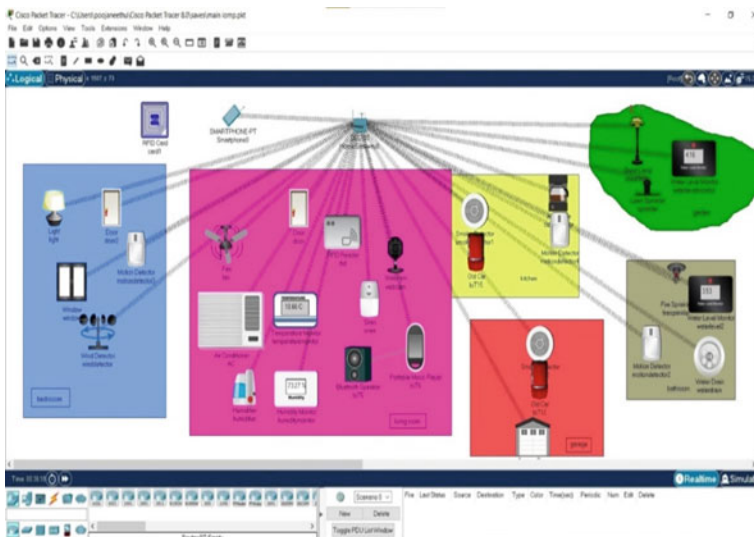


Fig. 10 Overall network layout of our smart home system

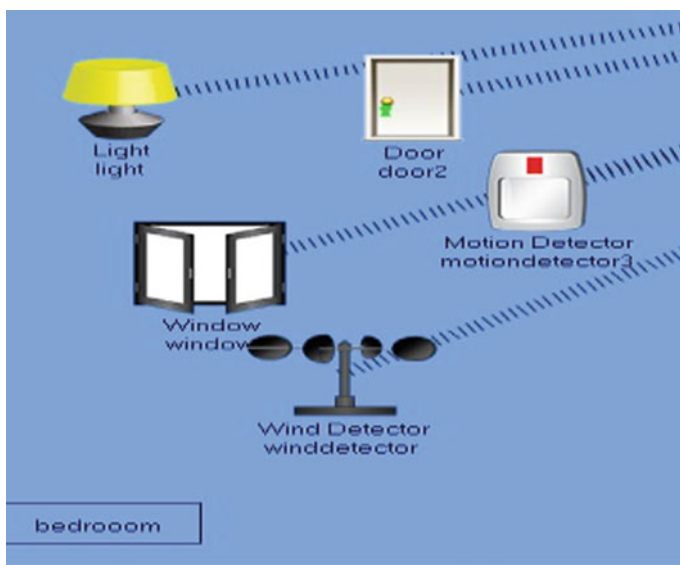


Fig. 11 Bedroom layout

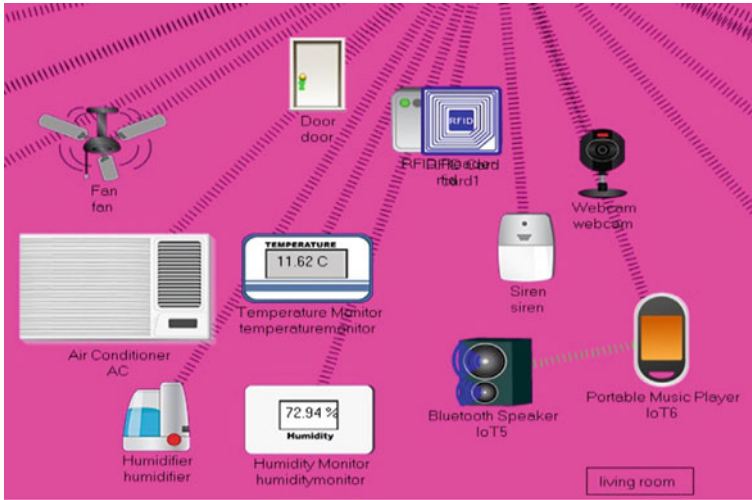


Fig. 12 Living room layout

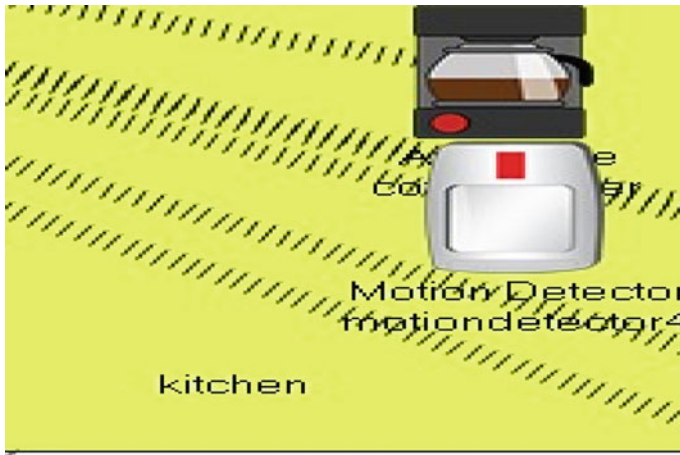


Fig. 13 Coffee machine in living room

Many users are fond of having a bath under shower during hot summer; to match this requirement, shower mechanism is established in this scenario which is visible in Fig. 14. Water sprinkler will start the water flow as the motion is detected in the motion detector. Once the person moves away from the shower, it stops the flow. For managing the outlet of water, water level indicators are used so that when a certain level of water is there, it must operate the drainage to have water out. Same water sprinkler mechanism is used in Fig. 15 to have sufficient water flow in the garden area.



Fig. 14 Shower operations in bathroom

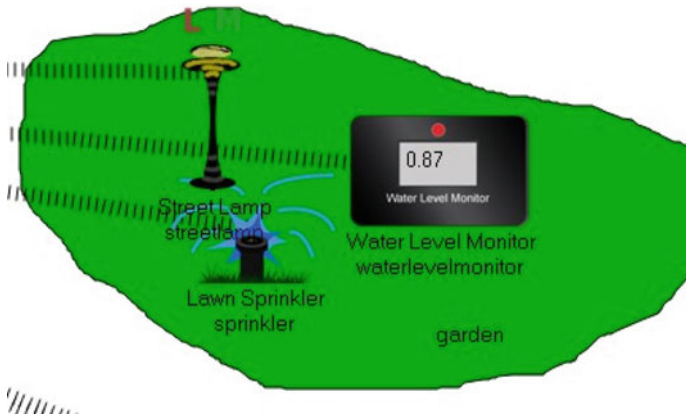


Fig. 15 Water sprinkler in garden area

Figure 16 is displaying garage layout. In this, we are detecting CO₂ emitted by the vehicle. Because in some regions of earth, vehicle must be kept in ON condition to keep engine in warm conditions. To prevent damages occurring due to excess amount of CO₂ we have used a smoke detector. This smoke detector will open and close the garage door as per the values recorded by the smoke detector.

5 Conclusion

In this paper, the need of home automation and its implementation is explained. This project has incorporated all possible aspects of different rooms and areas of a normal

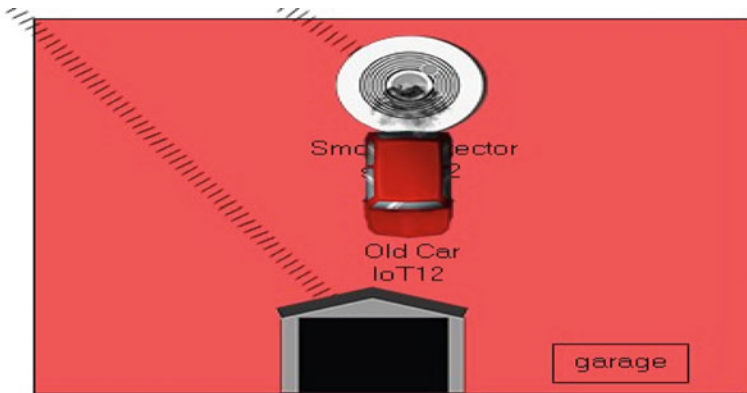


Fig. 16 CO₂ detection in garage

home structure. In this paper, we have shown how to configure the Home Gateway and all the other appliances in home. But, we have not touched the concept of power in line and range extender. This work can be carried out by adding one range extender in this scenario as it will also have more advantages compared to this scenario. All these appliances are operated using a smart phone, and it is in the range of a Home Gateway. In future work, external network may be included and remote operations can be done using them.

