M. Seetha
Sateesh K. Peddoju
Vishnu Pendyala
Vedula V. S. S. S. Chakravarthy   *Editors*

Intelligent Computing and Communication

Proceedings of 6th ICICC 2022

Springer

Advances in Intelligent Systems and Computing

Volume 1447

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).

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

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

M. Seetha · Sateesh K. Peddoju · Vishnu Pendyala ·
Vedula V. S. S. S. Chakravarthy
Editors

Intelligent Computing and Communication

Proceedings of 6th ICICC 2022

Editors
M. Seetha
Department of Computer Science
G. Narayanamma Institute of Technology
and Science (Autonomous)
Hyderabad, India

Vishnu Pendyala
Faculty of Data Science
San Francisco Bay Area, CA, USA

Sateesh K. Peddoju
Department of Computer Science
Indian Institute of Technology Roorkee
Roorkee, India

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

Conference Committee

International Conference on Intelligent Computing and Commination (ICICC-2022)

Department of Computer Science and Engineering, G. Narayanamma Institute of Technology and Science for Women, (Autonomous), Shaikpet, Hyderabad, Telangana-500104, India

Chief Patrons

Shri. P. Subba Reddy, Chairman, GNITS, Hyderabad, Telangana
Smt. Srividya Reddy Gunampalli, Vice-chairman, GNITS, Hyderabad, Telangana
Prof. N. V. Ramana Rao, Director, NIT Warangal, Telangana

Patron

Dr. K. Ramesh Reddy, Principal, GNITS, Hyderabad, Telangana

Honorary Chairs

Dr. Roman Senkerik, Tomas Bata University in Zlin, Czech Republic
Dr. Uday Tupakula, Deputy Director, ACSRC, University of Newcastle, Australia

General Chairs

Prof. R. B. V. Subramanyam, Professor, CSE, NIT Warangal, Telangana
Prof. Jinshan Tang, George Mason University, USA
Dr. Suresh Chandra Satapathy, KIIT, Odisha

Conference Chairs

Dr. S. Ravi Chandra, Professor and Head, CSE, NIT Warangal, Telangana
Dr. M. Seetha, Professor and Head, CSE, GNITS, Hyderabad

Editorial Board

Dr. M. Seetha, Professor and Head, CSE, GNITS, Hyderabad, Telangana, India
Dr. Sateesh K. Peddoju, Associate Professor, Department of CSE, IIT Roorkee, India
Dr. Vishnu Pendyala, Professor, San Jose State University College of Engineering, California, USA
Dr. Vedula V. S. S. S. Chakravarthy, Professor, ECE, Raghu Institute of Technology, Vizag, Andhra Pradesh

Publication Chairs

Dr. N. Kalyani, Professor, Department of CSE, GNITS, Hyderabad, Telangana, India
Dr. Vikrant Bhateja, Shri Ramswaroop Memorial Group of Professional Colleges, Lucknow, Uttar Pradesh

Program Chairs

Dr. A. Sharada, Professor, Department of CSE, GNITS, Hyderabad, Telangana, India
Dr. O. Obulesu, Associate Professor, Department of CSD, GNITS, Hyderabad, TS, India

Technical Program Chairs

Mrs. Jayashree S Patil, Associate Professor of CSE, GNITS, Hyderabad, Telangana
Dr. D. V. Lalitha Parameswari, Associate Professor of CSE, GNITS, Hyderabad, Telangana
Dr. K. Prasanna, Associate Professor of CSM, GNITS, Hyderabad, Telangana

Steering Committee

Dr. P. V. D. Soma Sekhar Rao, Professor and Dean Academics, GNITS, Hyderabad, Telangana
Dr. S. S. Tulasi Ram, Professor and Dean-PG Studies, GNITS, Hyderabad, Telangana
Dr. K. Rama Linga Reddy, Professor and Head, ETE, GNITS, Hyderabad, Telangana
Dr. G. P. Prasada Reddy, Professor and Controller of Examinations, GNITS, Hyderabad, Telangana
Dr. I. Ravi Prakash Reddy, Professor and Head, IT, GNITS, Hyderabad, Telangana
Dr. B. Venkateshulu, Professor and Head, ECE, GNITS, Hyderabad, Telangana
Dr. N. Malla Reddy, Professor and Head, EEE, GNITS, Hyderabad, Telangana
Mr. Venkata Ramana Reddy, Head, ME, GNITS, Hyderabad, Telangana
Dr. P. Aparna, Head, HM, GNITS, Hyderabad, Telangana
Dr. T. Charan Singh, Head, BS, GNITS, Hyderabad, Telangana
Dr. G. Narendra Babu Reddy, Training and Placement Officer, GNITS, Hyderabad, Telangana
Mr. T. V. Rammohan Reddy, Head, CE, GNITS, Hyderabad, Telangana
Dr. K. Bharatha Lakshmi Devi, Librarian, GNITS, Hyderabad, Telangana
Dr. M. V. L. Surya Kumari, Physical Directress, GNITS, Hyderabad, Telangana

International Advisory Committee

Dr. Aweng, Professor, Universiti Malaysia Kelantan
Dr. Marmelo V. Abante, Professor, World Citi Colleges, Philippines
Dr. Erry Yulian, Professor, IIUM, Malaysia
Dr. George Barber, Country Manager, Terra Energy and Resource Technologies, Indonesia
Dr. Sheryl, Professor, MinSCAT Bongabong Campus, Philippines
Mr. Menchita F. Dumlao, Professor, Women's University (PWU), Philippines
Dr. Neil P. Balbo, Vice President, Philippine Society of IT Education (PSITE)
Dr. I. Made Putrawan, Professor, State University of Jakarta, Indonesia
Dr. Ricky, Professor, FCPC, Philippines
Dr. Sasitharan Nagappan, Professor, UTHM, Malaysia

Dr. T. Kishore Kumar, Professor, ECE, NIT Warangal, Telangana
Dr. N. Bheema Rao, Professor, ECE, NIT Warangal, Telangana
Dr. P. Rajesh kumar, Professor and HoD, ECE, AU College of Engineering, Vizag, Andhra Pradesh
Dr. D. Srinivasa Rao, Professor of ECE, JNTUH, Hyderabad, Telangana
Dr. M. Asha Rani, Professor, ECE, JNTUH, Hyderabad, Telangana
Dr. Subhash Kulkarni, Professor and HoD, PES University, Bengaluru, Karnataka
Dr. Sumantra Dutta Roy, Professor, EEE, IIT Delhi
Dr. A. G. Keskar, Prof. Visvesvaraya National Institute of Technology, Nagpur, Maharashtra
Dr. G. Bhuvaneshwari, Professor (HAG), IIT, New Delhi
Dr. K. Shanti Swarup, Professor, IIT, Madras
Dr. M. Surya Kalavathi, Professor of EEE, Yogi Vemana University, YSR, Andhra Pradesh
Dr. N. K. Kishore, Professor, EEE, IIT Kharagpur
Dr. D. M. Vinod Kumar, Professor (HAG), EEE, NIT Warangal, Telangana
Dr. N. Yadaiah, Professor, EEE, JNTUH, Hyderabad, Telangana
Dr. P. V. N. Prasad, Professor, EEE, Osmania University, Hyderabad, Telangana
Dr. G. Yesuratnam, Professor, EEE, Osmania University, Hyderabad, Telangana
Dr. Siva Sarma D. V. S. S., Professor (HAG),EEE, NIT Warangal, Telangana
Dr. M. Vijaya Kumar, Professor of EEE, JNTUA, Anantapur, Andhra Pradesh
Dr. S. Ashok, Professor, EEE, NIT Calicut, Kerala
Pradeep Nirgude, Additional Director, CPRI, Hyderabad, Telangana
Dr. Udaykumar R. Yaragatti, Professor, EEE, NIT Surathkal, Karnataka
Dr. Amit Jain, Head of Division, CPRI, Bengaluru, Karnataka
Dr. Mini Shaji T., Professor, EEE, Jamia Millia Islamia (Central University), New Delhi
Dr. G. Tulasi Ram Das, Professor, EEE, JNTUH, Hyderabad, Telangana
Dr. Malaji Sushama, Professor, EEE, JNTUH, Hyderabad, Telangana
Dr. K. Siva Kumar, Professor, EEE, IIT Hyderabad, Telangana
Dr. A. Raghu Ram, Professor, EEE, JNTUH, Hyderabad, Telangana
Dr. B. Padmaja Rani, Professor, CSE, JNTUH, Hyderabad, TS, India
Dr. D. Vasumathi, Professor, CSE, JNTUH, Hyderabad, TS, India
Dr. R. Sridevi, Professor, CSE, JNTUH, Hyderabad, TS, India
Dr. I. V. Murali Krishna, Dr. Raja Ramanna, Distinguished Fellow Former Professor and Head Spatial Information Technology and Director [R&D], JN Technological University, Hyderabad, India

Conference Committee

Dr. G. Annapurna, Professor of EEE, GNITS
Dr. Raj Kumar L. B., Professor of ETE, GNITS
Dr. Swapna Raghunath, Professor of ECE, GNITS

Dr. Supriya Vaddi, Associate Professor of IT, GNITS
Dr. S. Ravi Kumar Raju, Associate Professor of CSM, GNITS
Dr. B. Sasidhar, Assistant Professor of CSM, GNITS
Dr. S. Vasundhara, Assistant Professor of HM, GNITS
Dr. S. Uday Bhasker, Associate Professor of BS, GNITS

Local Organizing Committee

Dr. G. Malini Devi, Associate Professor of CSE, GNITS, Hyderabad, Telangana
Dr. Raghavender K. V., Associate Professor of CSE, GNITS, Hyderabad, Telangana

Preface

The maiden International Conference on Intelligent Computing and Communication (ICICC-2022), organized by the Department of Computer Science and Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India, is held during 18th and 19th November 2022. G. Narayanamma Institute of Technology and Science (Autonomous), Hyderabad, is commemorating 25 years of Silver Jubilee celebrations in 2022. G. Narayanamma Institute of Technology and Science (GNITS) is a leading engineering college for women and was founded by the late Sri. G. Pulla Reddy in 1997, with the objective to provide quality education in Engineering. GNITS received UGC Autonomous status for 10 years from 2018, and it is affiliated to Jawaharlal Nehru Technological University Hyderabad (JNTUH), Hyderabad, with accreditation from NAAC, NBA, and ISO certified. College is listed in the rank band of 251–300 in NIRF-2022 approved by MHRD. GNITS offers eight B.Tech. (CSE, CSM, CSD, IT, CST, ECE, EEE, and ETM) and five M.Tech. (CSE, DECE, PEED, CNIS, and WMC) programs with 3200 students enrolled in the B.Tech. stream and 180 in the M.Tech stream. This is the first college in India established an Apple Lab, inaugurated by Tim Cook. 28 MoUs with Industry are in place to support good internships and industry projects for students. The college has three research centers affiliated with JNTUH in CSE, ECE, and EEE departments.

The aim of AICTE sponsored 6th International Conference on Intelligent Computing and Communication (ICICC-2022) is to provide a platform for researchers, scientists, technocrats, academicians, and engineers to exchange their innovative ideas and new research findings in the field of computational intelligence and communication. The conference will boost excellent innovations in terms of day-to-day life and academics. The proceedings of the conference will unleash a package platform for innovative ideas into a single unit. Prospective authors are invited to submit manuscripts reporting original unpublished research and recent developments in the topics related to the field of computational intelligence and communication. Experts from different parts of the globe are involved in the interaction on respective fields of the conference theme. Out of 250 submissions from all over the world, only 65 papers were selected after thorough reviewing, for publication in the Springer Book Series on Advances in Intelligent Systems and Computing (AISC).

Selected papers were divided into six tracks, well balanced in the content, and created enough discussion space for trending concepts. The purpose of the conference has been served satisfactorily through international and national speakers, 65 oral presentations by delegates, exchanging and sharing research knowledge among peers. This conference created an ample opportunity for discussions, debate, and exchange of ideas and information among participants. We are very much grateful for the international and national advisory committee, session chairs, peer reviewers who provided critical reviews in selecting quality papers, organizing committee members, student volunteers, and faculty of the Department of Computer Science and Engineering, who contributed to the success of this conference. We are also thankful for all the authors who submitted quality papers and communicated with their peers through the presentation of the work, which led to the grand success of the conference.

We are very much thankful to AICTE, New Delhi, for sanctioning the grant under GOC scheme, and to the Director, NIT Warangal, for their technical support in conducting the conference with a great success. We are very much thankful to the management of G. Narayanamma Institute of Technology and Science for their support in every step of the journey toward the success of this conference, which inspired organizers and motivated many others.

Hyderabad, India   M. Seetha
Roorkee, India   Sateesh K. Peddoju
San Francisco Bay Area, CA, USA   Vishnu Pendyala
Visakhapatnam, India   Vedula V. S. S. S. Chakravarthy

Contents

About the Editors

Dr. M. Seetha, was awarded Ph.D. from JNTU Hyderabad. She has 30 years of teaching experience in reputed institutions. The department received a total grant worth Rs. 9,611,419/- under various schemes like TIDE, DDP, AICTE—CAYT, MODROBS, and PMKVY schemes in her headship for conducting research and organizing conferences. Two DST projects worth Rs. 5,814,719/- were sanctioned under Device Development Programme (DDP) and Technology Interventions for Disabled and Elderly (TIDE) schemes. A grant of Rs. 9.8 Lakhs was received under PMKVY Scheme, a grant of Rs. 7.1 Lakhs was received under MODROBS-AICTE (AQIS) to support research in artificial intelligence, a SEED amount of Rs. 425,000/- was sanctioned from GNITS Management for two research projects, a grant of Rs. 366,700/- was sanctioned from AICTE for conducing international conference in November 2022 as part of Silver Jubilee celebrations of GNITS, MODROBS of Rs. 2.65 lakhs was received to establish "IoT and Analytics Centre of Excellence" under AICTE-(AQIS) scheme. AICTE—CAYT grant of Rs. 10.5 Lakhs were received in 2009. 81 research projects were applied to different funding agencies like DST, AICTE, DRDO, NRSC, TSCOST, and BTBP till date. She is the recipient of Best HOD of the Year Award by CSI Mumbai TechNext, Acharya Ratna Award by Information Technology Association of Andhra Pradesh in 2019, Distinguished Faculty in Engineering Award by Venus International Foundation, Chennai, in 2017 and Star Achiever Award from Accenture Services Pvt. Ltd. in 2016. She has published 175 research papers in national/international journals and conferences. She is the incharge of college website design and maintenance. She delivered sessions as a resource person in various FDPs/workshops in the areas of machine learning, deep learning, and IoT. She is recognized as supervisor and co-supervisor for various universities like JNTUH, VTU, JNTUA, KLU, and JNTUK. Twelve research scholars have been awarded with Ph.D. and three are pursuing under her guidance. She is the member of Governing Body, GNITS. Curriculum development is done for two specialization courses in CSE (artificial intelligence and machine learning), and CSE (data science) under her headship. She serves as an Editorial Board Member, Reviewer, Advisory/Program/Steering committee member of various national/international journals and conferences. She is the BOS

member, subject expert for Staff Selection Committee for other colleges and also a Life member of ISTE, IETE, CSI, and Fellow of IE.

Sateesh K. Peddoju is currently associated with the Indian Institute of Technology Roorkee, India, since 2010 and has 25+ years of teaching, research, and consultancy experience. He is the Senior Member of ACM, Senior Member of IEEE, and Member of various other councils, including IETE, CSI, ISTE, DMTF, OPF, and HPC Advisory Council. He receives the Cloud Ambassador Award from AWS Educate, IBM SUR Award, Microsoft Educate Award, the University merit scholarship, the best teacher award in his previous employment, the best student thesis award, the best student internship award, the best paper/presentation awards, and travel grants. Sateesh has more than 65 publications, of which 19+ are published in reputed journals, and 35+ are published in high-quality conferences. He is the co-author of the book *Security and Storage Issues in the Cloud Environment* published by Springer and co-editor of six books by Springer and two by others. Sateesh is the Treasurer for the IEEE Roorkee section and Founding faculty sponsor for ACM IIT Roorkee Student Chapter. He was the Program Chair, Organizing Chair, and Web Chair for IEEE MASS 2020 and Founding Steering Committee Chair for the SLICE workshop series. Sateesh was the Publicity Chair for BDA-2022, Program Chair for ICICC-2022, ICCISC-2022, and ICRTC-2022. He has contributed INR 700+ lakhs in grants from various funding agencies due to research and consultancy. Under his supervision, six doctoral and 44 master-level students were awarded the thesis and are currently guiding two master's and 11 doctoral students. Sateesh is involved in various national/international committees, including being the Chair for the Communications subgroup of the IoT Security Workgroup constituted by MEITY, Government of India, and an Expert member in the CERT-UK (Cyber Security Guidelines) committee constituted by the Department of IT, Government of Uttarakhand. His research interests include cloud computing, ubiquitous computing, and security aspects.

Vishnu Pendyala is an IEEE Distinguished Contributor and past ACM distinguished speaker with over two decades of experience in the software industry in Silicon Valley. Educational qualifications are Ph.D. (Computer Engineering) from Santa Clara University, California, USA, and Bachelor of Engineering (Computer Engineering), Osmania University, Hyderabad, TS, India. Professional experience includes being a faculty member in Applied Data Science from 2021 to present at San Francisco Bay Area, USA, and adjunct faculty member from January 2019 to 2021 at San Jose State University, San Francisco Bay Area, USA; Infosys technical management team member at Verizon during 2020 and 2021; Cisco Systems software engineer/technical leader during 2010–2020 at San Francisco Bay Area, USA; and Synopsys senior engineer during 2001–2010 at San Francisco Bay Area, USA. He received the Ramanujan Memorial Gold Medal at the State Math Olympiad. He is a senior member of IEEE and published papers in various international conferences. He served as a technical paper reviewer for journals and conferences from 2003 to 2007. He served as an area governor with Toastmasters International and received

the Distinguished Toastmaster, Area Governor of the Year, and Silver Scribe awards. He taught a one-week course sponsored by the Ministry of Human Resource Development, Government of India, on "Big Data Analytics for Humanitarian Causes" under the GIAN program to 60 research scholars of Computer Science in November 2017. During his recent 3-year term as an ACM Distinguished speaker and before that as a researcher and industry expert, he gave numerous (50+) invited talks in conferences, faculty development programs, and other forums. His book, *Veracity of Big Data*, is available in several libraries, including those of MIT, Stanford, CMU, and internationally. Two other books on machine learning and software development that he edited are also well-received and found place in reputed libraries.

Dr. Vedula V. S. S. S. Chakravarthy is a professor in the Department of Electronics and Communication Engineering and Dean R&D at Raghu Institute of Technology (A), Visakhapatnam. He is the member of Communication, Signal Processing, Antenna and Propagation, MTT, Education Societies, and WIE affinity group of IEEE. He is serving as Chair of IEEE Communication and Signal Processing Societies Joint Chapter of IEEE Vizag Bay Section. He is also serving as the branch counselor of IEEE SB at RIT. He is currently volunteering the Educational activities of APS and Section representative for DART group of IEEE India. During his officer role tenure, the SB established Computer Society student Chapter and Signal Processing student Chapter. He served as publication committee member of ICISGT conference organized by Vizag Bay Sub-Section. He co-edited three volumes of Lecture Notes in Electrical Engineering published by Springer Inc as part of ICMEET 2016, ICMEET 2019, ICMEET 2021, and ICMEET 2022 conferences. His research interests include computational intelligence, smart antenna, data modeling, machine learning, and evolutionary computing tools. He has 20 years of teaching and two years of industry experience. He has four research scholars working in the field of smart antennas, evolutionary computing tools, and high-impedance antennas, and one scholar got awarded. He is a senior member of IEEE, member of Applied Computational Electromagnetic Society (ACES), Senior life member of Wireless, Antenna and Microwave Symposium (WAMS) and life member of professional bodies like Instrumentation Society of India, the International Computer Science and Engineering Society (ICSES), and Soft Computing Research Society. He published more than 40 journal and conference papers along with one book chapter which are indexed in SCOPUS and SCI. He is the associate editor of *International Transactions on Evolutionary and Meta heuristic Algorithms* published by the International Computer Science and Engineering Society and guest editor to special issue in *Soft Computing Journal* published by Springer-Nature.

Cross-Site Scripting Recognition Using LSTM Model

Ishan S. Joshi and **Harsh J. Kiratsata**

Abstract Cross-site Scripting (XSS) is one of the most prominent types of web application assaults. It is one of the most serious hazards to online applications and one of the top vulnerabilities, according to the Open Web Application Security Project (OWASP). As a result, detecting and countering this attack is crucial. As a result, we've demonstrated how to leverage the Long Short-Term Memory (LSTM) model to detect the XSS attack script in this paper. First, we acquired several scripts for the dataset from the XSS vulnerability archives, and then we preprocessed the data by generalizing it, tokenizing it, and changing text to the sequence. The processed data then is trained and tested using the LSTM model, a Recurrent Neural Network. With a precision rate of 99.57% and an f1 score of 99.78%, the proposed approach can yield a substantial outcome.

Keywords Deep learning · Long Short-Term Memory · Cross-site scripting · Xss detection · Cyber security

1 Introduction

Due to ever-increasing advancements in computing power, data quantities, and software applications, the network-dependent application has become the attack vector, becoming more sensitive to cyber-attacks. One of the cyber-attacks, cross-site scripting (XSS) is a sort of injection attack in which malicious scripts are injected into a website in order to evade access restrictions or obtain information about other users, such as authentication cookies or tokens. As per the Open Web Application Security Project (OWASP), XSS attacks have been the major threat to web applications. There are generally three types of XSS vulnerabilities as shown in Fig. 1.

I. S. Joshi (✉)
Adani Institute of Infrastructure Engineering,Adalaj, India
e-mail: ishan25j@gmail.com

H. J. Kiratsata
Bhaskaracharya National Institute for Space Applications and Geo-informatics,Gandhinagar, India

1
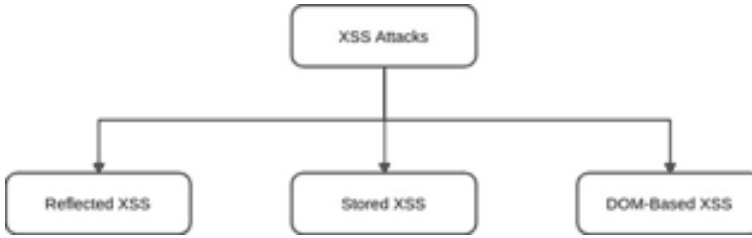
Fig. 1 Types of XSS vulnerabilities

Reflected XSS

Reflected XSS which is also known as non-persistent vulnerability, which is when the script is injected the effect is seen on the browser instantly and the script gets removed when refreshed. The script is activated through a request to a website with a vulnerability that enables execution of malicious scripts. Following are the chain of steps of exploiting the vulnerabilities of non-persistent XSS attack:

- The attacker first builds a URL with a malicious JavaScript string and sends it to the web browser.
- The victim is duped into requesting the URL address from the VWA by the attacker.
- In the HTTP response, the VWA includes the maliciously injected JavaScript string from the URL address.
- The victim's web browser executes the injected JavaScript within the HTTP response message, sending the victim's credentials (passwords, cookies, and so on) to the attacker's server.

Stored XSS

Stored XSS, also known as persistent XSS, is a vulnerability when the malicious scripts gets stored due to the susceptibility in the web application. It works by storing the script in the database, which afterwards gets executed every time when the web application is accessed by the user. Following are some sequence of steps which explains the scenario of exploiting the vulnerability of stored XSS attack:

- The attacker injects a malicious JavaScript string into the website's repository via one of the web application forms.
- The victim goes to the VWA and asks for a web page from the web application.
- In the HTTP response message, the web application includes the malicious JavaScript string from the web application's repository and sends it to the victim's browser.
- The malicious JavaScript code in the HTTP response message is executed by the victim's web browser, which then sends the victim's credentials to the attacker's web server.

DOM-Based XSS

DOM-Based XSS, also known as Type-0 XSS, is a vulnerability where the payload is executed as a result of modifying the DOM "environment" through client-side script execution. So, basically it uses a kind of malicious script that alters the DOM or in layman's language the HTML elements, tags, and attributes properties. Following are the chain of steps of exploiting the vulnerabilities of DOM-Based XSS attacks.

- Through one of the website's web application forms, the attacker injects a malicious JavaScript string into the repository.
- The victim visits the VWA and requests a page from the web application.
- The malicious JavaScript string from the web application's repository is included in the HTTP response message and sent to the victim's browser by the web application.
- The attacker's web server receives the victim's credentials once the malicious JavaScript code in the HTTP response message is performed by the victim's web browser.

Here is the example of a basic XSS script:

"'>><SCRIpt> alert("XSS attack")</scRIpt>

In this paper, we have shown the novel technique to recognize XSS script by means of deep learning's Recurrent Neural Network model named LSTM model. Our major work is as follows:

- We propose a model which can efficient detect the malicious XSS scripts, helping to prevent the attack.
- We have used tokenization to transform the text into the sequence having the length equal to the max length of the data using sequence padding.
- We have developed a model which can achieve a precision rate of 99.48% and an f_1 score of 99.74%.

2 Related Work

Kuppa et al. [1] proposed an approach named ConvXSS for the XSS detection using the convolutional neural network. In this approach, the data is preprocessed and sanitized for better accuracy and speed compared to other approaches.

Hussainy et al. [2] proposed a method to prevent web attacks such as SQL injection attacks and XSS attacks by using the deep learning model.

Gupta and Gupta [3] have proposed a brief description about the XSS attacks and defence mechanisms. In the paper, they have mentioned about XSS vulnerabilities, different types of XSS attacks, various incidents which have occurred earlier, real-world XSS worms, and the state-of-the-art techniques to defence against it.

Gupta et al. [4] have presented a cloud-based framework extracts the suspicious HTML5 strings from the latent injection points and performs the clustering of that string based on the similarity between them. In addition, it detects the variation in the HTML5 code which is embedded in the HTML response.

Yan et al. [5] have performed a survey on applying deep learning for code injection detection. In this paper, they have mentioned the limitation of conventional signature-based detection techniques and have mentioned the approaches such as network-based Intrusion Detection which would improved the performance as well as accuracy.

Yan et al. [6] proposed an approach which uses hybrid deep learning network, and for feature extraction, they used Abstract Syntax Tree of JavaScript to extract key features.

Wang et al. [7] have proposed a technique to discover malicious JavaScript code by using the sparse random projection, deep learning, and logistic regression. In this paper, auto-encoders are used to extract high level featured from JavaScript code, and the logistic regression is used as a classifier to distinguish between the malicious and generous JavaScript code.

3   Propose Solution

In this paper, we have introduced the model to recognize XSS script and have described the approach of how to gather data by using web scraping and web crawling, how to pre-process and filter out data by generalizing the data, transform data into proper UTF-8 format and tokenizing data for training as well as shown how to develop an efficient model using the LSTM model. Here, we have classified data as XSS and non-XSS. [8–17] are referred for the preprocessing as well as generalizing the data and for developing efficient fine-tuned model.

3.1   Data Gathering

In our work, we are using [18, 19] sites which are archives of different XSS scripts vulnerabilities. So, we prepared a dataset by gathering XSS scripts with a basic web scraper and web crawler, acquiring a site's HTML with a requests module in Python and filtering out XSS scripts with regular expression.

3.2   Data Pre-processing

After data gathering, the data is passed through the decoder to transform into the proper format. As, the data which is being extracted through web scraping consists of
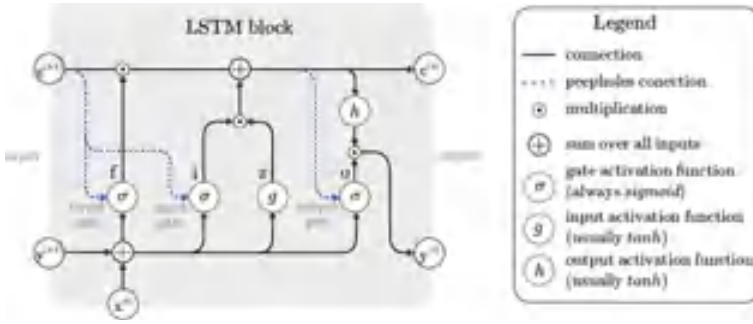
Fig. 2 Architecture of LSTM model

XSS scripts in multiple formats like hexadecimal digits such as %3D, %23, and %3F and HTML character entities such as &equal;, #, and ? which are after decoded into =, #, and ? by a custom decode function. Moreover, the XSS script consists of similar patterns that are generalized into some distinct formats. For example, <script>alert("hello")</script> and <script>alert("XSS attack")</script>have same pattern hence can be generalized as <script>alert(0)</script>. The data is sanitized by filtering and generalizing it to remove extraneous information, which reduces memory usage and saves computational resources. Following sanitization, the data is divided into training and testing data, each of which is tokenized separately from text into a sequence with a length equal to the maximum length of the data entity available via sequence padding. The word dictionary is built from this token to produce a sequence that may be used to train a model.