References

1. Robles RJ, Kim T-H (2010) Review: context aware tools for smart home development. *Int J Smart Home* 4(1):1–12
2. Vinay Sagar KN, Kusuma SM (2015) Home automation using internet of things. *Int Res J Eng Technol (IRJET)* 2(3):1965–1970
3. Ashok GLP, Saleem Akram P, Sai Neelima M, Nagasai Kumar J, Vamshi A (2020) Implementation of smart home by using packet tracer. *Int J Sci Technol Res (IJSTR)* 9(2):678–685
4. Amudha A (2017) Home automation using IoT. *Int J Electron Eng Res* 9(66):939–944
5. Majeed R, Abdullah NA, Ashraf I, Bin Zikria Y, Mushtaq MF, Umer M (2020) An intelligent, secure, and smart home automation system. *Special issue on Scientific programming towards a smart world 2020*, p 14
6. Satapathy LM, Bastia SK, Niharmohanty (2018) Arduino based home automation using Internet of things (IoT). *Int J Pure Appl Math* 118(17):769–778
7. Tabeidi RA, Masaad SM, Elshaikh BG (2019) Implementing smart college using CISCO packet tracer 7.2 simulator. *J Eng Res Appl* 9(4):34–39
8. Shemsi I (2018) Implementing smart home using cisco packet tracer simulator. *Int J Eng Sci Invent Res Dev* 4(7):261–269
9. Soliman M, Abiodun T, Hamouda T, Zhou J, Lung C-H (2013) Smart home: integrating internet of things with web services and cloud computing. In: 2013 IEEE International conference on cloud computing technology and science, pp 317–320

10. Chatteraj S (2015) Smart home automation based on different sensors and Arduino as the master controller. *Int J Sci Res Publ* 5(10):1–4
11. Lammle T (2011) *Cisco certified network associate study guide, 7th edn.* ISBN: 978-0-470-90107-6. Wiley

Interpretation of Wireless Communication Using OFDM Technology



Bipasha Chakrabarti, B. Roy, Priyanka S. Das, Prajit Paul, and A. K. Bhattacharjee

Abstract This paper presents analysis of wireless communication using the OFDM scheme. This OFDM is expecting skilled use of the obtainable band of frequencies for permitting to extend over and cover a part of between the carriers. Observing from various sides that the OFDM is the robustness modulation method which is susceptible to the higher data rate; accordingly, it will be capable of removing intersymbol interference. By using FFT or IFFT technique, we can calculate effective computation and operate. We have analyzed it in wireless communication using 16-QAM. From the simulation results, we have got this analytical result. The orthogonality among sub-carrier's frequency components can be adequately preserved. We likewise noted from the BER calculation that minimum bit error rate showed the least intersymbol interference (ISI) and gives an excellent performance for wireless communication, but the bit error rate calculation tool calculates the error rate every moment that depends on time. It does not fix the error rate.

Keywords OFDM · MIMO · QAM · SNR · BER

1 Introduction

In the technology of orthogonal frequency division multiplexing is such type of multiplexing utilizing that a higher data rate signal is split and change to many lower data rate signals, which subsequently at a similar time transmit by using different sub-carriers such that they do not impede by each other. Assume the receiver requires

B. Chakrabarti · P. S. Das

Department of Electronics and Communication Engineering, Gargi Memorial Institute of Technology, Kolkata, West Bengal, India

B. Roy (✉)

School of Electronics and Communication Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

e-mail: bappaditya13@gmail.com

P. Paul · A. K. Bhattacharjee

Department of Electronics and Communication Engineering, NIT Durgapur, Durgapur, West Bengal, India

a particular symbol within a casement of two microseconds. If the number of paths delays the carrier signal by more than two microseconds, the receiver section shall receive the symbol in the forthcoming symbol period and give rise to a significant symbol incorrectly [1, 2]. So, higher level of the data rate and higher level of the chance with multipath shall cause intersymbol interference (ISI). A distinct path to reduce the error rate should be to bring down the symbol rate; each symbol should be last longer or prolonged, and it is less unaffected by multipath. Unfortunately, this decreases the data rate also [2]. For reducing the symbol rate without slowing data rate, frequency division multiplexing (FDM) technology is used. But FDM is not a methodical bandwidth, to overcome the situation. So, OFDM is used instead of traditional FDM, and sub-carriers are generating in this way which do not require extra protection [3].

2 Proposed Idea

The OFDM concept is based on spreading the data to be transmitted over a large number of carriers, each being modulated at a low rate. The carriers are made orthogonal to each other by appropriately choosing the frequency spacing between them. In contrast to conventional FDM, the spectral overlapping among sub-carriers is allowed in OFDM since orthogonality will ensure the sub-carrier separation at the receiver, providing better spectral efficiency. This sub-carrier is as well orthogonal to each other over a set of time span (i.e., over the time span of OFDM symbol). The difference between the bandwidth use of FDM and OFDM is shown in Figs. 1, 2 and 3. MIMO transceiver block diagram is as shown in Fig. 4.

The transceiver scheme block diagram for MIMO schemes using OFDM technology contains more extra blocks than OFDM transceiver block diagram, such as sub-carrier mapping and MIMO channel estimation block. These blocks are required for MIMO processing. In accepted wireless communication manner with characteristically transmitter and single receiver, when the signal travels through a wireless domain, because of different types of a stumbling block in the way such as hill, tree, building, ship, vehicle, etc., the transmit signal goes through reflection, diffraction and the overall effects are called multipath fading; therefore, the aforesaid signal reaches the end of the line with different time and with various gains. Because of this productive and devastating thrusting, the signal phase will be shifted which takes occur [4]. This error is increased by different types of a path as the result will be faded and data rate will be reduced to overcome these outcomes of multipath fading, more than one input, and more than one output (MIMO) system represented or more than input antennas and more than output antennas will be used [5]. To get better performance and reduced fading effect, more than one antenna is used for the transmitter section as well as receiver section.

Figure 5 shows SISO is about the amply acquainted antenna arrangement than one Tx antenna and one Rx antenna; this arrangement is [MISO]. The arrangement contains over than one receiving antenna, and a single Tx aerial is called SIMO. And

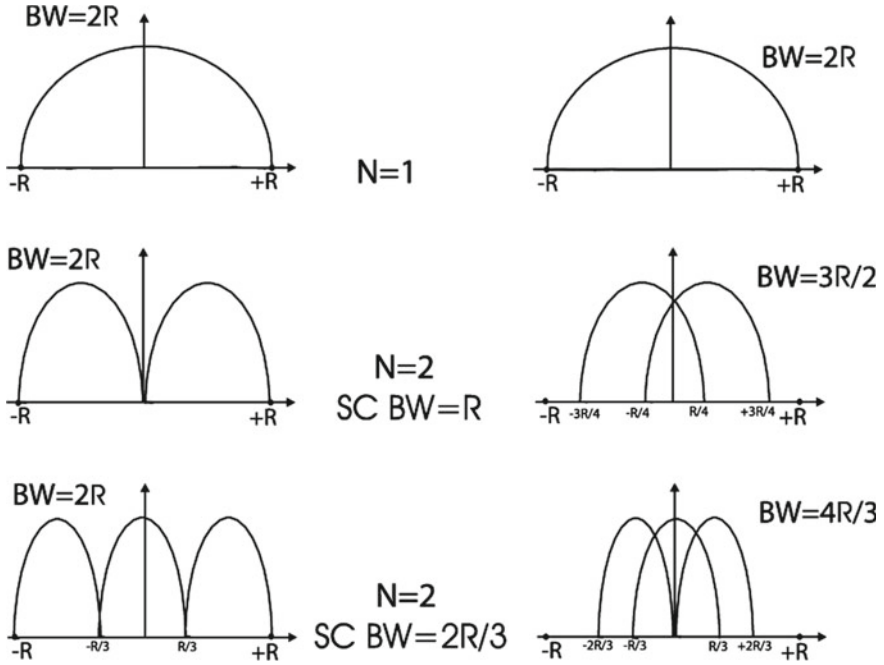


Fig. 1 Spectral proficiency

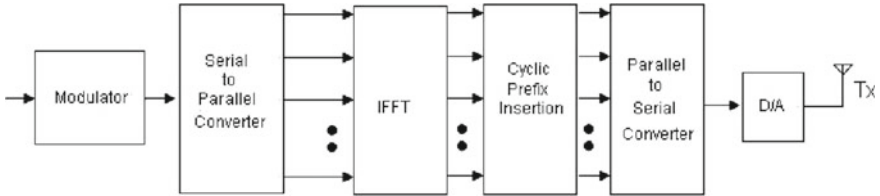


Fig. 2 OFDM Tx block diagram

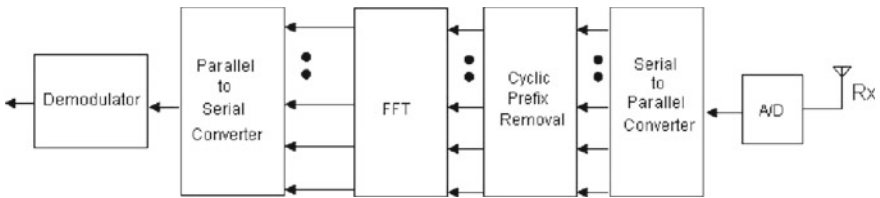


Fig. 3 OFDM Rx block diagram

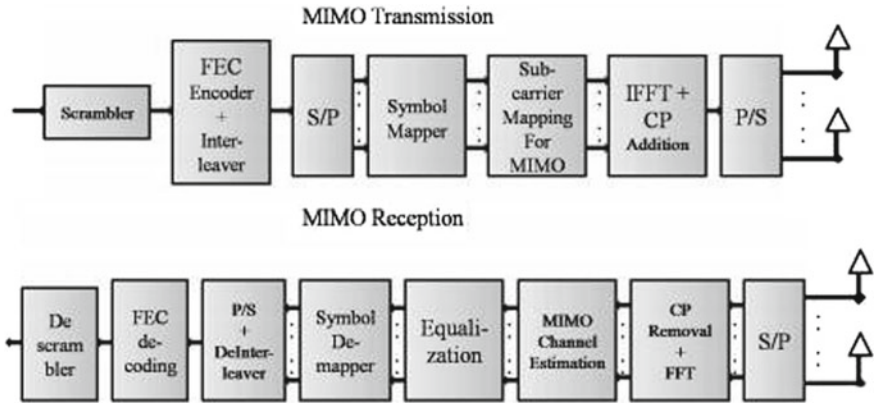


Fig. 4 Transceiver block diagram for MIMO schemes

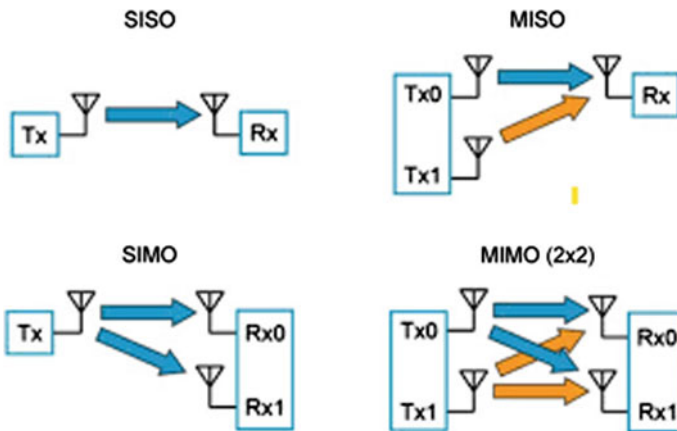


Fig. 5 Antenna configurations

finally, arrangement with more than a single transmitter section antenna and more than a single receiver section antenna is called MIMO that is multiple input multiple output [6, 7].

3 Problem Statement and Related Work

Figure 6 shows the functional block diagram of OFDM utilizing simulation:

At first describe the transmitter section; the random data generator generates the data with desirable amplitude, frequency and phase, but the generating data are changing with respect to time, then the output of random data generator is the input

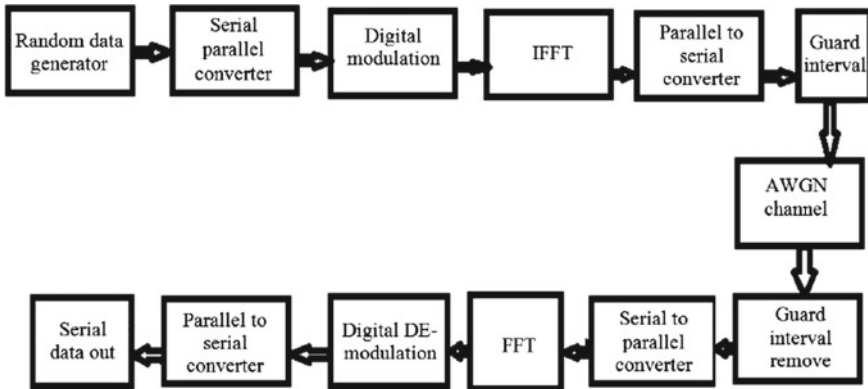


Fig. 6 Functional block diagram utilizing simulation

of serial to parallel converter; this converter is basically designed with the help of D-flip flop, then the parallel output applied to the modulator block. BER formula is used to measure the error rate from the source to the destination of digital data. Using this method or formula, we can determine how much rate of error exists in the transmitted symbol. The BER formula is given below [8].

$$BER = \frac{\text{Number of bits in error}}{\text{Total number of bits transmitted}}$$

There is the energy of each bit that will be increased for utilizing a higher robustness transmission procedure, but it has some drawbacks because it interferes with other systems. A higher bit rate reduces the energy per bit simultaneously enhance the capacity that is an important parameter for a system between this result; hopefully, the Eb/No is the shapely multiplier. That is why the bit error ratio is a figure of merit parameter in a digital communication system. Signal-to-noise ratio of the receiver and throughout the operating frequency range [9, 10]. SNR is contrary to maintain a relation with BER that is low bit error rate provides good communication signal. That is very difficult for the relation between them and BER within the multichannel environment. Therefore, SNR is utilizing to evaluate the representation of a communication method that is measured in decibels and is given below [11, 12].

$$SNR = 10 \log_{10} \left(\frac{\text{Signal Power}}{\text{Noise Power}} \right)$$

End-to-End Retransmissions (EER): Wherever the entity associations do not afford link-layer retransmissions plus fault revival–trustworthy envelope shift [13] is accomplished just through retransmissions commenced by the source node.

4 Results

We have used MATLAB software and utilized it for simulating the OFDM design. MATLAB/Simulink program will be used to simulate as well as calculate the effects with ISI and to measure the guard intervention. However, the outcome of intersymbol interference of the representation the BER becoming meaningful in several paths is fading thereby using Rayleigh channels [14, 15] as the measure of GI (guard interval) reduces, that at last, it produces the margin of error floor (Fig. 7).

- Sample time (ts) = 0.0000000005 s.
- Sample frequency = 1/ts = 2 GHz.
- Del F = 2 GHz/64 = 31.25 MHz.
- Bandwidth of sub-carriers = 31.25 MHz.
- Total bandwidth = [31.25 * 1000000] *52 = 1.625 GHz.
- Where, data sub-carriers = 48 and pilot sub-carrier = 4.
- Tg = 8 * 0.000000001 s.
- Discrete path delay = 20.
- Average path gains = 1.
- Highest Doppler shift (Hz) = 0.0001.
- Tfft = 32 * 0.000000001 s.
- Tframe = 4 * 0.000000001 s.
- Guard band(Tg) = 11.
- FFT size = 64.
- DC sub-carrier = 1.
- Sample per frame = 192.
- SNR (dB) = 20.
- Bit per symbol = 4.
- Bit rate = 2gbps.

This phenomenon of bit error rate calculations is given below.

- Error rate = 0.473.
- Number of error = 1.85e + 04 = 101.0065.

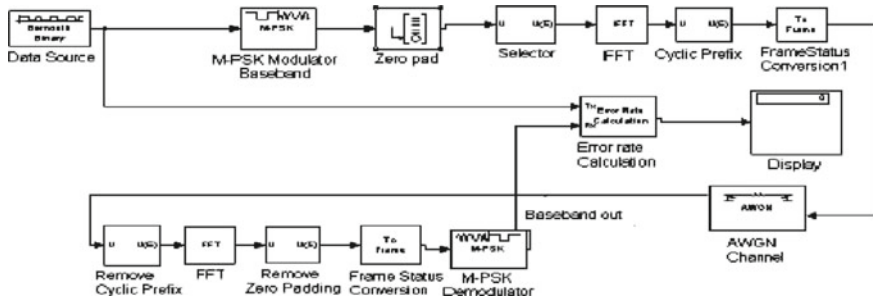


Fig. 7 OFDM system model

- Number of bits = $3.91e + 04 = 213.4787$.

It means that the error rate of bit (BER) is a measure of the number of errors of a bit that occurs within a given number of bit transmissions, but here the error rate calculator tools calculate the error rate. It will usually express as a ratio. For example, if 1-bit errors occur in 1000 bits transferred, the BER is 1/1000 or .001. This system model shows rate of error or error rate of bit (BER) = 0.473 where this numeral as BIT = $3.91e+04 = 213.4787$ Or BER = error number/numeral of bit = $101.0065/213.4787 = 0.473$. Therefore, this minimum bit error rate indicated the least inter symbol interference (ISI) and gives excellent performance for wireless communication. But this bit error rate calculation tools calculate error rate every moment that is dependent upon time. So, the error rate is not fixed like Error rate = 0.0118. Number of Error = 129. Number of Bits = 1088 (Figs. 8 and 9).

In Fig. 10, we make out that, at HHR, the gain percent is lofty with esteem to huge envelope size (i.e., $k = 512000$ bits) for Typical Go-Back-N ARQ with CRC-16, than an *Adapted MPC + PRPC System* by Product Code, i.e., (16129, 7680, 10) code at BER = 0.00001 But *Adapted MPC + PRPC System* with Product Code i.e. (16,129, 7680, 10) code up to triple-bit error explains improved outcome than *Adapted MPC + PRPC System* with Product Code, i.e., (16129, 7680, 10) code up to double-bit error.

In Fig. 11, it is seen that a reduced amount of energy E_{BER} is required for *Adapted MPC + PRPC System* with Product Code, i.e., (16129, 7680, 10) code up to triple-bit error & *Adapted MPC + PRPC System* with Product Code, i.e., (16129, 7680, 10) code up to double-bit error (for $k = 51200$ bits) with esteem to link error probability (p_{link}) when BER (0.00001) correspondingly than Typical Go-Back-N ARQ.