3.3 Training a Model

The processed data is then passed to the LSTM model. LSTM model is a Recurrent Neural Network model that comes under deep learning which is capable to learn order dependency in sequential prediction tasks. The LSTM model works by using the units consisting of a set of four gates, namely the input gate, the output gate, the forget gate, and the cell gate. The cell gates and the appropriate activate functions such as sigmoid functions as well as tanh functions are used for selecting proportional data. The sigmoid activation functions are used to operate the gates as open or close, while the tanh function is used as the unit state and output selection. Figure 2 shows the architecture of the LSTM model. Here, the input sequence of word indices is converted into embedding vectors to learn word embedding during the training process. After that, it is passed into the LSTM layer, and the dropout is set to prevent overfitting. The XSS recognition is a binary classification problem, so only one dense layer is set for the output. The sequence of steps which is explained above can be seen in Fig. 3.
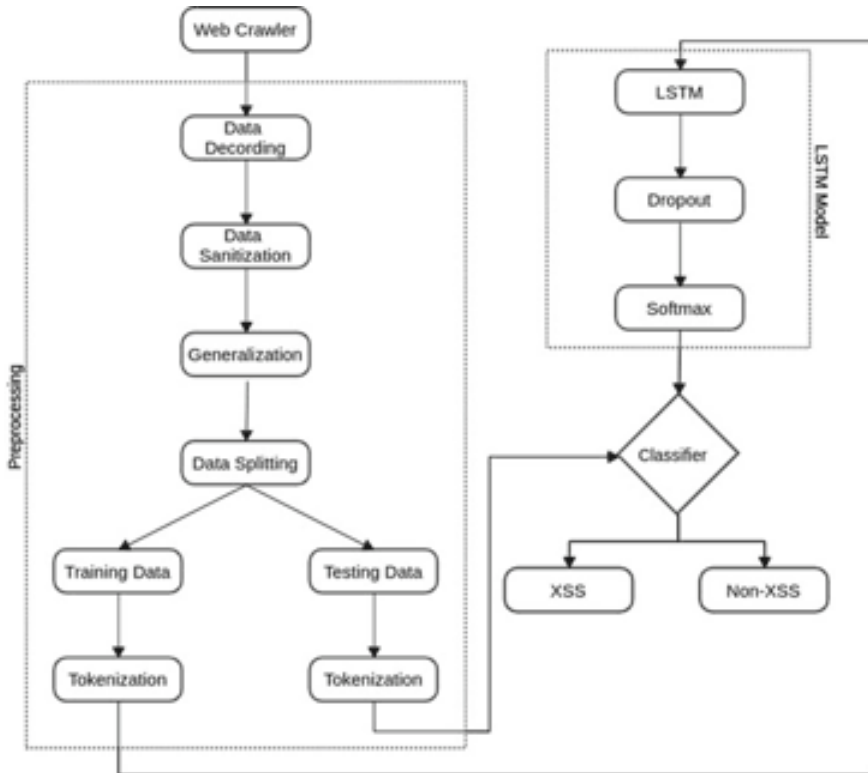
Fig. 3 Flow diagram of the model

3.4 Experimentation Result

For this experiment, our modal has three LSTM Layers, three dropout layers and a single dense layer as an output along with the k-fold cross-validation method where we have kept the value of k equal to 10. The dataset is split into 80% training and 20% testing set. So, in tenfold cross-validation, the training set is further equally divided into ten sub-sets where nine sets are taken for training and the final one is kept for the validation of the model. The cross-validation is repeated ten times where each time the validation of the model takes place, and at the end, the average of the all the validations is taken as a metrics for improving modal. The classifier is tested by feeding it with randomized test split data. Our modal has an outstanding test accuracy of 99.67%. Moreover, the epoch accuracy and the epoch loss while training the model can be seen in Fig. 4.

The XSS recognition middleware is developed, which receives JSON as an input and pairs the XSS key with the payload value. The payload is made up of inputs such as a title, a description, an email, a phone number, a name, an address, or an XSS

(a) Epoch Accuracy

(b) Epoch Loss
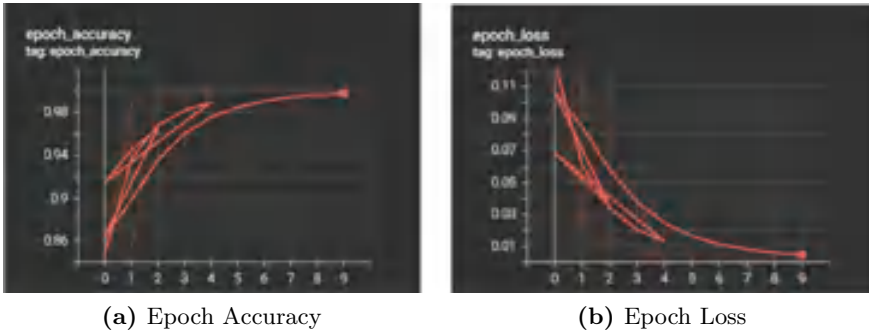
Fig. 4 Epoch accuracy and loss
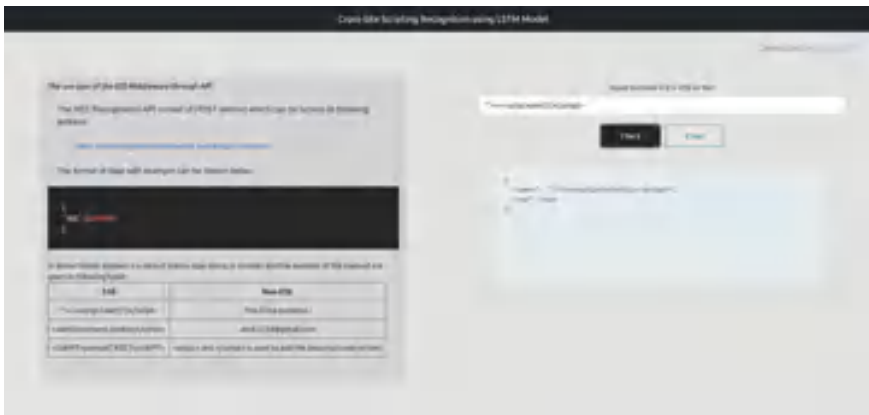
Fig. 5 XSS recognition middleware

script. This payload is delivered via request to the flask-based middleware, where the input is checked using a pre-trained deep learning LSTM model, which classifies the payload as XSS or non-XSS by returning a true or false response along with the given payload in the format described below. Figure 5 shows the XSS recognition middleware which is used for detection. Figure 6 illustrates the results when the XSS is used as an input payload, while Fig. 7 shows the output when the non-XSS strings is passed as an input payload.

4 Limitations

- Due to limited dataset, the model may not be able to detect all the XSS attacks. So, the model may not predict the obscure XSS attacks.

Fig. 6 XSS recognition middleware—XSS payload

Fig. 7 XSS recognition middleware—non-XSS payload

- The model can detect the XSS attack, but for now the external mechanism is required to prevent it or filtering it

5   Conclusion and Future Work

The most significant threat to the web application has been an XSS Vulnerability Attack. So, its detection has become mandatory. We propose an effective recognition technique for XSS detection in this paper, based on one of the deep learning models—Long Short-Term Memory (LSTM) and a Recurrent Neural Network (RNN) model. To begin, we collect data using a web crawler, which is subsequently encoded and sanitized in order to transform it into a more general form. Data preparation plays an important role in the accuracy and performance of any model. Second, the processed data is divided into two sets: training and testing, each of which is tokenized separately. The tokenized training data is subsequently sent to the model's word embedding layer. Then, using the Long Short-Term Memory (LSTM) model, the classifier is generated. In a real dataset, tenfold cross-validation is used for parameter tuning and model evaluation. The model is evaluated on test split set with a significant test accuracy of 99.67%. The model has a precision rate of 99.57% and a f1 score of 99.78% after the evaluation. Furthermore, this model can not only predict conventional XSS attacks, but it can also detect obfuscated XSS attacks to a degree.

For the future work, we can gather much more variety of XSS data to increase the knowledge of the model and to avoid obscurity of XSS attack. Futhermore, the prevention or the filtering mechanism can be incorporate to filter out malicious script.

References

1. Kuppa K, Dayal A, Gupta S, Dua A, Chaudhary P, Rathore S (2022) ConvXSS: a deep learning-based smart ICT framework against code injection attacks for HTML5 web applications in sustainable smart city infrastructure. Sustain Cities Soc 80:103765
2. Hussainy AS, Khalifa MA, Elsayed A, Hussien A, Razek MA (2022) Deep learning toward preventing web attacks. In: 2022 5th International conference on computing and informatics (ICCI), 2022, pp 280–285, https://doi.org/10.1109/ICCI54321.2022.9756057
3. Gupta S, Gupta BB (2015) Cross-site scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art. Int J Syst Assur Eng Manage 8. https://doi.org/10.1007/s13198-015-0376-0
4. Gupta B, Gupta S, Chaudhary P (2017) Enhancing the browser-side context-aware aanitization of suspicious HTML5 code for halting the DOM-based XSS vulnerabilities in cloud. Int J Cloud Appl Comput (IJCAC) 7(1):1–31. http://doi.org/10.4018/IJCAC.2017010101

5. Yan R, Xiao X, Hu G, Peng S, Jiang Y (2018) New deep learning method to detect code injection attacks on hybrid applications. J Syst Softw 137:67–77

6. Yan R, Xiao X, Hu G, Peng S, Jiang Y (2018) New deep learning method to detect code injection attacks on hybrid applications. J Syst Softw 137:67–77

7. Wang Y, Cai WD, Wei PC (2016) A deep learning approach for detecting malicious JavaScript code. Secur Commun Netw 9(11):1520–1534

8. Pan Y, Sun F, White J, Schmidt DC, Staples J, Krause L (2018) Detecting web attacks with end-to-end deep learning. Vanderbilt Univ, Melbourne, FL, USA, pp 1–14

9. Yao L, Guan Y (2018) An improved LSTM structure for natural language processing. In: 2018 IEEE international conference of safety produce informatization (IICSPI), pp 565–569. IEEE

10. Khazal IF, Hussain MA (2021) Server side method to detect and prevent stored XSS attack. Iraqi J Electr Electron Eng 17(2)

11. Gupta BB, Gupta S, Chaudhary P (2017) Enhancing the browser-side context-aware sanitization of suspicious HTML5 code for halting the DOM-based XSS vulnerabilities in cloud. Int J Cloud Appl Comput 7(1):1–31

12. Likarish P, Jung E, Jo I (2009) Obfuscated malicious javascript detection using classification techniques. In: Proceedings of 4th international conference malicious unwanted software (MALWARE), pp 47–54

13. Wang R, Jia X, Li Q, Zhang S (2014) Machine learning based crosssite scripting detection in online social network. In: Proceedings of IEEE international conference high performance computer communication IEEE 6th international symposium cyberspace safety security IEEE 11th international conference embedded software system (HPCC, CSS, ICESS), pp 823–826

14. Rathore S, Sharma PK, Park JH (2017) XSSClassifier: an efficient XSS attack detection approach based on machine learning classifier on SNSs. J Inf Process Syst 13(4):1014–1028

15. https://towardsdatascience.com/deploy-a-machine-learning-model-using-flask-da580f84e60c

16. Sari WK, Rini DP, Malik RF, Azhar ISB (2020) Sequential models for text classification using recurrent neural network. In: Sriwijaya international conference on information technology and its applications (SICONIAN 2019), pp 333–340. Atlantis Press

17. https://towardsdatascience.com/multi-class-text-classification-with-lstm-1590bee1bd17

18. Xssed.com (2022) XSSed | cross site scripting (XSS) attacks information and archive (online). Available at: http://www.xssed.com. Accessed 11 Mar 2022

19. GitHub (2022) xss-payload-list/README.md at master . payloadbox/xss-payload-list. (online) Available at: https://github.com/payloadbox/xss-payload-list/blob/master/README.md. Accessed 11 Mar 2022

Development and Implementation of Efficient Recruitment Portal Using Blockchain

Vijay Mane, Parag Satpute, Kaustubh Wakodkar, Sameeran Zingre, and Datta Sarwade

Abstract Blockchain is an immutable and distributed ledger that can be used to create decentralized applications or Dapps. Dapps are the applications where the data is not centrally stored at a single source rather are stored in a decentralized network. Our data is sensitive and its safety is crucial when we put it on the Internet so, by creating records in Blockchain data access can be controlled and prevent fraud as all the data in the Blockchain is end-to-end encrypted. In this paper, we present a Recruitment Portal developed using Blockchain. The current recruitment process is long and hectic as most of the work has to be done manually. Using Blockchain for this can eliminate the time-consuming and paper heavy traditional process. The proposed application is developed using the Ethereum Blockchain where candidates and recruiters can register to the Blockchain. Recruiters can upload their job openings with all job-related details and job descriptions and candidates interested can read job details and apply for them with their resumes.