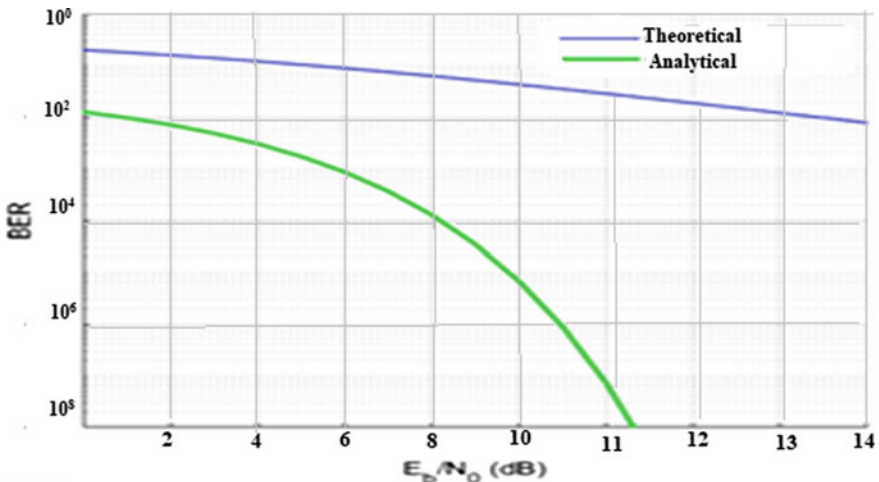


Fig. 8 Performance of MIMO Rayleigh fading channel 10^0

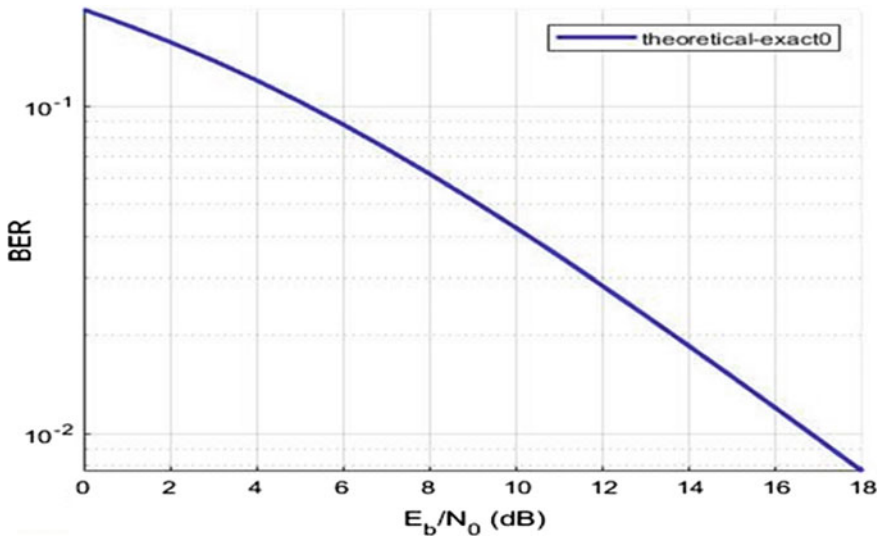


Fig. 9 Graph of BER versus SNR

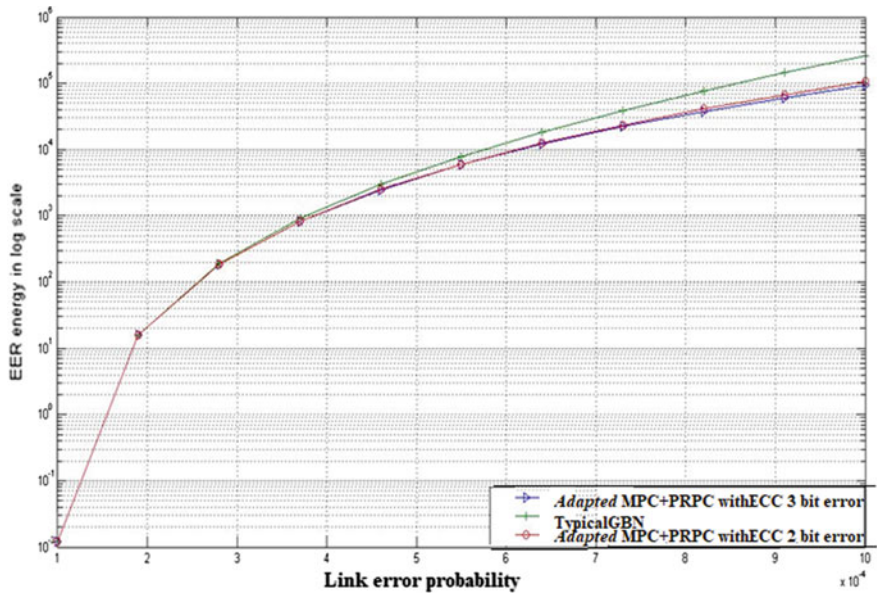


Fig. 10 Energy (E_{EER}) with esteem to link error probability (p_{link}) envelope size $k = 51,200$ Bits with BER = 0.00001 using Product Code, i.e., (16129, 7680, 10)

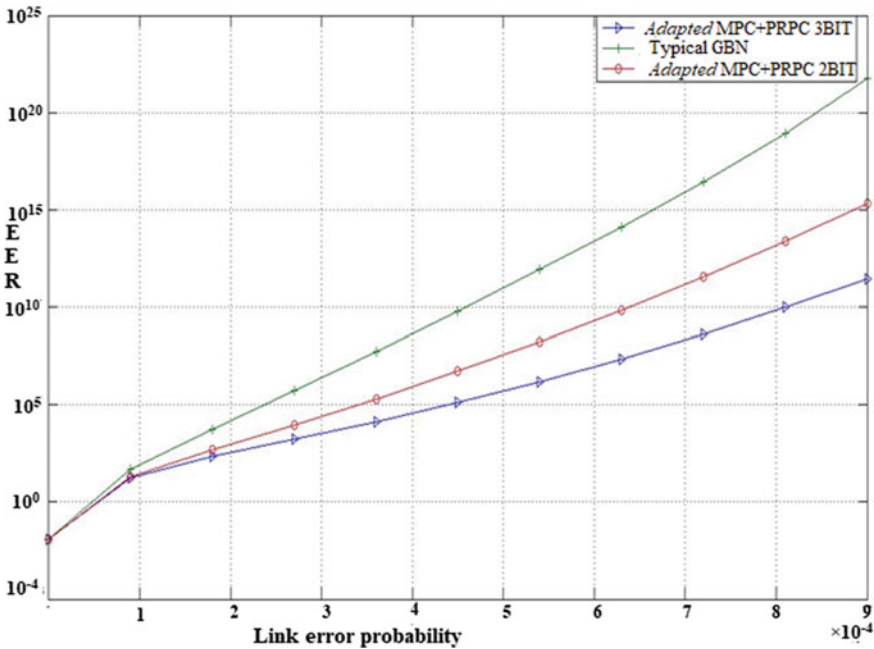


Fig. 11 Energy (E_{EER}) with esteem to link error probability (p_{link}) envelope size $k = 512,000$ Bits with BER = 0.00001 using Product Code, i.e., (16129, 7680, 10)

4.1 BER versus SNR of MIMO-OFDM Scheme

This is the representation of SNR versus BER, as displayed below. The error probability will be calculated in opposition to the SNR. BER versus SNR plot shows the log scale graph, and to overcome this problem, MIMO system simulation is used and obtained with MIMO-Rayleigh fading and plotted between BER and SNR. The graph shows the better qualities of the signal. MIMO antenna used on both sides and hence reduced the bit error rate (Fig. 12).

$$SNR = 0dB, 8dB \ \& \ BER = 0.01, 0.001.$$

With the cellular radio channels, the distribution of Rayleigh fading is generally utilizing for narrating the measurable time fluctuating character to the signal being received in the form of wrapping of flat fading signal or the independent number of path components will be an envelope. This is the familiar that wrapping of the addition of more than one that is two quadrature Gaussian noise signals followed the Rayleigh distribution. It is to calculate the mean squared fading rate. Rayleigh distribution is two types.

- i. NON-LOS
- ii. LOS.

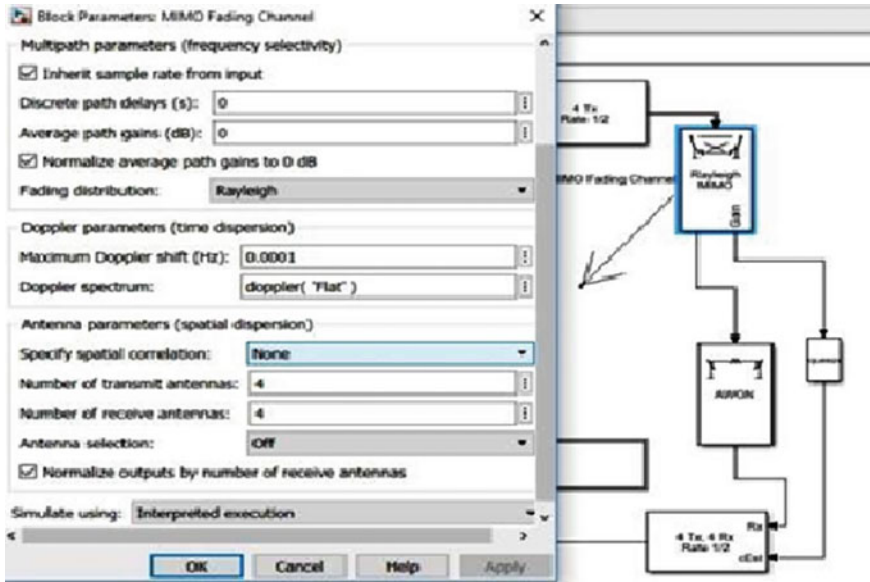


Fig. 12 Rayleigh MIMO fading channel

4.2 Doppler Spectrum

DOPPLER (‘Flat’) in a multipath channel, the channel response is varying in frequency. However, in an OFDM system, we have got a flat response for each sub-carrier bandwidth. Therefore, we have to get a collection of flat fading MIMO transmissions. It helps for channel estimation.

4.3 AC and STBC

STBC is a space–time block code. It does not need CSI (THE INFORMATION OF CHANNEL STATE) or consciousness of the environment coefficient among communication systems. Transmit Antenna = 4, Receiver Antenna = 4.

The picture has been shown the complete 192 numerals to the un-modulated and delineate bits will be transmitted also received by the MIMO-OFDM method, and each of the values of the parameter will be taken within the simulation. **Spectrum analyzer** as shown in Fig. 13 helps to measure direct reading (assumes spectrum analyzer measurement band with is greater than the signal bandwidth) of total bandwidth = $[31.25 * 1000000] * 52 = 1.625$ GHZ.



Fig. 13 No. of bit spectrum analyzer and signal constellation

4.4 Transmit Signal Parameter

Average power = 23.98dBm. Max. Power = 30 dBm. Sample rate = 192 GHz. The symbol of OFDM bearing ZP has power spectral density to the smallest ripple within the band, and the bigger outside to band power accommodate high power has been using for transmitting but the permanent of peak transmission power.

5 Conclusion and Future Work

From the simulation results we observed, there is the OFDM method that will merely subject to a flat fading channel because of longer the CP is larger adequate. This analytical result has got in the channel of Rayleigh fading. However, the accomplishment of bit error rate will produce to ISI is very poor in such number of the path of Rayleigh fading channel concerning guard interval. The orthogonality among sub-carrier's frequency components can be completely preserved. It is also observed from BER calculation that with minimum bit error rate showed that least intersymbol interference (ISI) and gives an excellent performance for wireless communication. But this bit error rate calculation tool calculates the error rate every moment that is dependent upon time. So, the error rate is not fixed like error rate = 0.118, number of error = 129, number of Bits = 1088. The graph representation is also frame-up in the form of BER versus SNR.

We are able to implement this system in Adapted MPC (Modified Packet Combining) + PRPC (Packet Reverse Packet Combining) System with Product Code,

i.e., (16129, 7680, 10) code up to triple-bit error demonstrates extra energy efficient, superior gain percent and with a reduction of transmitting power than Adapted MPC + PRPC System with Product Code, i.e., (16129, 7680, 10) code up to double-bit error; also it affords healthier steadfastness than the prior systems. Our preliminary hypothetical effort and realistic accomplishment with three-bit error correction code furnish heartening consequences.

References

1. Pokle SB, Kulat KD (2006) Matlab simulation of wireless communications using OFDM principle. IETE Tech Rev 23(3):187–197
2. Langton C Orthogonal frequency division multiplexing (OFDM) tutorial. <http://www.complextoreal.com/chapters/ofdm2.pdf>
3. Wangt C-L, Changt C-H, Fan JL, CiofJi JM. Discrete Hartley transform based multicarrier modulation. Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 300, R.O.C; Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305
4. Hanzo L, Munster M, Choi BJ, Keller T (2000) OFDM and MC-CDMA for broadband multi-user communications. IEEE, New York
5. Pokle SB, Kulat KD (2005) OFDM techniques: a novel approach for design of wireless communication system. Proceeding of National level conference on advanced communication techniques, Act-2005 at BIT Durg, April 5–6, 2005, pp 140–150
6. Pokle SB, Kulat KD, Keskar AG (2004) MATLAB simulation of a fuzzy controller for attitude control of a geostationary satellite. IETE J Educ 45(4):203–209
7. Haykin S (1988) Digital communications. Wiley Publications Ltd., Singapore
8. Proakis JG, Manolakis DG Digital signal processing, principles, algorithms, and applications. Prentice-Hall Publications
9. Multiple-access techniques: www.wtec.org/loyola/wireless/02_04.htm
10. Mabiala E, Coinchon M, Maouche K Study of OFDM modulation
11. Wang Y QPSK modulation and demodulation, ELE 791 Software Radio Design
12. Matia D Introduction to OFDM, II edition. OFDM as a possible modulation technique for multimedia, applications in the range of mm waves, 10/30/98/TUD-TVVS
13. MATLAB 6.5 Software. Modeling tool for communication, M/s The Mathworks Inc., USA
14. Hammuda H (1994) Spectral efficiency of multiple access techniques. Department of Electronic and Electrical Engineering, DeMontfort University, Leicester
15. Wikipedia the Free encyclopedia: <http://www.wikipedia.org/>

Ergodic Capacity Analysis of RIS-aided System Relying on User Grouping and Fixed Power Allocation



Kaveti Umamaheswari, Fazal-E-Asim, and Dinh-Thuan Do

Abstract In this paper, we focus on the ergodic capacity of RIS-assisted NOMA in wireless communication systems. The closed upper bound of ergodic capacity analysis is obtained for the proposed technique by using both Rayleigh and Rician fading channel environments for multiple users. Compared to simulations, numerical results demonstrate the tightness and effectiveness of our closed-form expressions.

Keywords Ergodic capacity · Reconfigurable intelligent surface (RIS) · Non-orthogonal multiple access (NOMA) · Rayleigh fading channel · Rician fading channel

1 Introduction

Reconfigurable intelligent surfaces (RIS) have recently gained extensive attention due to their potential to achieve significant improvements in established techniques, throughput, and energy efficiency [1–3]. They are envisioned as an innovative technology for the beyond fifth-generation (B5G) data transmission. According to Ref. [1], RIS is made up of a number of low-cost reflecting elements that are used for reflected signal propagation and can be smartly rearranged by varying the phase shifts of all representing factors to achieve significant improvements in specific skills. RIS is a single-layer meta-surface made up of many passive elements controlled by a smart controller to allow changes occurring in reflections for a range of applications, including signal force enhancement and disturbance suppression. When compared to older techniques such as active relaying and beam shaping, by utilizing just lightweight passive elements, RIS not only reflects signals in a full-duplex and without generating self-interference, but it further greatly reduces energy

K. Umamaheswari (✉) · D.-T. Do
Department of Computer Science and Information Engineering, College of Information and Electrical Engineering, Asia University, Taichung 41354, Taiwan
e-mail: kavetiuma466@gmail.com

Fazal-E-Asim
Department of Tele Informatics Engineering, Federal University of Cear´a, Fortaleza, Brazil

consumption, and finally, hardware/deployment expenditures. In this field, a substantial amount of work has recently been published. For example, the authors of Ref. [4] investigated the role of RIS in improving the physical layer security of wireless systems, whereas the authors of Ref. [5] demonstrated the superiority of an RIS-supported system over one that only consists of an amplify-and-forward relay, and they investigated the joint active and passive beam forming problem in an RIS-supported system. Furthermore, the authors of [1] demonstrated an RIS's resistance to phase-shift errors. A few authors have concentrated their investigations on demonstrating the effectiveness of RISs and/or trying to compare their results to well-known technologies such as relays or multi-antenna systems [5–15].

Non-orthogonal multiple access, or NOMA, is a promising technology for future wireless communication systems [6, 16, 17]. NOMA has also received a lot of attention, and in conventional wireless networks including RIS, it has demonstrated to be faster to orthogonal multiple access (OMA) [18–21]. By increasing spectrum efficiency, balancing user fairness, and expanding network connections. Before interpreting the message in the downlink NOMA, the stronger channel user employs the successive interference cancellation (SIC) technique to cancel out the co-channel impedance from the weaker channel users [19]. Inspired by this fact, we investigate the improved performance of an innovative scheme that combines the features of RIS-NOMA in this paper.

The following are brief summaries of our major contributors inside this paper:

We acquire tighter upper bounds on the possibility of achieving ergodic rates to use both Rayleigh and Rician channel scenarios and show that they closely resemble simulation models.

Notations: Scalars are denoted by lower-case italic letters a , vectors by bold lower-case letters \mathbf{a} matrices by bold upper-case letters \mathbf{A} . The absolute and expected values of x are represented by $|x|$ and $E[x]$, respectively. The transpose of an is a^T , and the i th element of an is $[a]_i$. $\text{diag}(a)$ is a diagonal matrix whose diagonals are the elements of a , whereas the $\arg(x)$ is the complex number x 's phase.

2 System Model

We assume the RIS-assisted NOMA shown in Fig. 1, where the base station (S) wants to connect with two different destinations. The faraway user (U_2) and the near user (U_1) are two types of users. We assume that all nodes have a single antenna, and the RIS consists of N metasurfaces. The signal is transmitted from the source to the two types of users via the RIS, which is designed to improve signal strength at the locations. In an ideal situation, the RIS would be able to obtain the channel state information (CSI) of such users. Fortunately, the RIS may use of such CSI to increase the received SNR at these locations.

RIS utilizes its ability to receive mode and also data transmission; it uses the reflective surface mode. However, the RIS is indeed a mechanism that passively reflects signal. As it is a novel form of RIS implementation used on NOMA network,

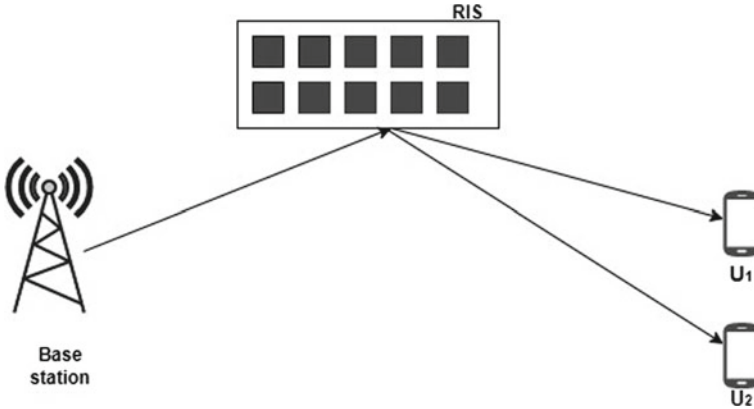


Fig. 1 RIS-NOMA system model

the system does not require a direct link from the source to each user. Each link also considers attenuation caused by distance, which is a first objective in this study.