Keywords Blockchain · Decentralized application · Ethereum · IPFS · Smart contracts · Web3.js

V. Mane (✉) · P. Satpute · K. Wakodkar · S. Zingre · D. Sarwade
Vishwakarma Institute of Technology, Pune, India
e-mail: vijay.mane@vit.edu

P. Satpute
e-mail: parag.satpute18@vit.edu

K. Wakodkar
e-mail: kaustubh.wakodkar18@vit.edu

S. Zingre
e-mail: sameeran.zingre18@vit.edu

D. Sarwade
e-mail: datta.sarwade18@vit.edu

1 Introduction

Blockchain is one of the most anticipated and famous technology in this century. Due to its great usage, various applications and advantages over traditional methods, this technology is rising exponentially in recent years [1]. Information drives the business. If this information is faster and accurate it is always better for business needs. For delivering that information, Blockchain technology is ideal since it provides immediate and transparent information which is stored on an immutable ledger that can only be accessed by the people who have the permissions [2]. Orders, payments, accounts, production and much more can be tracked on a Blockchain network. Members also have a single view of the truth, allowing them to see the details of a transaction end to end, which gives them greater confidence and opens up new opportunities. Blockchain consists of three key concepts: blocks, nodes and miners [3]. Blockchain technology has been incorporated by many businesses around the world in recent years [4]. So, what is it and how exactly does it work? Does it make a significant difference or is it only an addition? Blockchain technology is still in its infancy but has the potential to revolutionize many industries in the future. A peer-to-peer network containing a shared ledger is a means to store the transactions and records of the network. Blockchain contains three important and leading technologies. (1) Cryptographic keys. (2) A means of computing, to store the transactions and record of the network. (3) A peer-to-peer network containing shared ledger.

The private key and public key are the two components of a cryptographic key. Together, they enable two parties to transact successfully. These two keys are used by every individual to create a secure digital identity reference. Blockchain technology provides a secure identity. In the world of cryptocurrencies, this type of identity is said to be a digital signature and it is used to identify and authorize various transactions. The peer-to-peer network and the digital signature are merged. A large number of people acting as main authorities use the digital signature to do various transactions because of its usage and advantages over traditional methods of signature [5]. A deal between a peer-to-peer network and a digital signature is secured through mathematical verification, which results in a successful transaction. In short, Blockchain users use cryptography keys to perform different types of digital interactions. Blockchain will eventually saturate the commercial world, posing significant challenges for HR and the workforce [6]. There was also a feeling that technology has degraded trust in recent years, particularly with the rise of cyber dangers [7]. Blockchain technology proposes a method for regaining lost trust through the use of technology. HR functions must be restructured. The entry of the Industrial Revolution, in which the digital era continues to develop, has had no effect on the company's employee recruitment system [4]. The old manual recruitment process is being phased out in favors of a database-driven digital system. It is possible in the coming years, the limited manual recruitment method will be abandoned in favors of a more efficient digital-based recruitment trend and everything will be incorporated into a Blockchain-based system.

Blockchain is a distributed or open ledger that can record transactions between two parties in an efficient and verifiable manner and it is also a permanent. Blockchain is a technological wonder that establishes trust through agreement by ensuring that all parties permitted to access the Blockchain, agree on any modifications made, which will assure the data's validity [8]. Surprisingly, with the Blockchain technology system, even the tiniest modification must be confirmed by the network in advance. This indicates that all information which is entered is reliable and has successfully passed the authentication process. If a modification or some changes are made in the data, then all the people who has access to the database are communicated about the changes [9]. Old data storage systems, on the other hand, tended to be centralized on a single computer. Blockchain technology has a bright future worldwide. Blockchain technology has been observed in the financial industry to have incredible scope. The financial sector was unable to cope with the heavy workload after demonetization. It brought to light the difficulties of having a centralized approach of specialist handling financial transaction. RBI has therefore encouraged banks to digitize as much as possible. In addition to this, the company released a statement highlighting Blockchain's potential to combat fakes. In addition, it highlighted its potential to bring about specific changes to the way financial markets, collateral identification and payments work. The incorporation of Blockchain with financial transactions yields many benefits, including the ability to save substantial amounts of time and money [10]. In addition, there is a drastic reduction in the time needed to process and validate transactions. A distributed database powers the Blockchain, ensuring great amount of security and safeguarding it from various digital attacks such as cybersecurity, phishing.

2 Literature Survey

Extensive research has been conducted on Blockchain and various scholars, students and enthusiasts have done excellent work on Blockchain. This section dissusses briefly about the literature survey done for the project to study work done on Blockchain. Ahmed, Olov and Karl presented their study of Blockchain from the perspective of applications, challenges and the future [1]. The paper gives detailed information about Blockchain architecture and also discusses scalability, privacy, challenges and the future. Akhil, Anjana, Ritika and Shantagouda presented their paper "Simplified Recruitment Process" [2] and in this paper they have given a solution for the tedious recruitment procedure in a simplified manner with the help of Blockchain. They believe by using the solution they proposed the companies can save their time and money.

Md Onik, M. H. Miraz and Chul-soo Kim presented two systems, BcRMS and BcHRMS in paper "A recruitment and human resource management technique using Blockchain technology for industry 4.0" [3]. The solution proposed in the paper holds definite advantages compared to the conventional recruitment system. Umesh, Sundeep and their colleagues have done an extensive study on Blockchain for industry

4.0 [4]. In their paper, they analyzed the current industry 4.0 and the influence as well as the future of Blockchain in the current industry. The paper also provides information on various Blockchain technologies and their applications. Toqeer and his colleagues in their paper presented a comparative analysis of core Blockchain architecture. The paper also presents the applications in 3 domains Healthcare, IoT and the Vehicular industry [5].

Qalab and his colleagues presented a paper on Blockchain discussing various applications and technologies that are taking place in this domain [6]. When some technologies are being used in a large scale, one of the most important point is to avoid difficulties and challenges while using that technique. With this point, Wazheng Li and his colleagues in their paper discussed about the future trends and challenges while using Blockchain [8]. The paper is useful to get idea about the future trends as well as the challenges of Blockchain technology. Today, use of cryptocurrency has increased a lot. Henry and his colleagues presented a detailed literature survey on Blockchain in their paper [10].

Mr. Qin, Mr. Wang and Mr. Jiang in their paper "RPChain: A Blockchain-Based Academic Social Networking Service for Credible Reputation Building" [11], presented a Blockchain-based academic social network model which has peer review records that are irreversible, reputation building that can be traced and proper incentives for content contribution. T. Poongodi, R. Sujatha and the team in their paper "Blockchain in Social Networking" [12] presented a non-centralized or decentralized approach to Blockchain technology in the area of social networking. The paper states the disadvantages and the limitations of the conventional centralized way of social networking and how it can be made effective with the help of Blockchain via a decentralized way. Chian and Pinyaphat in their paper, they discussed the applications and challenges of Blockchain technology and presented general ideas about the usability of Blockchain [13].

Chandan, Preeti and Amit presented a paper on Blockchain in the educational sector [14]. The paper provides a systematic overview of Blockchain technology in the educational sector. The paper states that the use of Blockchain technology in the educational sector is still a new sector and is evolving. In the coming years, we will see the use of Blockchain technologies in the educational sector. IoT and Blockchain are the most popular and useful technologies in this century. Mayra and Ralph presented a paper to integrate two technologies, i.e., IoT and Blockchain [15]. To overcome the key challenge of hosting location the paper evaluates the use of fog and cloud as hosting platforms. Gareth R. T. White presented a paper on the future of Blockchain in business and management [16]. The paper presented the study in Blockchain and its future usability in business and management. Papers [7, 17] discuss about crypto currencies, their scope and use in the domain of Blockchain.

3 Methodology

3.1 Tools and Technologies

Solidity: Solidity is a high-level object-oriented, curly braces and statically typed programming language specially developed by Ethereum Network team for creating smart contracts. Smart contracts contains all the functions required to interact with the data on the Ethereum Blockchain. Smart contracts written in solidity programming language are programs that are run on a peer-to-peer network with no central authority.

Truffle: Truffle is an Ethereum programming environment and testing framework aimed at making contract testing and deployment simpler for Ethereum developers. With Smart contract compilation, linking, deployment and binary management are all embedded into Truffle. It's a scriptable deployment and migrations framework, as well as a configurable build pipeline with support for bespoke build processes. It also has an interactive console for direct contract negotiation and the ability to recreate assets in real time throughout development.

IPFS: IPFS or Interplanetary File System is a platform where files, websites, apps and data are stored and accessed through a distributed system. It is a decentralized and immutable file storage system where contents cannot be changed but can attach new version of file to old one. When file is uploaded a hash is provided which is the ultimate way of accessing file. When someone requests a file or a webpage, their node caches a copy of the file. As more users seek that data, there will be more cached copies available. The decentralized web employs content-based routing instead of address-based routing, which requires knowing the location of the data and providing a precise URL to that data.

Web3.js: Web3.js is among the JavaScript libraries that allow us to communicate with an Ethereum node via the internet or locally. Our data is on our Blockchain and user communicate with the frontend so this gap is fulfilled by Web3.js. Web3.js has its own defined functions to interact with the Blockchain like web3.eth.Contract, web3.eth.accounts, web3.eth.abi which helps in getting contract deployed on Blockchain, accounts and abi of the contract deployed.

Ganache: For Ethereum development, Ganache is a customizable Blockchain. It is locally running Ethereum Blockchain that may be used to check how the chain works when running tests, executing commands and inspecting states. It has ten dummy accounts with 100 ethers each and this can be used to test the smart contracts.

3.2   Proposed Method

The project's main objective is to create a portal using Ethereum Blockchain where recruiters can upload job details and candidates can apply for the job. Figure 1, shows the system model for the project which clearly explains the model of the project.

The recruiter can register to the Blockchain and all their job openings would be recorded on the Blockchain. Even though the Blockchain is a decentralized network, whenever the data is submitted to the Blockchain it will be validated. To add data in the Blockchain means to perform a transaction and to validate this transaction a secured identity is created using the public key and private key. After a successful transaction, it needs to be authorized which is done by other computers on the network. Now candidates can see all job openings present on the Blockchain and see if any opportunity excites them. If they are interested, they can apply for the jobs with their resumes. The resumes they share will be stored on IPFS. The candidates can keep track of all the jobs they have applied for on this portal. Similarly, recruiters can see all the candidates applying for their jobs. They can see all the applied candidates' resumes which would be fetched from IPFS.
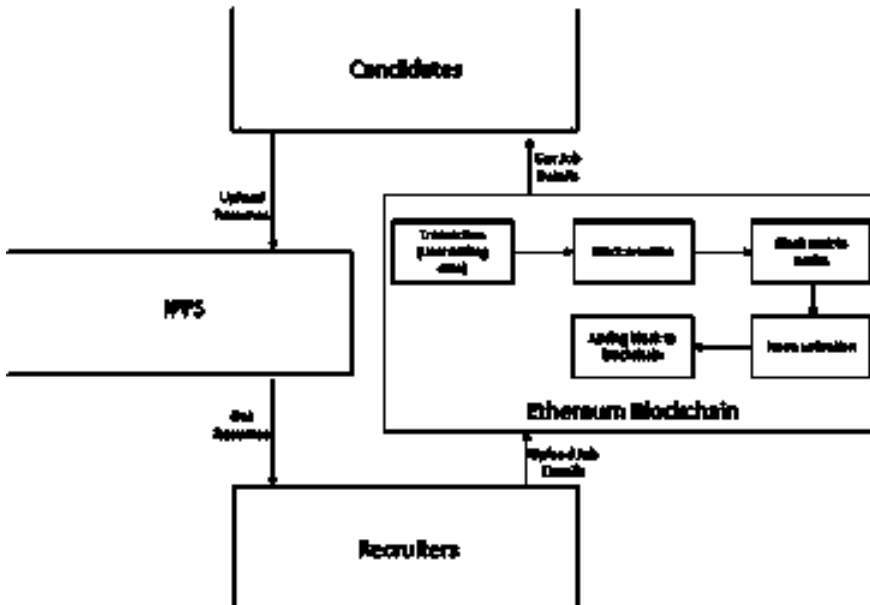
Fig. 1 System model

3.3 Implementation and Working

The previous part explains our proposed method for the system and this section discusses about how the project is implemented and how it works. Figure 2, explains the implementation and flow of all the platforms in the project. The heart of the project is the smart contracts which are written in Solidity programming language. It contains all the functions required to interact with the data on the Blockchain. Truffle as explained in libraries used in the earlier part of this section is used to compile and deploy the smart contracts to the Blockchain which is done using "truffle compile" and "truffle migrate" commands, respectively. As the contract is deployed on the Blockchain a hash is provided which is the contract address. The Blockchain used for the project is Ganache which is a local Ethereum-based Blockchain used for testing purposes. The users using the project would be interacting with the frontend but all the data is present on the Blockchain and IPFS so to interact with them two a special JavaScript library is used called Web3.js. Web3.js has its own defined functions to interact with the Blockchain and render all the data on frontend.

The candidate will have an HTML page that shows all the jobs available on the Blockchain and he wants to apply for a job. When he applies with his resume first, he needs to approve the transaction from his cryptocurrency wallet which is metamask in this project and then the resume would be stored in IPFS and candidate applications would be recorded on the Blockchain through the smart contract deployed on Blockchain. A hash would be generated for an uploaded resume and is stored on the Blockchain. Later using the same hash company can access the candidate's resume.
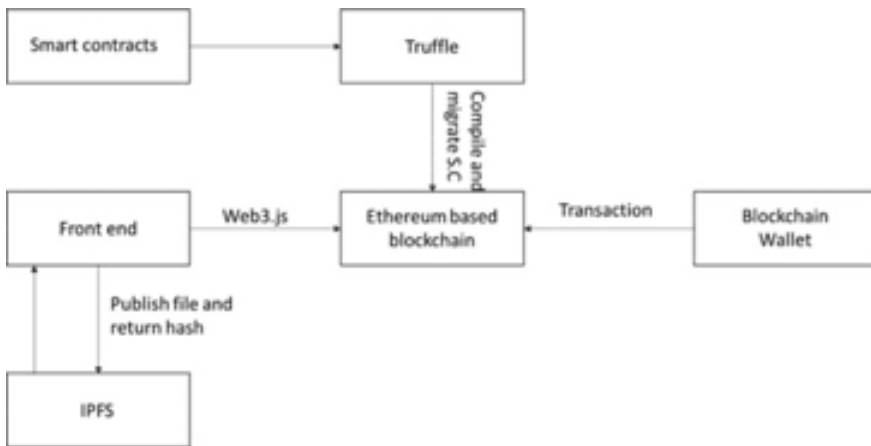
Fig. 2 Implementation

4 Results and Discussion

Using Blockchain-based applications have multiple benefits over normal websites like it provides more transparency, more security, speed and efficiency.