From source (S) to receivers U_1 and U_2 there is no direct link; we define $h_{SR} \in C^N$, $h_{RU_1} \in C^N$ and $h_{RU_2} \in C^N$ as the channels between $S \rightarrow R$, $R \rightarrow U_1$ and $R \rightarrow U_2$ respectively, so each connection begins to experience both small- and large-scale fading; we assume variant fading paths to and from the RIS, as described in detail and evaluated in the following sections. First, the source (S) sends its signal to RIS (R), which reflects the occurrence signal back to the users. We assume that the power of given varies by the RIS multiple times is very low and could be ignored [5].

By using NOMA method, the overlaid signal $(a_1x_1 + a_2x_2)$ sent by the BS is expected to serve distant mobile users with the presence of RIS. The NOMA concept is used in this study is to provide user equality, with a_1 and a_2 being power allocation factors for users U_1 and U_2 , respectively. We have $a_1 < a_2$ and $a_1^2 + a_2^2 = 1$ because there is less strength needed to sustain for near user. The received signal at the user U_1 is given as,

$$y_1 = \sqrt{P[(h_{SR}^T \Theta h_{RU_1})(a_1x_1 + a_2x_2)] + n_1} \tag{1}$$

Similarly, the received signal at the user U_2 is,

$$y_2 = \sqrt{P[(h_{SR}^T \Theta h_{RU_2})(a_1x_1 + a_2x_2)] + n_2} \tag{2}$$

where P is transmitted power, n_1, n_2 is the additive white Gaussian noise (AWGN) at U_1 and U_2 with zero mean and unit variance of σ_n^2 where $n \in [1, 2]$. Θ is a matrix to control the reflection coefficients as well as phase shifts of the RIS aspects for the user U_1 , $\text{diag}(\eta_1 e^{j\theta_1}, \eta_2 e^{j\theta_2}, \dots, \eta_N e^{j\theta_N})$ where $\eta_i \in [0, 1]$ and $\theta_i \in [0, 2\pi]$ are to be the reflection coefficient and phase shift at the RIS for the i th reflecting element.

Regarding NOMA concept, the x_2 signal is decoded using successive interference cancellation (SIC) of the received superposition signal at x_1 . At U_1 , we consider U_2 's signal as noise, and the maximized signal-to-noise ratio (SNR) to detect signal x_2 is given by [7].

$$\rho_{U_1(x_2)} = \frac{P \left| \sum_{i=1}^N \eta_i e^{j\theta_i} [h_{SR}]_i [h_{RU_1}]_i \right|^2 a_2^2}{\left| \sum_{i=1}^N [h_{SR}]_i [h_{RU_1}]_i \right|^2 a_1^2 + \sigma_1^2} \quad (3)$$

After removing the noise signal from user U_2 by applying SIC, then we compute SNR to detect the signal for U_1 to decode its own signal x_1 . It can be written as,

$$\rho_{U_1(x_1)} = P \sum_{i=1}^N \eta_i e^{j\theta_i} [h_{SR}]_i [h_{RU_1}]_i a_2^2 \quad (4)$$

For far user U_2 , the $\text{diag}(\varsigma_1 e^{j\varphi_1}, \varsigma_2 e^{j\varphi_2}, \dots, \varsigma_N e^{j\varphi_N})$ where ς_i and φ_i is to be the reflection coefficient and phase shift at the RIS; therefore, SNR of U_2 is given by,

$$\rho_{U_2} = \frac{P \left| \sum_{i=1}^N \varsigma_i e^{j\varphi_i} [h_{SR}]_i [h_{RU_2}]_i \right|^2 a_2^2}{\left| \sum_{i=1}^N [h_{SR}]_i [h_{RU_2}]_i \right|^2 a_1^2 + \sigma_2^2} \quad (5)$$

The achievable rate for the proposed schemes can be given as

$$R_1 = \log_2[1 + (\rho_{U_1})] \quad (6)$$

$$R_2 = \log_2[1 + (\rho_{U_2})] \quad (7)$$

Assume the amplitude of reflection coefficient would be established to unity. For reality, although these reflectivities were frequently or less 1, they're often affected by the phase-shift parameters, as shown in [8]. As a result, the two users' phase $\theta_i = \arg([h_{SR}]_i [h_{RU_1}]_i)$ and $\varphi_i = \arg([h_{SR}]_i [h_{RU_2}]_i)$ where $\eta_i = \eta, \forall i$ and $\varsigma_i = \varsigma$ hence, Eqs. (3) and (5) can be rewritten as,

$$\hat{\rho}_{U_1} = \frac{P \left| \eta \sum_{i=1}^N [h_{SR}]_i [h_{RU_1}]_i \right|^2 a_2^2}{\left| \sum_{i=1}^N [h_{SR}]_i [h_{RU_1}]_i \right|^2 a_1^2 + \sigma_1^2} \quad (8)$$

$$\hat{\rho}_{U_2} = \frac{P \left| \varsigma \sum_{i=1}^N [h_{SR}]_i [h_{RU_2}]_i \right|^2 a_2^2}{\left| \sum_{i=1}^N [h_{SR}]_i [h_{RU_2}]_i \right|^2 a_1^2 + \sigma_2^2} \quad (9)$$

3 Ergodic Capacity Analysis

The ergodic capacity is done by assuming two different types of fading channels in this section, that is Rayleigh and Rician fading channels. The upper bounds for ergodic analysis in our suggested method are obtained in the upcoming sections by utilizing Rayleigh and Rician fading channels.

$$h_j = \begin{cases} \sqrt{\frac{K_j}{K_j+1}} \bar{h}_j, & \text{LOS(Rician)} \\ \bar{h}_j, & \text{NLOS(Rayleigh)} \end{cases} \quad (10)$$

where K_j is the Rician-K factor, $j \in (\text{SR}, \text{RU}_1 \text{ and } \text{RU}_2)$.

3.1 Upper-Bound Ergodic Capacity Using Rayleigh Fading Channel

Considering the non-line-of-sight (NLOS) to the users U_1 and U_2 , where $\tilde{h}_l = \tilde{g}_l d_l^{-\alpha/2}$, $l \in \text{SR}, \text{RU}_1 \text{ and } \text{RU}_2$. Where \tilde{g}_l is the complex-Gaussian small-scale fading channels having zero mean and unit variance, d is the distance between two nodes and α is the path-loss exponent,

where $\rho_m = \frac{P}{\sigma_n^2}$, $n \in [1, 2]$ and $m \in [1, 2]$ by defining the $\sqrt{\tilde{\psi}_{\text{SRU}_1}} = \frac{\eta}{N} \sum_{i=1}^N |h_{\text{SR}}|_i |h_{\text{RU}_1}|_i$ and $\sqrt{\tilde{\psi}_{\text{SRU}_2}} = \frac{\zeta}{N} \sum_{i=1}^N |h_{\text{SR}}|_i |h_{\text{RU}_2}|_i$.

For many reflecting elements N , both $\sqrt{\tilde{\psi}_{\text{SRU}_1}}$ and $\sqrt{\tilde{\psi}_{\text{SRU}_2}}$ are considering making use of a non-central chi-square distribution with expected values for two users written in below (11),

$$E[\tilde{\psi}_{\text{SRU}_1}] = \frac{\eta^2 \left[\pi^2 + \left(\frac{1}{N} \right) (16 - \pi^2) \right]}{16 d_{\text{SR}}^\alpha d_{\text{RU}_1}^\alpha} \quad (11)$$

$$E[\tilde{\psi}_{\text{SRU}_2}] = \frac{\zeta^2 \left[\pi^2 + \left(\frac{1}{N} \right) (16 - \pi^2) \right]}{16 d_{\text{SR}}^\alpha d_{\text{RU}_2}^\alpha} \quad (12)$$

The expected value of the observed SNR at users can be calculated using Eqs. (11) and (12). And it must be observed that its second term in exists whenever the holder inequality is applied. Where ρ_{U_1} is received SNR at U_1 , by putting the $E[\rho_{U_1}]$ we can write in below equations,

$$E[\hat{\rho}_{U_1}] = E \left[\rho_1 \left(N \sqrt{\tilde{\psi}_{\text{SRU}_1}} \right)^2 \right] \quad (13)$$

$$E[\hat{\rho}_{U_1}] = \rho_1 \left(N^2 [\tilde{\psi}_{SRU_1}] \right) \tag{14}$$

$$E[\hat{\rho}_{U_1}] \leq \rho_1 \left(\frac{1}{16} d_{SR}^{-\alpha} d_{RU_1}^{-\alpha} \eta^2 N (16 + (N - 1)\pi^2) \right) \tag{15}$$

Furthermore, an expected value of U_2 is provided in Eq. (16) here after performing the same methods in U_1 ; after this, we can compute the expected value for user U_2 .

$$E[\hat{\rho}_{U_2}] \leq \rho_2 \left(\frac{1}{16} d_{SR}^{-\alpha} d_{RU_2}^{-\alpha} \zeta^2 N (16 + (N - 1)\pi^2) \right) \tag{16}$$

The ergodic upper-bound equations for U_1 and U_2 are,

$$R_{upper}^1 = \log_2(1 + \hat{\rho}_{U_1}) \tag{17}$$

$$R_{upper}^2 = \log_2(1 + \hat{\rho}_{U_2}) \tag{18}$$

3.2 Upper-Bound Ergodic Capacity Using Rician Fading Channel

In this section, we assume that line-of-sight (LOS) channel exists between the links in SRU_1 and SRU_2 , and that they are characterized as Rician fading channels in this scenario. $\tilde{h}_k = \tilde{g}_k d_k^{-\alpha}/2$ the links are represented by k . Where $\tilde{g}_k = (e^{j\phi_{k,1}}, e^{j\phi_{k,2}}, \dots, e^{j\phi_{k,N}})^T$ is a fixed-component vector with unit-power components, and $\phi_{k,n} \in [0, 2\pi]$ is the n th channel's phase, the angle of arrival/departure is controlled by the wavelength, distance between reflecting elements and wavelength.

Now, we assume the $\sqrt{\bar{\psi}_{SRU_1}} = \frac{\eta}{N} \lambda_{SRU_1} \sum_{i=1}^N |[h_{SR}]_i [h_{RU_1}]_i|$ and $\sqrt{\bar{\psi}_{SRU_2}} = \frac{\zeta}{N} \lambda_{SRU_2} \sum_{i=1}^N |[h_{SR}]_i [h_{RU_2}]_i|$. Where the values of $\lambda_{SRU_1} = \sqrt{\frac{K_{SRU_1}}{(K_{SR}+1)(K_{RU_1}+1)}}$ and $\lambda_{SRU_2} = \sqrt{\frac{K_{SRU_2}}{(K_{SR}+1)(K_{RU_2}+1)}}$ and K is the Rician factor. In addition, the $\bar{\psi}_{SRU_1}$ and $\bar{\psi}_{SRU_2}$ have the constant mean values that is $E[\bar{\psi}_{SRU_1}] = \frac{\eta^2 \lambda_{SRU_1}^2}{d_{SR}^\alpha d_{RU_1}^\alpha}$ and $E[\bar{\psi}_{SRU_2}] = \frac{\zeta^2 \lambda_{SRU_2}^2}{d_{SR}^\alpha d_{RU_2}^\alpha}$. Therefore, the expected values SNR of U_1 by using Rician channel can be written as,

$$E[\bar{\rho}_{U_1}] = E \left[\rho_1 \left(N \sqrt{\bar{\psi}_{SRU_1}} \right)^2 \right] \tag{19}$$

$$E[\bar{\rho}_{U_1}] = \rho_1 (N^2 [\bar{\psi}_{\text{SRU}_1}]) \quad (20)$$

$$E[\bar{\rho}_{U_1}] \leq \rho_1 \left(\eta^2 \frac{N^2 \beta_{\text{SRU}_1}^2}{d_{\text{SR}}^\alpha d_{\text{RU}_1}^\alpha} \right) \quad (21)$$

Similarly, the expected value for ρ_{U_2} is,

$$E[\bar{\rho}_{U_2}] \leq \rho_2 \left(\zeta^2 \frac{N^2 \beta_{\text{SRU}_2}^2}{d_{\text{SR}}^\alpha d_{\text{RU}_2}^\alpha} \right) \quad (22)$$

Therefore, the upper bound for the proposed scheme using Rician fading channel for U_1 is,

$$\tilde{R}_{\text{upper}}^1 = \log_2(1 + \bar{\rho}_{U_1}) \quad (23)$$

Similarly, the upper bound for the proposed scheme using Rician fading channel for U_2 is written as,

$$\tilde{R}_{\text{upper}}^2 = \log_2(1 + \bar{\rho}_{U_2}) \quad (24)$$

4 Numerical Analysis

In this section, we use mathematical derivations to simulate ergodic capacity. The transmitter is placed at the origin, at the beginning $(x_S, y_S) = (0, 0)$, $(x_R, y_R) = (80, 0)$, through RIS at near user $(x_{U_1}, y_{U_1}) = (90, 0)$ and far user $(x_{U_2}, y_{U_2}) = (100, 0)$. The reflection coefficient for RIS ($\eta = \zeta$) is 0.8. For LoS link, α was set to 2 and 2.5 for NLoS channels, noise variance of σ_n^2 is unity, and the Rician-K factor is 10 dB. The final simulations evaluated combined multiple channel realizations, and the SNR is defined as, $\rho_m = \frac{P}{\sigma_n^2}$ where m represents the links for U_1 and U_2 .

Figure 2 depicts the ergodic capacity performance of two users over Rayleigh fading channels. The reason behind it is that RIS receives a transmitted signal as from BS in the format of electromagnetic (EM) radiation and directs it to the users according to their positions, enhancing the SNR at the users. The near user U_1 has better performance than the far user U_2 .

The performance of ergodic capacity for multiple users over Rician fading channels is depicted in Fig. 3. The U_1 significantly outperformed the U_2 in terms of performance.

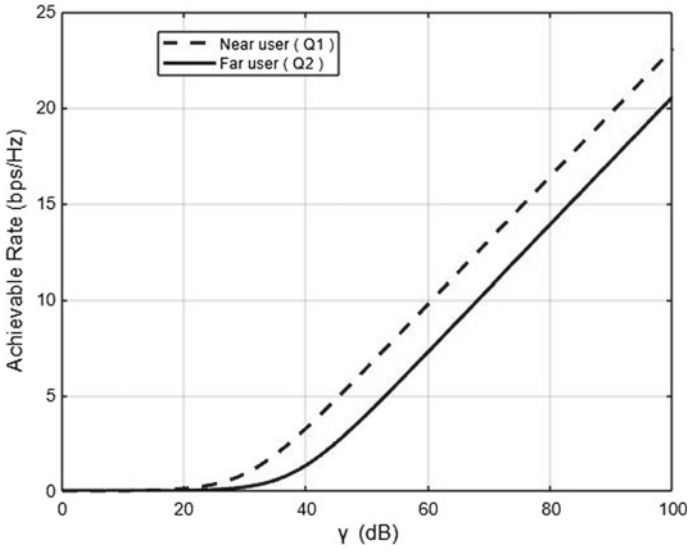


Fig. 2 Rayleigh fading channel for U_1 and U_2

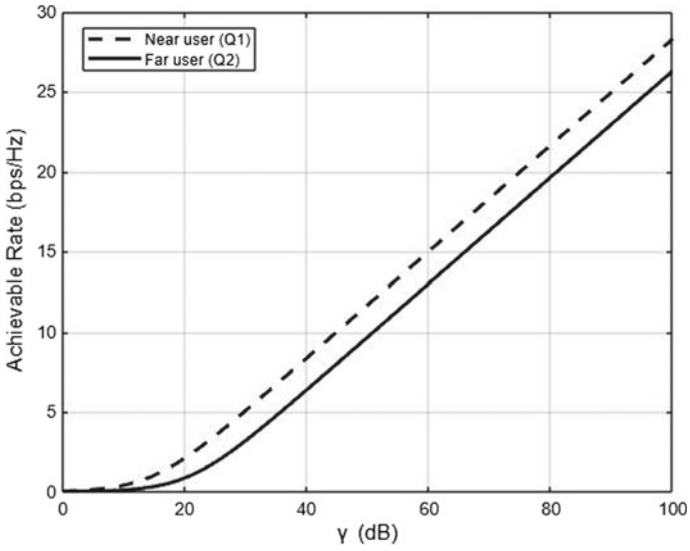


Fig. 3 Rician fading channel for U_1 and U_2

5 Conclusion

With the use of RIS-NOMA and numerous channels, this paper explores the ergodic capacity analysis for multiple users. In the ideal scenario of employing reflecting elements, upper bounds based on possible rates were successfully determined in both Rayleigh and Rician fading systems. The simulation results indicate that the Rayleigh channel outperforms the Rician channel. In this scenario, the near user U_1 has better performance than the far user U_2 .