1. Higher Transparency: In normal websites data is stored in a database to which someone who is administrating it has access on. Whereas all the data uploaded and used in the complete project is stored on Blockchain meaning in a decentralized network and not on one single server thereby eliminating the chances of misuse of the data. Also, complete data is accessible anytime and anywhere allowing complete transparency to the users.
2. Higher Security: All the data once recorded cannot be changed and is E2E encrypted with powerful cryptographic algorithms. Data is stored in network of computers unlike normal websites where data is stored at single network. Also, in Blockchain all transactions performed are stamped with time and date and are complete history is stored eliminates any opportunity for fraud.
3. Improved Speed and Efficiency: Uploading the resume and other required documents on IPFS eliminate the need of manually submitting it thus reducing the paper work and making complete process faster and efficient. The smart contracts used are used to automate the transaction which even more increase the speed and efficiency. As all the specified conditions are validated the next process is automatically started and all this reduces the dependencies on third party services.

5 Conclusions

This paper discussed about a project which demonstrate how to use Blockchain to develop a recruitment portal. Blockchain is one of the most trending subjects today and has the potential to revolutionize people's lives in almost every area. We have presented a flawless, easy-to-interact model for the recruitment process. The model designed is developed using the Ethereum Blockchain. The model is suitable for both the recruiters and candidates. Recruiters can upload various job openings and candidates can apply for relevant jobs based on their profiles or preferences. In the project smart contracts were created by Solidity programming language and Truffle was used to compile and deploy the smart contracts in the Blockchain whereas Web3.js was used to connect Blockchain to the frontend.

References

1. Monrat AA, Schelen O, Anderson K (2019) A survey of blockchain from the perspectives of applications, challenges and opportunities
2. Bhavirisetty A, Pradeep A, Chadha R, Kuntoji S (2018) Simplified Recruitment Process

3. Onik MMH, Miraz MH, Kim CS (2018) A recruitment and human resource management technique using Blockchain technology for industry 4.0. In: Smart Cities symposium 2018, ISBN: 978-1-78561-942-7
4. Bodkhe U, Tanwar S, Parekh K, Khanpara P, Tyagi S, Kumar N, Alazab M (2020) Blockchain for industry 4.0: a comprehensive review
5. Syed TA, Alzahrani A, Jan S, Siddiqui MS, Nadeem A, Alghamdi T (2019) A comparative analysis of blockchain architecture and its applications: problems and recommendations. IEEE Access
6. Abbas QE, Sung-Bong J (2019) A survey of blockchain and its applications. IEEE ICAIIC 2019, Electronic ISBN:978-1-5386-7822-0.
7. Rajput S, Singh A, Khurana S, Bansal T, Shreshtha S (2019) Blockchain technology and cryptocurrencies. In: 2019 AICAI, Electronic ISBN:978-1-5386-9346-9
8. Li W, He M, Haiquan S (2021) An overview of blockchain technology: applications, challenges and future trends. In: IEEE, ICEIEC 2021, ISBN:978-1-6654-0385-6.
9. Dhanala NS, Radha D (2020) Implementation and testing of a Blockchain based recruitment management system. In: ICCES 2020, Electronic ISBN: 978-1-7281-5371-1.
10. Andrian HR, Kurniawan NB, Suhardi (2018) Blockchain technology and implementation: a systematic literature review. In: ICITSI 2018, Electronic ISBN: 978-1-5386-5693-8.
11. Qin D,Wang C, Jiang Y (2018) RPChain: a blockchain-based academic social networking service for credible reputation building. Springer International Publishing, ICBC 2018, LNCS 10974, pp 183–198
12. Poongodi T, Sujatha R, Sumathi D, Suresh P, Balamurugan B (2020) Blockchain in social networking
13. Tasatanattakool P, Techapanupreeda C (2018) Blockchain: challenges and applications. ICOIN 2018, ISBN: 978-1-5386-2290-2.
14. Bhaskar P, Tiwari CK (2021) Blockchain in education management: present and future applications. Emerald Publishing Limited, pp 1741–5659, https://doi.org/10.1108/ITSE-07-2020-0102
15. Samaniego M, Jamsrandorj U, Deters R (2016) Blockchain as a service for IoT. In: 2016 IEEE international conference on internet of things (iThings) and IEEE green computing and communications (greencom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (smartdata)
16. White GRT (2017) Future applications of blockchain in business and management: a delphi study. Strateg Change 26(5):439–451
17. Shakya V, Pavan Kumar PVGN, Tewari L, Pronika (2021) Blockchain based cryptocurrency scope in India. In: ICICCS 2021, Electtronic ISBN:978-1-6654-1272-8.

Machine Learning Classifiers Performance Comparison for Breast Cancer Detection

Vijay Birchha and Bhawna Nigam

Abstract Breast cancer, worldwide is the top reason of mortality amongst women. Breast cancer early detection can minimize mortality and increase the likelihood of survival at a reduced cost. This study assesses the performance of machine learning (ML) classifiers using the Wisconsin (original) breast cancer (WBC) dataset and proposes the classifier with the best breast cancer diagnosis evaluation criteria. On the WBC dataset, the machine learning classifiers K-nearest neighbor, Naive Bayes, Random forest, Neural network (NN), Support vector machine, and Logistic regression were chosen for comparison. Logistic regression scored the most excellent AUC value of 0.9943, while the NN classifier achieved the highest accuracy score of 0.989. The research will aid in selecting a ML classifier for a computer-aided breast cancer diagnosis system.

Keywords Neural network · Breast cancer · Stratified sampling · Wisconsin dataset · Machine learning

1 Introduction

The top cause of mortality for women globally is breast cancer. The World Health Organization (WHO) reports that the largest number of new instances of breast cancer were reported in 2020, and 6.85 thousand individuals passed away as a result [1]. Early diagnosis can help reduce breast cancer mortality and enhance the probability of survival with a reduced cost of treatment [1]. Therefore, machine learning (ML) techniques are becoming valuable for healthcare analysis; the achievement of the computer-aided diagnostic system (CAD) depends on the selection of ML algorithms for a breast cancer diagnosis; there are various ML algorithms available, and

V. Birchha (✉)
Department of Computer Engineering, IET, DAVV, Indore 452001, India
e-mail: vijaybirchha@gmail.com

B. Nigam
Department of Information Technology IET, DAVV, Indore 452001, India
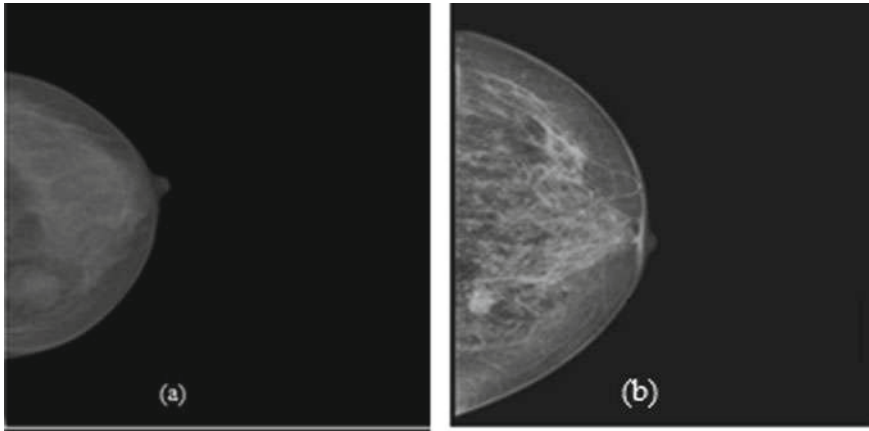
Fig. 1 Breast images **a** benign tumor and **b** malignant tumor [2]

the success of each of them depends on their characteristics and available dataset. This research aims to compare the machine learning classifier's performance on the WBC dataset and suggest best classifier with evaluation parameters for breast cancer diagnosis (Fig. 1).

2 Related work

The following relevant works of literature were studied to discover other authors' contributions to machine learning-based breast cancer diagnosis using the Wisconsin dataset.

In [3], the author compared the ML classifiers random forest (RF), linear and kernel-based support vector machines, k-nearest neighbors (KNN), multi-layer perceptron (MLP), logistic regression (LR), and decision tree (DT) performance [4]; MLP and LR classifier scored the accuracy score of 98%. Similarly, in [5] researchers studied five ML algorithms, Naïve Bayes (NB), DT (J48), KNN, and MLP; the KNN scored the highest accuracy with k-fold dataset split method. In [6] the author compared two ML techniques, Bayesian Networks and J48; The Bayesian Networks achieved 97.80% accuracy. To improve the classifier performance to detect breast cancer, the author in [7] used dimension reduction using Gray Wolf Optimization (GWO). In [8], a CAD system was proposed by combining genetic programming and ML algorithms to differentiate between benign and malignant breast tumors accurately. In [9] the author compared the classifier algorithm's performance on three distinct databases of Wisconsin breast cancer. In [10] author used NB and neural network (NN) ML classifier on the WBC dataset to detect breast cancer. In

Table 1 Recent related work

Aim	Author	Dataset	Methodology
Breast cancer detection using machine learning technics	Houfani et al. [3]	WDBC	Practical comparison between RF, SVM, L-SVM, DTs and MLP
	Kaklamanis and Filippakis [5]	WDBC	Comparative study between NB, IBK, SMO, J48 and MLP.
	Rodrigues [6]	WBC((Diagnostic)	Compared Bayesian Network and J48
	Kamel et al. [7]	WBC	SVM with feature selection using GWO
	Dhahri et al. [8]	WDBC	Hybrid model using genetic algorithm and machine learning
	Salama et al. [9]	WBC, WDBC, WPBC	Compare classification algorithms J48, MLP, NB, SMO and IBK accuracy with other parameters on tenfold cross-validation
	Bohacik [10]	WBC	Data mining techniques NN, MLP, C4.5 and Bayes
	Osman et al. [11]	WBC, WBCD, BCD, BCP,	RBFNN, LR, KNN, SVM, DTs, CNN and NB

[11] the author found that Radial-Based Function Neural Network models (RBFNN) performed well compared to LR, KNN, SVM, DT, CNN, and NB.

Most of the researchers have compared supervised machine learning classifiers; they have commonly used SVM, KNN, MLP, and DT [3, 7]. In addition, some authors used the J48 modified version of DT [5, 6, 9].

During the related literature study following points were observed 1. Authors have used well-known machine learning classifiers, 2. Some of the authors [3, 5, 7] used train-test dataset splitting techniques; [8, 9, 11] utilized k-fold cross-validation techniques. 3. To maintain class balance in dataset splits, nothing was done (Table 1).

In this proposed research work, famous ML classifiers NN, KNN, RF, SVM, NB, and LR were selected to address all the points mentioned above. Furthermore, the k-fold cross-validation dataset splitting technique with stratified sampling is planned to use with the WBC dataset. This research aims to evaluate the various ML classifiers, compare their performance and propose a classifier that scores high on the WBC dataset.

The paper has systematized into many sections; Sect. 1 talks about the basic knowledge of the issue, research objective and earlier pertinent literature work, and Sect. 2 includes the details of the material, machine learning techniques and measures of performance evaluation parameters used; Sect. 3, shows the prediction results of all six classifiers with graphical charts. Section 4 includes the research findings and discuss the limitation. Finally, the future recommendation with a conclusion was given in Sect. 5.

3 Material and Methods

This segment has explained the dataset, preprocessing dataset, and ML methods.

3.1 Breast Cancer Wisconsin Dataset

Breast Cancer WDBC dataset: the WDBC dataset contains 32 features and 569 records; the dataset features include real and multivariate values with zero missing values [12]. Breast Cancer WBC dataset: The WBC dataset has ten features and 699 records; the features are multivariate, and integer values with some missing values in the seventh feature, Bare Nuclei; the dataset also contains 54 duplicate records. Breast cancer features are multivariate and real numbers; it is most suitable for classification and regression problems [13]. All the datasets mentioned above are most appropriate to train the machine learning model for classification. The WBC dataset was selected for the research work because it contains fewer features and a more extensive record set when compared with the WDBC and WPBC datasets; the WBC dataset is most suitable for ML model training due to fewer features; it will consume less computation time for model training. The dataset class feature has two values: benign and malignant. The dataset contains 241 malignant and 458 benign samples. The features name of WBC dataset with statistical information given in the Table 2.

3.2 Machine Learning Algorithms

K-nearest neighbour (KNN) classifier: The KNN is a supervised and non-parametric ML classifier. It utilizes proximity to do forecasts about the group of an individual data point. The working principle is that similar points can be found near one another. The advantages of the KNN algorithm are that implementation is effortless, easily modified and has a small number of hyperparameters [14] (Fig. 2).

Table 2 Breast cancer Wisconsin (original) dataset (WBC)

S. No.	Attribute name	Value	Mean	Range
1	Sample code number: id	–	–	–
2	Clump thickness	1–10	4.42	9
3	Uniformity of cell size	1–10	3.13	9
4	Uniformity of cell shape	1–10	3.21	9
5	Marginal adhesion	1–10	2.81	9
6	Single epithelial cell size	1–10	3.22	9
7	Bare nuclei	0–10	3.46	10
8	Bland chromatin	1–10	3.44	9
9	Normal nucleoli	1–10	2.87	9
10	Mitoses	1–10	1.59	9
11	Class	2 or 4	–	–

Fig. 2 K-nearest neighbour [14]
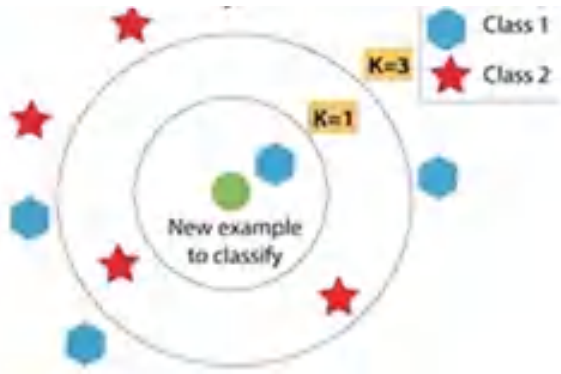
Support vector machine (SVM) classifier: SVM can successfully deal with high numbers of dimensions and avoid the curse of dimensionality, and it is fast to compute. Furthermore, the algorithm maximizes the distance between two sets of data [15] (Fig. 3).

Random Forests (RF) Classifier: when one or many decision trees are trained using a randomly selected training set or input features, their predictions are finally combined; this concept worked behind the random forest algorithm. The working principle increases their overall prediction accuracy. It does not use formulas but is comparatively slower than the decision tree [16] (Fig. 4).

Naïve Bayes (NB) classifier: The Naive Bayes classifier is a set of classification methods based on Bayes' Theorem. Furthermore, in the Eq. (1) given below P(X) and P(Y) is the probability of events X and Y occurring and P(Y) ≠ 0. P(X|Y) denotes the probability of Y happening, given evidence that X has already happened [17].
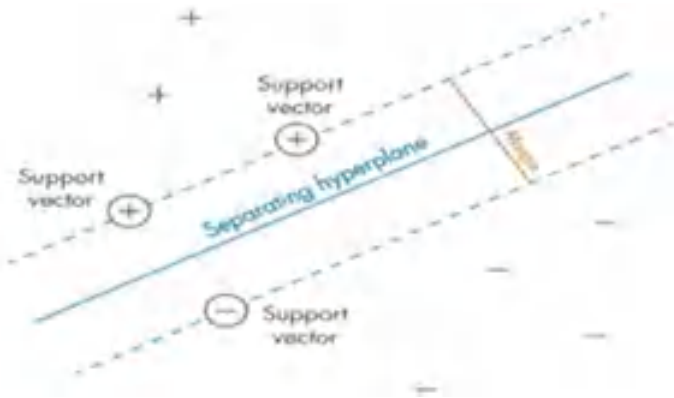
Fig. 3 Support vector machine [15]
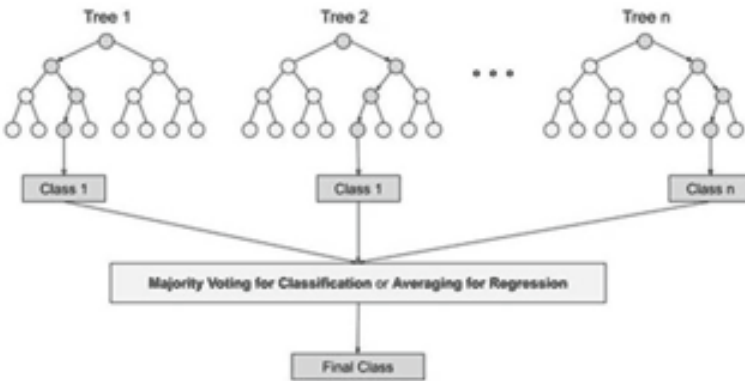
Fig. 4 Random forests [16]

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{1}$$

Artificial Neural Networks (NN) Classifier: Neural networks, which include several linked processing nodes, are models of the way the human brain operates. A neural network's capacity for pattern detection is useful for a variety of tasks, including natural language translation, image production, image identification, and audio recognition. [18] (Fig. 5).

Logistic Regression Classifier—Logistic regression is a supervised ML classifier. LR is an easy and efficient technique for two-class and linear classification problems. It is capable of providing a solution for nonlinear classification problems [19] (Fig. 6).
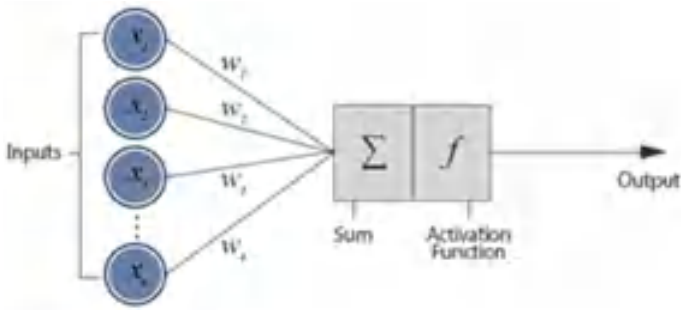
Fig. 5 Artificial neural networks [17]
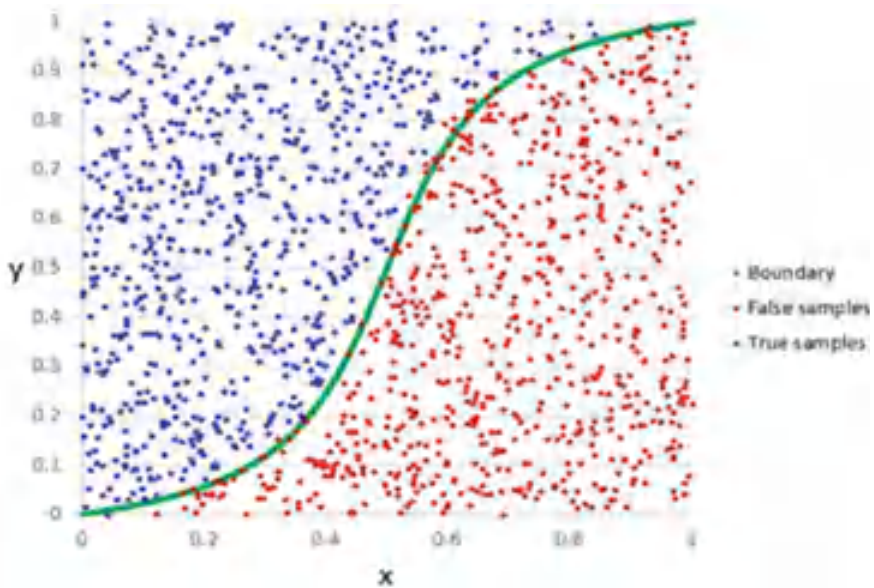
Fig. 6 Logistic regression [18]

3.3 Evaluation Parameters

The performance of the ML algorithm is evaluated using the following parameters.

- Confusion matrix: The matrix summarizes the classifier performance; it presents the count of correctly and incorrectly classified instances. The evaluation parameters of the machine learning classifiers were calculated based on tp, tn, fp and fn values mentioned in the confusion matrix.
- Area Under the Curve (**AUC**): "The area under the receiver operating characteristic (ROC) curve" [20] (Fig. 7).

Fig. 7 Confusion matrix

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}} \qquad (2)$$

$$\text{Precision(PR)} = \frac{\text{tp}}{\text{tp} + \text{fn}} \qquad (3)$$

$$\text{Sensitivity(SE)} = \frac{\text{tp}}{\text{tp} + \text{fp}} \qquad (4)$$

$$F_1\text{-score} = \frac{2 * \text{PR} * \text{SE}}{\text{PR} + \text{SE}} \qquad (5)$$

$$\text{Specificity} = \frac{\text{tn}}{\text{fn} + \text{tn}} \qquad (6)$$

4   Experiment Design and Implementation

4.1   Dataset Preprocessing

Preparing raw data for ML algorithms is known as data preprocessing; before applying the WBC dataset to machine learning algorithms, the duplicate records were removed with the help of the sample id column. The WBC dataset searched for missing values using the attribute value for nil, ? and empty space, the work was done using Microsoft Excel. Some missing values were found in the seventh column Bare Nuclei; the missing values were filled using the mean value of that column. The sample id column is not helpful for further research work, which was removed; the rest of 10 columns were used for performance evaluation. The dataset was normalized
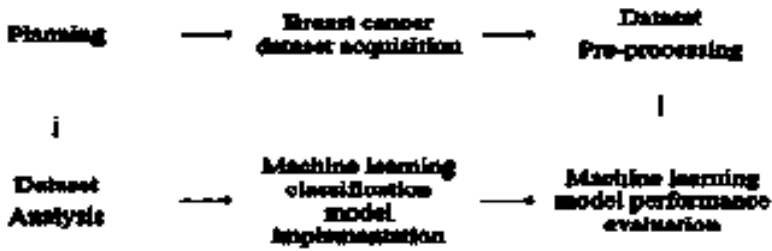
Fig. 8 Research process

by applying the MaxMin method; in the normalization process, the class feature was not considered; at last, the dataset was split into k-folds of equal size.

4.2 Design and Implementation

The experiment was done using a WBC dataset with six different machine learning classifiers; the data analysis was done using Python programming language. Finally, the steps mentioned in Fig. 8 were followed to implement the research process and disseminate the results. The Orange data mining toolkit was utilized for research process implementation and execution.

5 Results and Discussion

In the breast cancer WBC dataset selected for the research work, 645 unique instances out of 699 were employed for the experiment; the dataset was split into 10-folds of the same size, and the stratified sampling technique was used to maintain class balance as the original dataset in folds. The WBC dataset has 11 features, sample id is useless for machine learning model training, and class feature is utilized as a target to check the model performance; the other nine features are applied to predict the breast cancer class benign or malignant.

For this comparative study, six classifiers mentioned in the previous section were utilized. Table 3 stores the performance scores. Table 4 compare ML models by the area under ROC curve. Figure 9 shows the machine learning comparison chart based on the accuracy score, and Fig. 10 represents the different classifier performances based on AUC score.

The research work started to compare the ML classifier's performance on WBC dataset and suggested best classifier to detect breast cancer. The result mentioned in Table 3 indicates that the neural network machine learning classifier achieved the

Table 3 Machine learning classifiers evaluation parameters scores

Model	AUC	Accuracy	F_1 Score	Precision	Recall	Specificity
KNN	0.9904	0.9845	0.9845	0.9846	0.9845	0.9846
SVM	0.986	0.9793	0.9794	0.9805	0.9793	0.9692
RF	0.9902	0.9689	0.9689	0.9689	0.9689	0.9769
NN	0.9934	0.9896	0.9897	0.99	0.9896	0.9846
NB	0.9926	0.9845	0.9845	0.9852	0.9845	0.9769
LR	0.9943	0.9741	0.974	0.974	0.9741	0.9846

Table 4 Machine learning classifiers comparison using area under ROC curve

Model	KNN	SVM	RF	NN	NB	LR
KNN	–	0.825	0.440	0.175	0.217	0.165
SVM	0.175	–	0.279	0.175	0.191	0.167
RF	0.560	0.721	–	0.190	0.171	0.203
NN	0.825	0.825	0.810	–	0.746	0.254
NB	0.783	0.809	0.829	0.254	–	0.254
LR	0.835	0.833	0.729	0.746	0.746	–
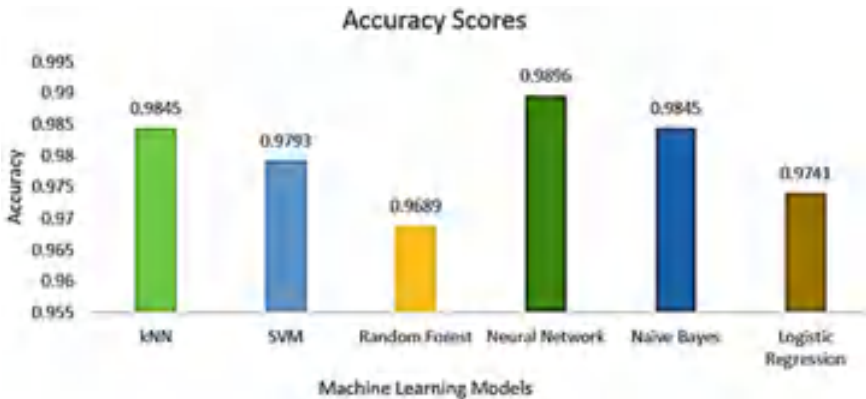
Fig. 9 Machine learning models accuracy score comparison chart

highest accuracy score of 0.99; it can diagnose benign and malignant cases more accurately. The precision and recall scores show that the neural network classifier performed well in all respects; it scores minimum false-positive and false-negative cases. When comparing the classifiers based on AUC score mentioned in Table 4, the logistic regression got the top position; KNN, NN, and LR scored the maximum score of 0.984 in the specificity parameter. Kamel et al. [7] presented a research work to improve SVM accuracy by using by feature selection Gray Wolf algorithm and reported an accuracy of 100%; the said model feels overfitted; due to this, not
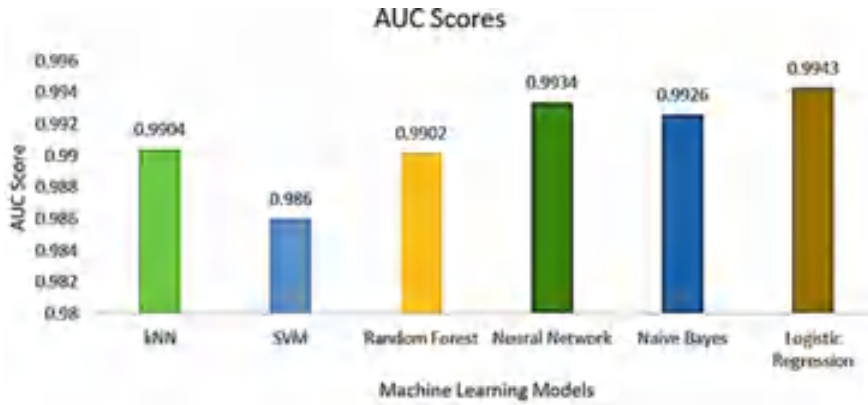
Fig. 10 Machine learning models AUC score comparison chart

Table 5 Research work outcomes comparison with relevant research articles

Author	Machine learning algorithm	Scored accuracy (%)	Research findings
Dhahri and Al Maghayreh [8]	Genetic programming with machine learning algorithms	98.24	The artificial neural network classifier achieved an accuracy score of 0.989
Bohacik [10]	Naive Bayes classifier, neural network	95.99	
Osman and Aljahdali [11]	Radial based function neural network models (RBFNN) with ensemble boosting	97.4	

included in Table 5 for comparison. All the classifiers were configured with parameters mentioned in Table. The stratified sampling technique maintained the class balance in this way helped to reduce fp and fn cases, thus improved the accuracy of classification algorithms. Entioned the other authors' researchon the WBC dataset, and the proposed comparative study results suggested that the NN is the best algorithm to predict breast cancer. However, the research effort's limitation is that it only applies to predicting features mentioned in the breast cancer Wisconsin (original) dataset to detect breast cancer.

6   Conclusion

The leading cause of mortality for women globally is breast cancer. The research aims to study the various machine learning classifier performance to forecast breast cancer in the breast cancer Wisconsin (original) dataset. In the study, six supervised machine

learning classifiers were utilized. As a result, the artificial neural network classifier attained the highest accuracy score of 0.989, and logistic regression achieved the highest AUC score of 0.9943. In future, probabilistic feature selection methods will apply to enhance the classifier prediction accuracy.