References

1. Wu Q, Zhang R (2019) Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming. *IEEE Trans Wirel Commun* 18(11):5394–5409
2. Di Renzo M et al (2019) Smart radio environments empowered by reconfigurable AI metasurfaces: an idea whose time has come. *EURASIP J Wirel Commun Netw* 2019(1). Article no. 129
3. Wu Q, Zhang R (2020) Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network. *IEEE Commun Mag* 58(1):106–112
4. Liu Y, Qin Z, Elkashlan M, Ding Z, Nallanathan A, Hanzo L (2017) Nonorthogonal multiple access for 5G and beyond. *Proc IEEE* 105(12):2347–2381
5. Cui M, Zhang G, Zhang R (2019) Secure wireless communication via intelligent reflecting surface. *IEEE Wirel Commun Lett* 8(5):1410–1414
6. Liaskos C, Nie S, Tsioliaridou A, Pitsillides A, Ioannidis S, Akyildiz I (2018) A new wireless communication paradigm through software-controlled metasurfaces. *IEEE Commun Mag* 56(9):162–169
7. Nguyen T-L, Do D-T (2018) Power allocation schemes for wireless powered NOMA systems with imperfect CSI: an application in multiple antenna-based relay. *Int J Commun Syst* 31(15):e3789
8. Lee D, Lee JH (2011) Outage probability of decode-and-forward opportunistic relaying in a multicell environment. *IEEE Trans Veh Technol* 60(4):1925–1930
9. Basar E, Di Renzo M, De Rosny J, Debbah M, Alouini M, Zhang R (2019) Wireless communications through reconfigurable intelligent surfaces. *IEEE Access* 7:116753–116773
10. Gradshteyn I, Ryzhik I (2007) Table of integrals. Series and Products Academic Press
11. Badiu M-A, Coon JP (2020) Communication through a large reflecting surface with phase errors. *IEEE Wireless Commun Lett* 9(2):184–188
12. Tse D, Viswanath P (2005) Fundamentals of wireless communication. Cambridge University Press
13. Abeywickrama S, Zhang R, Wu Q, Yuen C (2020) Intelligent reflecting surface: practical phase shift model and beamforming optimization [Online]. Available: <https://arxiv.org/abs/2002.10112>
14. Abdullah Z, Chen G, Lambotaran S, Chambers JA (2020) A hybrid relay and intelligent reflecting surface network and its ergodic performance analysis. *IEEE Wirel Commun Lett* 9(10):1653–1657
15. Huang C, Zappone A, Alexandropoulos GC, Debbah M, Yuen C (2019) Reconfigurable intelligent surfaces for energy efficiency in wireless communication. *IEEE Trans Wirel Commun* 18(8):4157–4170
16. Zheng B, Wu Q, Zhang R (2020) Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA. *IEEE Commun Lett* 24(4)

17. Ding Z, Yang Z, Fan P, Poor HV (2014) On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process Lett* 21(12):1501–1505
18. Ding Z, Vincent Poor H (2020) A simple design of IRS-NOMA transmission. *IEEE Commun Lett* 24(5):1119–1123
19. Do D-T, Le A, Lee BM (2020) NOMA in cooperative underlay cognitive radio networks under imperfect SIC. *IEEE Access* 8:86180–86195
20. Hou T, Liu Y, Song Z, Sun X, Chen Y, Hanzo L Reconfigurable intelligent surface aided NOMA networks. *IEEE J Sel Areas Commun.* <https://doi.org/10.1109/JSAC.2020.3007039>
21. Zhang J, Bjrnson E, Matthaiou M, Ng DWK, Yang H, Love DJ (2019) Multiple antenna technologies for beyond 5G. *Arxiv*, 1910.00092

Author Index

A

Abhijit Sarkar, 145, 153
Abhijyoti Ghosh, 299
Adireddy Ramesh, 603
Ajeya Jha, 99, 137
Ajeya K. Jha, 145, 153
Ajith Bala, B., 211
Ajitha, D., 1, 545
Akash Kumar Bhoi, 223, 471
Akhtar Ismail Nadaf, 87
Akshatha, M. S., 71
Amrita Biswas, 369
Anh-Tu Le, 569
Anjali Jha, 447
Anuj Kumar, 413
Anup Sharma, 15
Arjun Chakravarthi Pogaku, 569
Arjun Kar, 427
Arvind Choubey, 289
Athulya, P., 173
Avinash Bhute, 111

B

Badarla Sri Pavan, 593
Baibaswata Bhattacharjee, 25, 45
Bappaditya Roy, 347, 535
Bhagya Lakshmi, K., 1
Bhattacharjee, A. K., 625
Bibeth Sharma, 137, 153
Bipasha Chakrabarti, 625
Boyapati Bharathidevi, 309

C

Chitragada Roy, 161

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

S. Dhar et al. (eds.), *Advances in Communication, Devices and Networking*, Lecture Notes in Electrical Engineering 902, <https://doi.org/10.1007/978-981-19-2004-2>

Chitrapriya Ningthoujam, 427

D

Debasree Mitra, 447, 461
Debosmita Ghosh, 495
Deepak Betadur, 279
Deepika Koundal, 471
Devender Sharma, 319
Dewan, R., 347
Dharanidaran, R. B., 535
Dinh-Thuan Do, 569, 637
Divya Chaudhari, 111
Donepudi Tata Rao, 603
Dristi Dugar, 447
Dwaipayana Saha, 483

E

Elamaran, V., 199, 211

F

Fazal-E-Asim, 637

G

Gautham Reddy, 377
Gopal Thapa, 339
Gopalakrishnan, E. A., 437
Gurjinder Kaur, 265
Gurmehr Singh, 265
Gursharan Sandhu, 265

H

Haraprasad Mondal, 15

Hari Theivaprakasham, 437
 Harigovindan, V. P., 593
 Hima Bindu, K., 545

I

Indrani Mukherjee, 483
 Ira Nath, 387
 Ishika Dhall, 233

J

Jana Shafi, 471
 Jaya Rani Rani, 145, 153
 Jaya Sharma, 129
 Jayendra Kumar, 309
 Jesmin Roy, 483
 Jeyakumari, D., 199, 211
 Jitendra Kumar, 145, 153
 Joseph Bamidele Awotunde, 251

K

Kamanashis Goswami, 15
 Kamran Iqbal, 447
 Kapila Sharma, 339
 Karri Manoz Kumar Reddy, 603
 Karthik K. Bhat, 377
 Kartik Aggarwal, 251
 Kashish Mahindroo, 319
 Kaveti Umamaheswari, 637
 Kirti Vyas, 129
 Koushik Karmakar, 583
 Kuldeep Bhalerao, 413

L

Laasya Lata Anumakonda, 545, 611
 Litisha Mohapatra, 529
 Lolit Kumar Singh, L., 299

M

Manam Ravindra, 603
 Manoj Sarkar, 299
 Meenu Garg, 265
 Meghna Chhabra, 413
 Mehaboob Mujawar, 503
 Mohammed Ismail, B., 173
 Mohammed Mahaboob Basha, 611
 Mohammed Nasir Ansari, 519
 Monika Arora, 413
 Muhammad Fazal Ijaz, 265

N

Nagabushanam, M., 71
 Narayana Rao Palepu, 309
 Naveen Kumar, P., 279
 Neeha Cintury, 161
 Nhan Duc Nguyen, 569
 Nikita Agarwal, 401
 Niranjan Girhe, 111
 Niranjana Krupa, 377
 Nirmal Rai, 339
 Nishant Gupta, 529
 Nishant Saxena, 355

P

Pandya Vyomal Naishadhkumar, 611
 Partha Sarathi Chakraborty, 495
 Pinakshi De, 387
 Pooja Reddy Bathula, 611
 Potaraju Yugender, 35
 Prachi Channe, 111
 Prajit Paul, 625
 Pranati Rakshit, 447, 461, 483, 495
 Pratik Mahapatra, 369
 Praveen Kumar, 289
 Pravesh Pal, 289
 PremaLatha, V., 223
 Priyanka Saha, 35
 Priyanka S. Das, 625
 Pronit Sarkar, 495
 Pushpita Chatterjee, 233

R

Rachit Saxena, 355
 Rahul Kumar, 355
 Ranjeet Kumar, 289
 Rashmi Sinha, 289
 Ravi Kumar, 511
 Rayapudi Srinivasa Rao, 603
 Rijhi Dey, 511, 529
 Rinkila Bhutia, 339
 Ritam Dutta, 401
 Ritika Saini, 511
 Ritu Jain, 355
 Roy, B., 625
 Rudra Sankar Dhar, 35, 511, 529
 Rupsa Roy, 57

S

Sahil Sejwal, 251
 Saibal Kumar Saha, 99, 137
 Samrat Kumar Mukherjee, 145, 153

Sangita Saha, 99
 Sanjay A. Pardeshi, 87
 Sankha Subhra Debnath, 327
 Santosh Kumar Mahto, 289
 Sarmistha Neogy, 583
 Sarthak Singhal, 319
 Sasi, G., 211
 Sheifali Gupta, 265, 471
 Shekhar, R., 183
 Shilpi Barman Sharma, 233
 Shivam Raj, 369
 Shreya Sakshi, 427
 Shridhar Sharma, 559
 Shubhankar Roy, 495
 Sivaarunagirinathan, S., 199
 Sivakumar, S., 23
 Snigdha Pv, 611
 Sohail Saif, 583
 Soman, K. P., 437
 Somu, G., 279
 Sonia Munjal, 137
 Soumya Ranjan Nayak, 223, 233, 251, 265,
 471
 Soumyarupa Saha, 161
 Sourav Dhar, 57
 Sowmya, V., 437
 Sreedevi, E., 223
 Sri Krishna Mishra, 355
 Sricharan, B., 545
 Sridhar, T., 183
 Subhadip Talukder, 495

Sudipta Chattopadhyay, 299
 Sujatha, Y., 1
 Sujit Kumar Verma, 355
 Sundar Prakash Balaji, M., 199, 211
 Suparna Biswas, 583
 Surajit Bosu, 25, 45
 Swagat Nanda, 35
 Swarup Sarkar, 57
 Swastika Chakraborty, 327, 339
 Swetha, K., 503

T

Thanapal, P., 199

U

Ujjwal Mondal, 511, 529
 Umang Lunia, 377
 Uttam Ghosh, 223

V

Vani Sadadiwala, 319
 Vatsala Anand, 471
 Venkata Lakshmaiah, Y., 347
 Vijaya Saradhi, D., 503
 Vimlesh Singh, 319
 Vinayakumar Ravi, 437
 Vivek Pandey, 145