References

1. Cancer (2022) https://www.who.int/news-room/fact-sheets/detail/cancer
2. Inbreast, Biōkeanós (2022) https://biokeanos.com/source/INBreast. Accessed
3. Houfani D (2020) Breast cancer classification using machine learning techniques: a comparative study. Med Technol J 4(2):535–544
4. Matloob F (2021) Software defect prediction using supervised machine learning techniques: a systematic literature review. Intell Autom Soft Comput 29(2):403–421
5. Kaklamanis MM, Filippakis ME (2019) A comparative survey of machine learning classification algorithms for breast cancer detection. In: Proceedings of the 23rd pan-hellenic conference on informatics, pp 97–103
6. Rodrigues L (2015) Analysis of the wisconsin breast cancer dataset and machine learning for breast cancer detection. In: and others (ed.) XI workshop de Visão computacional. p 6
7. Kamel SR, Yaghoubzadeh R, Kheirabadi M (2019) Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. J Big Data 6(1):90
8. Dhahri H, Maghayreh E, Mahmood A, Elkilani W, Nagi M (2019) Automated breast cancer diagnosis based on machine learning algorithms. J Healthc Eng 2019:1–11
9. Salama GI, Abdelhalim MB, Zeid MA (2012) Breast cancer diagnosis on three different datasets using multi-classifiers. Int J Comput Inf Technol 01(1):8
10. Bohacik J (2017) Data mining applied to breast cancer wisconsin data. J Inf Technol 10(1):7
11. Osman AH, Aljahdali HMA (2020) An Effective of Ensemble Boosting Learning Method for Breast Cancer Virtual Screening Using Neural Network Model. IEEE Access 8:39165–39174
12. UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set (2022). https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+diagnostic
13. Mangasarian OL, Street WN, Wolberg WH (1995) Breast cancer diagnosis and prognosis via linear programming. Oper Res 43(4):570–577
14. Cunningham P, Delany SJ (2021) K-nearest neighbour classifiers–a tutorial. ACM Comput Surv. 54(6). https://doi.org/10.1145/3459665
15. Pisner DA, Schnyer DM (2020) Chapter 6—Support vector machine. Academic Press. https://doi.org/10.1016/B978-0-12-815739-8.00006-7
16. Genuer R, Poggi JM (2020) Random forests. In: Random Forests with R. Springer International Publishing, pp 33–55
17. Berrar D (2019) Bayes' theorem and Naive Bayes classifier. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C (eds) Encyclopedia of bioinformatics and computational biology. Academic Press, pp 403–412
18. Buscema PM, Massini G, Breda M, Lodwick WA, Newman F, Asadi-Zeydabadi M (2018) Artificial neural networks. In: Artificial adaptive systems using auto contractive maps: theory, applications and extensions. Springer International Publishing, pp 11–35
19. Schober P, Vetter TR (2021) Logistic regression in medical research. Anesth Analg 132(2)
20. Montolío A, Cegoñino J, Garcia-Martin E, Palomar APD (2022) Comparison of machine learning methods using spectralis OCT for diagnosis and disability progression prognosis in multiple sclerosis. Ann Biomed Eng 50(5):507–528

Comparison of Pollard's rho Algorithm Based on Cycle Finding Methods

P. Deepthi and Nagaratna P. Hegde

Abstract The cryptosystems which require small key size to implement a public key cryptosystem, and which are more efficient and secure are elliptic curve cryptosystems. Many public key schemes such as Diffie–Hellman and El Gamal schemes solve the elliptic curve discrete logarithm problem (ECDLP). The security of the Elliptic Curve Cryptographic System depends on the difficulty of the elliptic curve discrete logarithm problem (ECDLP). The major attention of any public key systems is the problem to solve ECDLP. Best way to solve ECDLP is to have an exponential time complexity of within the underlying field size. This study shows comparison of Pollard's rho algorithm based on cycle finding methods and parallelization.

Keywords Elliptic curves · ECDLP · Pollard's rho · Floyd · Random walks · Kangaroo method

1 Introduction

There are several attacks for solving discrete logarithms like Baby Step Giant Step, Pollard rho, etc. One of the most well-known algorithms for solving discrete logarithms is Pollard's rho [1]. Elliptic curves are very important in the field of cryptography. The main use of ECC is in small devices like smart cards as it provides high level of security. In reasonable amount of time, ECC will solve ECDLP.

P. Deepthi (✉)
Computer Science and Engineering, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

N. P. Hegde
Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, Telangana, India

2 Literature Review

Many attacks were made for checking the strongness of the cryptographic techniques [2]. Major work was done on finding the solutions for integer factorization problem, discrete logarithmic problem, group points, etc.

Modular arithmetic is an integer arithmetic to perform reduced modulo M operation [3]. The number of variables in a set is limited in modular arithmetic which is basic of ECC.

The fields of number systems are rational, real and complex numbers. All of them must satisfy the properties of the group which are abelian in nature. The basic addition and multiplication operations are in the set F.

Elliptic curve (E) is satisfied by the following equation which is defined by a finite field (Mq):

$$E : y^2 a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6 3 \tag{1}$$

where the coefficients of the equation are taken from the finite field. The underlying finite field for the general curve equation was given by

$$y^2 = x^3 + ax + b \tag{2}$$

where $a, b \in Mq$ and $4a^3 + 27b^2 \neq 0$, with point O (which is infinite), serving as the specific point for finite group.

The basics of providing security is on the rigidness of discrete logarithmic problem (DLP [4]). All the cryptographic systems are built on security of the discrete logarithmic problem.

Cycle finding algorithms: "If G is set which is finite of some order k and if f: $M \rightarrow M$ is a function which behaves arbitrarily that are better to mix with the elements in M. Take an arbitrary element $x0 \in M$, for computing the number of iterations

$$x_i + 1 = f(x_i) \text{ for } i > 0$$

such that a succession $x_0, x_1, x_2 \ldots$ is calculated. Some elements from the set M must be appeared twice in the succession" [5]. The figure formed will be look as a Greek letter(ρ) (Fig. 1).

The sequence consists of a tail which is of length T. M is the length of infinitely repeating cycle. The purpose of cycle detection problem to find M and T.

3 Pollard's rho for ECDLP

This method can be given by two algorithms:

One is the iterating function and second is Floyd's cycle detection algorithm [6].
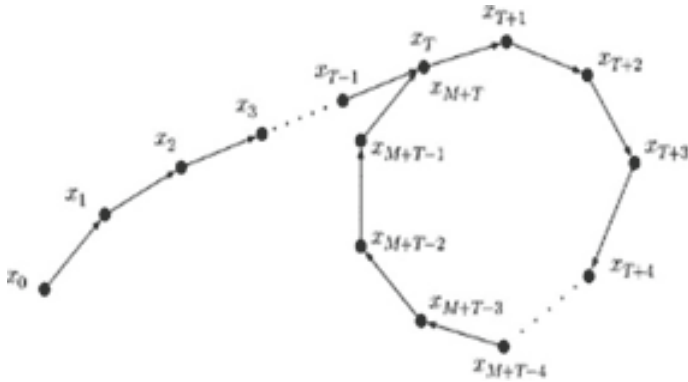
Fig. 1 Rho symbol formed by *Xi* values

Algorithm 1 Iteration function

  function f (R): R$_1$
  if R $\in$ S$_1$ then
  R$_1$ $\leftarrow$ R + P
  else if R $\in$ S$_2$ then
     R$_1$ $\leftarrow$ 2R
  else
     R$_1$ $\leftarrow$ R + Q
  end if
  return R$_1$
  end function
  function f(a, b): a$_1$, b$_1$
  if R $\in$ S$_1$ then
     a$_1$ $\leftarrow$ a + 1
  else if R $\in$ S$_2$ then
     a$_1$ $\leftarrow$ 2a
     b$_1$ $\leftarrow$ 2b
  else
     b$_1$ $\leftarrow$ b + 1
  end if
  return a$_1$, b$_1$
  end function

Algorithm 2 Pollard's rho (using Floyd's cycle detection)

  Require: P, Q, S$_1$, S$_2$, S$_3$
  Ensure: Integer 1 where Q = 1P
  Consider:

$a_0 \leftarrow$ random $\in]0; n[$

$b_0 \leftarrow$ random $\in]0; n[$

$j \leftarrow 0$

$R_0 \leftarrow a_0 P + b_0 Q$

for all j such than $R_j \neq R_{2j}$ do

$\quad (R_{j+1}, a_{j+1}, b_{j+1}) \leftarrow f(R_j), f(a_j, b_j)$

$\quad (R_{2(j+1)}, a_{2(j+1)}, b_{2(j+1)}) \leftarrow f(f(R_{2j})),$

$\quad f(f(a_{2j}, b_{2j}))$

$j \leftarrow j + 1$

If $R_j = R_{2j}$ and $b_j \neq b_{2j}$ then

$1 = \frac{a_{2j} - a_j}{b_j - b_{2j}}$ mod (n)

else if $b_j = b_{2j}$ then

$\quad a_0 \leftarrow$ random $\in 0; n[$

$\quad b_0 \leftarrow$ random $\in]0; n[$

$\quad j \leftarrow 0$

end if

end for

Return 1

4 Modified Pollard's rho

In order to reduce the time complexity, Pollard's original algorithm has undergone a number of modifications [7]. This paper presents the enhancements to the original Pollard's rho algorithm.

The proposal of modified Pollard's rho implementation is it should use the iterating function for generation and storing the items using stack.

The detailed procedure of the modified Pollard's rho method [8] is given as follows:

Three sets Z_1, Z_2, Z_3 which are disjoint are taken from the elliptic curve $E(F_n)$. Z_1, Z_2, Z_3 are the pointed on the elliptic curve which range between the value of the y-coordinate given by $Z_1 \in (0, (n/3) - 1), Z_2 \in (n/3, (2n/3) - 1), Z_3 \in (2n/3, n - 1)$.

Calculate $P_0 = a_0 M + b_0 N + 1$ for two random integers a_0 and b_0, where $a_0, b_0 \in [1, n - 1]$.

The iteration function g is defined by

If $P_k < P_j$, all the $P_j's$ are popped from the stack else P_k is pushed on to the top of the stack. This process is continued till $P_j = P_k$ is found. If the match is found, the discrete logarithm is solved.

The cycle detecting algorithm used in the study uses stack for storing all the intermediate values. These values are used for detecting the cycle. The stack uses a finite set of sequences which are used for the formation of the cycle.

5 The Modified Pollard's rho Implementation

Input of the algorithm: M, N, Z_1, Z_2, Z_3

where M, N are located on the elliptic curve $E(F_n)$ and the points are divided into the sets Z_1, Z_2, Z_3.

Output of the algorithm: To find integer h such that $N = hM$.

The pseudocode for the algorithm is as follows:

- a_0, b_0 are two random integers, which range between $[1, n - 1]$
- Initialize $k \leftarrow 1, j \leftarrow 0$
- Calculate: $P_0 \leftarrow a_0M + b_0N + 1$
- Find: $(P_k, a_k, b_k) \leftarrow (g(P_0), g(a_0, b_0))$
- Push (j, a_j, b_j, P_j) on to the stack, $\forall k$ such that if $P_k \neq P_j$
- For all k do, if $P_k < P_j$ then pop the stack else if $P_k \geq P_j$ then push (j, a_j, b_j, P_j) on to the top of the stack, end if
- If $P_k > P_j$ then find $(P_{k+1}, a_{k+1}, b_{k+1}) \leftarrow (g(P_k), g(a_k, b_k))$
- Else $h = \frac{a_k - a_j}{b_j - b_k}$ mod (n), end if
- $k \leftarrow k + 1$, and return h

6 Result Analysis

In computer programming, the algorithms implemented must be analysed [9], to know the best methods.

Time taken for the enhanced Pollard's rho method.

The modified Pollard's rho method provides better performance (Table 1). As the modified algorithm uses stack for the iterating function, the time taken to detect the cycle is reduced. The time taken for the modified Pollard's rho algorithm for cycle detection is reduced compared to the original Pollard rho algorithm.

Iterations required for modified Pollard's rho.

Table 1 Comparing of run time for modified Pollard's and original

No. of digits	Pollard's rho with Floyd's cycle (original) time (ms)	Pollard's rho with stack (modified) time (ms)
100	0.143455	0.078954
110	0.456321	0.231233
115	1.234563	1.013451
120	4.543781	3.903451
125	6.333490	4.653567
130	8.908321	7.454322

Table 2 Number of iterations in modified Pollard's rho and original Pollard's rho	No. of digits	Pollard's rho with Floyd's cycle (original)	Pollard's rho with stack (modified)
		Number of iterations	Number of iterations
	100	1020	489
	110	3456	2134
	115	123,124	54,367
	120	378,579	23,410
	125	543,671	412,389
	130	1,321,456	567,834

The iterations of the modified Pollard's rho are less when compared to original Pollard's rho algorithm (Table 2).

7 Random Walks

In this paper, Pollard's rho method uses an iterative function of a random walk to form a series of random terms [10]. The idea behind the algorithm's iterative functions is that the generated terms are random, like a random walk.

7.1 Random Walks Using Stack

Random walks, which were developed at the cost of a little extra work, were used in the initial iteration function [11]. As a result, in the Pollard's rho method, this paper employs both Stack and Random Walks.

The following are the steps in Pollard's rho mixing random walks using stack algorithm:

- $E(Fp)$ is divided into 30 sets by the sets $Z_1, Z_2, \ldots Z_{30}$.
- $Z_1, Z_2, \ldots, Z_{30}$ are places where the x-coordinate mod 30 is 0, 1, 2, … 29.
- Calculate the points $Mj = ajP + bjQ + 1$ by generating 30 random pairs (aj, bj).
- Define the iterative function g (as in Eq. 3) as $R_0 = a_0P + b_0Q$, where a_0 and b_0 are two integers which are random in the range [1, n − 1].

$$g(R_j) = R_{j+1} = \{M_k + R_k + 1 \text{ if } R_j + S_k\} \tag{3}$$

The *aj* and *bj* sequence are computed as (in Eq. 4) follows:

$$(a_j, b_j) = \{(a + a_{k+1}, b + b_{k+1})\} \text{if } R_j + S_k \tag{4}$$

If $Rj + 1$ is less than Rj pop the item from the stack where $Rj + 1$ is the least else push $Rj + 1$ on the top and continue the iteration.

- When $Rj = Rk$, stop the process and identify the fixed point; the discrete logarithm is resolved as (in Eq. 5)

$$1 = \frac{a_k - a_j}{b_i - b_k} \mathrm{mod(n)} \tag{5}$$

7.2 Performance Analysis of Mixing Stack and Random Walks with Runtime

With regard to runtime, the Pollard's rho mixing stack and random walks approach outperforms the original Pollard's rho method (Table 3).

8 Pollard Kangaroo Method

Pollard's kangaroo technique can use the stack-based cycle detection approach [12]. Compute the discrete logarithm using the kangaroo method in the cyclic group, given that the value must fall inside a specific interval, $N = kM$.

The lambda, or the kangaroo's method, on a single processor works as

- The kangaroo method involves two kangaroos, a tame and a wild one, jumping along a number line marked M, $2\,M$, $3\,M$, … until the wild kangaroo gets the tame one. $Z_{i+1} = Z_i + a(Z_i)M + 1$, where $a(Z_i)$ is a function that outputs the kangaroo's next jump size and the kangaroo advances by this value, is an iterating function that determines the kangaroo's next location.
- The $a(Z_i)$ function only takes into account the current position Z_i when randomly choosing values from a set A. The power of two values in the set A range from

Table 3 Original rho and mixing stack and random walks run time comparison

No. of digits	Pollard rho with Floyd's cycle (original) (time in ms)	Pollard rho with mixing stack and random walks (time in ms)
100	0.143455	0.023123
110	0.456321	0.342356
115	1.234563	0.760982
120	4.543781	1.234896
125	6.333490	2.345890
130	8.908321	4.342123

2^0 to a maximum value, with the largest value being such that the set's mean is an optimal value. A tame kangaroo begins the collision detection method at $Z0$ = bM and hops for an optimum value while considering the distance travelled, to find the collision of the kangaroo [13]. Then, a wild kangaroo starts at $Z0'$ = N and is permitted to leap using the same method for iteration, while considering its distance travelled. The wild kangaroo tries to land on the same position of the tame kangaroo at the end. By subtracting the distance travelled by the two kangaroos, the discrete logarithm can be determined. If the wild kangaroo travels a long distance, it has failed to reach the tame kangaroo's final point. If the wild kangaroo fails, a new wild kangaroo is launched from $Z0'$ = $N + tQ$, with a small known value of t.

In this study, the number of kangaroos in the algorithm is increased by using two kangaroos, one tame and one wild, and to keep the tame kangaroo's last four steps rather than just the last step. Then compare the w_i's to the previous stage to determine if there is a match. From the fourth to the last phase, wi's are compared. This entails comparing the w_i's to the last four phases. This algorithm has been changed to include the steps of storing X_{y-3}, X_{y-2}, X_{y-1}, and X_y. Then, whenever w_i is discovered, compare it to the four values X_{y-3}, X_{y-2}, X_{y-1}, and X_y. Collision is detected if a match is identified. Repeat the process for the following w_i if the collision is not detected.

8.1 Pollard's Kangaroo Method Performance Analysis

When the run time is compared to original Pollard's rho, the Pollard's kangaroo method takes less time. Because the kangaroo approach is the parallelized variant, it takes less time to implement (Table 4).

Table 4 Time taken for Pollard's rho and Kangaroo method

No. of digits	Pollard rho with Floyd's cycle (original) (time in ms)	Pollard's Kangaroo method (time in ms)
100	0.143455	0.053234
110	0.456321	0.210926
115	1.234563	0.980321
120	4.543781	2.342158
125	6.333490	3.563267
130	8.908321	6.567375

9 Implementation

The discrete logarithm problem is the most difficult mathematics problem in the case of elliptic curves. Till now there are no known rampant attacks on ECDLP, the study approaches are the most efficient known attacks for obtaining discrete logarithms.

9.1 GNU Multi-precision (GNU MP)

The algorithms in the paper were implemented in C, using GNU multi-precision (GNU MP) libraries. GNU MP library is portable which is used to carry out computations on large numbers (like arbitrary precision integers and rational numbers). It provides the fastest possible arithmetic for all the applications like cryptography.

9.2 Evaluation of Results

In computer programming, analysing algorithms is crucial since there are frequently numerous methods which are already present for any application that is known.

The following is an analysis of the results of the paper:

1. It was discovered that modified Pollard's rho time taken is very little compared to original Pollard's rho approach.
2. When compared to the iteration function of modified Pollard's rho with the original Pollard's rho approach, the iterations are less in number.
3. By combining stack and random walks, modified Pollard's rho modified technique outperforms the original rho method in terms of run time.
4. When compared to Pollard's rho, the Kangaroo technique takes less time to run (due to parallelization) than original rho method.
5. The algorithms' performance is assessed.

9.3 Performance Analysis of the Algorithms Implemented

Various methods were examined in this study. In general, each of them has a distinct advantage over the others in certain areas. To put it another way, there is no universal algorithm that works well in all circumstances. Everything is dependent on the environment in which it is used.

The Pollard rho with mixing stack and random walks performs better than the other algorithms since it takes less time (Table 5).

Table 5 Performance analysis of the algorithms

No. of digits	Pollard rho with Floyd's cycle (original) (time in ms)	Pollard rho with stack (modified) (time in ms)	Pollard rho with mixing stack and random walks (time in ms)	Pollard's Kangaroo method (time in ms)
100	0.143455	0.078954	0.023123	0.053234
110	0.456321	0.231233	0.342356	0.210926
115	1.234563	1.013451	0.760982	0.980321
120	4.543781	3.903451	1.234896	2.342158
125	6.333490	4.653567	2.345890	3.563267
130	8.908321	7.454322	4.342123	6.567375

10 Conclusion

The results of the research will be of use when studying Pollard's rho algorithm. Furthermore, since the security of the elliptic curve cryptography lies in the difficulty to solve the elliptic curve discrete logarithm problem the results will be of use when studying methods to break the elliptic curve cryptosystems.

References

1. Raya A, Mariyappn K (2020) Security and performance of elliptic curve cryptography in resource-limited environments: a comparative study. In: 15th International conference for internet technology and secured transactions (ICITST)
2. Abdaoui A, Erbady A, Al-Ali A, Mohamed A, Guizani M (2021) Fuzzy elliptic curve cryptography for authentication in Internet of Things. IEEE Internet of Things J (Early Access) 1
3. Meneghetti A, Sala M, Taufer D (2020) A new ECDLP-Based PoW model. Mathematics 8(8)
4. Verri Lucca A, Mariano Sborz GA, Leithardt VRQ, Beko M, Albenes Zeferino C, Parreira WD (2020) A review of techniques for implementing elliptic curve point multiplication on hardware. J Sens Actuator Netw 10
5. Sengupta A, Ray UK (2016) Message mapping and reverse mapping in elliptic curve cryptosystem. Secur Commun Netw 9:5363–5375
6. Stein A, Teske E (2002) The parallelized Pollard kangaroo method in real quadratic function fields. Math Comp 71:793–814
7. Neamah AA, Najaf (2015) New collisions to improve pollard's rho method of solving the discrete logarithm problem on elliptic curves. J Comput Sci 11(9):971–975
8. Qi B, Ma J, Lv K (2020) Improved algorithm for solving discrete logarithm problem by expanding factor. China Commun 17(4):31–41
9. Duta CL, Gheorghe L, Tapus N (2016) Framework for evaluation and comparison of integer factorization algorithms. In: SAI computing conference (SAI)
10. Sah CP, Jha K, Nepal S (2016) Zero-Knowledge proofs technique using integer factorization for analyzing robustness in cryptography. In: 2016 3rd International conference on computing for sustainable global development (INDIACom)
11. Lara-Nino CA, Diaz-Perez A, Morales-Sandoval M (2018) Elliptic curve lightweight cryptography: a survey. IEEE Access 6:72514–72550

12. Nickerson D (2020) Collision detection and Pollard's rho algorithm for the discrete logarithm problem
13. Mahto D, Yadav DK (2017) RSA and ECC: a comparative analysis. Int J Appl Eng Res 12(19):9053–9061. ISSN 0973-4562

Smart Street Parking System for Smart Cities Based on the IoT Prototype

Parthasarathi Pattnayak and Sanghamitra Patnaik

Abstract The Internet of Things (IoT), which provides selected sub-categories of data with free access to a wealth of digital services, is able to seamlessly incorporate a wide range of heterogeneous end systems. Smart City is a vision aimed at integrating people residing in the cities with services that are essential and affect everyday life. Smart parking within all streets with multiple information and communications solutions is one such example of essential service. The Internet of Things (IoT) is a new and unique step forward to manage efficiently and effectively the parking system through its ability of smart compute intelligent parking systems. The most important reason for using IoT for parking is to collect the data on vehicle occupancy and use free parking spaces effectively. This paper proposes a prototype IoT-based real-time smart street parking system, with easy accessibility to appropriate data and provide solution to the people to locate free parking spot easily and effectively.

Keywords Smart cities · Smart parking · IoT · Waiting time

1 Introduction

The concept of Smart cities is based on the model of urban development. It focuses on how to utilize the human and technological capital in an optimum way for growth and prosperity. It is a conscious effort on the part of smart cities to exploit strategically the Information and Communications Technology (ICT) that would lead to achieve effective growth and competitiveness at multiple socio-economic levels. The increased volume of traffic congestion necessitates for a smart parking systems.

P. Pattnayak (✉)
School of Computer Applications, KIIT Deemed to be University, Bhubaneswar, Odisha 751024, India
e-mail: parthakiit19@gmail.com

S. Patnaik
School of Law, KIIT Deemed to be University, Bhubaneswar, Odisha 751024, India
e-mail: sanghamitra18patnaik@gmail.com

Smart parking is a conglomeration of technology and human innovation with less possible use of resources such as fuel, time and space. It would make the parking easier and faster for maximum number of vehicles. The drivers can avail the facilities by paying marginal fees to the smart parking service providers. The use of 'smart parking' is a technique that revolves round Internet of Things (here after IoT) to deal with the challenges/crisis. The IoT is a recent paradigm of communication in the field of microcontrollers, digital transceivers and appropriate protocol stakes for communicating between users, making it an integral part of the Internet [1]. Consequently, the IoT concept aims to create an all-in-one Internet. Moreover, by easily accessing and interacting with a wide range of device such as home devices, surveillance cameras, sensors, actuators, displays and automobiles, IOT supports the development of several applications using the potentially huge amount and range of data that is generated by such objects. In fact, this paradigm applies in a variety of areas such as home automation, industrial automation, medical aid, Mobile Health, Senior Aid, Smart Energy and Smart Strengths, automotive, traffic and much more [2]. This work centres basically around the clever IoT leaving because of vehicle clog in the city. The primary justification utilizing IoT for stopping is to effortlessly gather information with the expectation of complimentary stopping places. In IoT objects, the sensors, actuators, blue instruments, radio repeat IDs, etc., will be communicated. This work can upgrade the adequacy of this arrangement of leaving, through IoT, where items can impart without human obstruction. This is unreasonable, automated road parking. Nowadays, numerous vehicles are guided by the GPS, and sensors can be handily placed in leaving spaces to screen the opening information by means of a robot and IoT to deliver a road database. Every hub is located in the heart of a car park and card and is located near the roadside sensor hubs. When a hub recognizes the vehicle's entry, it communicates with the customer by switches. You move the packet to the main server and collect information from hubs on the main server and show the parking direction on the display board.

2   Related Work

A webcam-based parking system that makes use of several sensors and IRIS infrastructure components. Compo-sensors collect information during component organization and provide offices with checking verified sensor information afterwards [6]. At the beginning, of the parking lot showcases ultrasonic sensors to identify and recognize vehicles [7]. Vehicle identification is based on its characteristics, such as shape and design. These qualities could be data based for each vehicle and their examples coordinated by the calculation of identification with a capacity to distinguish and recognize special kinds of vehicles. The parking algorithm [8] provides a sustainable vehicle parking strategy. There is no costly calculation needed, and the calculation has been shown to stop at any underlying positions and to correctly calculate parking space that can control the vehicle in a confined space. Considering WSN imaginative measures to give impelled features, for example, distant stopping

observing, PC estimated course and stopping booking tools [9]. The Parking Guidance Information System (PGIS) developed [10] has three sorts of centre points like screen intersections, centre points for guiding and intersection junction. These centre points converse with the distant channel and orchestrate themselves into a framework uncommonly planned. Display Park Sense shows if a resident has removed the car from the car park [11]. The system uses an energetic Wi-Fi signature to coordinate the approach of the driver of the stopped vehicle. A comprehensive system engineering made up of several devices [12]. The leaving heading and information framework utilized picture preparing strategies to just distinguish vehicles at a particular space as indicated by Reddy and others [13]. In the event that nothing else exists than a vehicle, it isn't booked. They utilized a microcontroller ARM9, a webcam and the GSM to screen vehicle leaves and a LCD contact screen to show the results. As recently referenced, the webcam can generate a large amount of data that is difficult to transport across the WSN. In spite of the fact that it extensively limits end client determination obstructions in the light of a legitimate concern for key security and assurance issues, acknowledgment that depends on the unwavering quality application of an IoT generally relies on the implementation of optimal security and security practices in the immediately visible IoT gadget [14]. Parking is becoming a scarce resource in any major city on the planet [15], and its limited availability addresses a concomitant cause of city gridlock and air pollution. Notwithstanding the way that an ideal chance for discovering free parking spaces could be decreased and furthermore vehicles action enhanced [16], showed an outdoors model of leaving, which utilizes WSN as the possible reaction for the issue of leaving. People-group-based approaches were used to improve the smart parking system's effectiveness (SPS) and in this manner to decrease the blockage of movement in metropolitan circumstances while expanding stopping constructions' capability and profitability [17]. Drivers filtering for open stopping roads address a gigantic measure of development in metropolitan regions. At the point when the vehicle enters the leaving region, the capacity interaction begins and the utilization of the lifting vehicles stops without blocking different vehicles [18]. IoT perspective requirements cloud-based sharp vehicle leaving experts in brilliant metropolitan areas [19]. This sort of organizations will, because of its unadulterated business parts, become a major segment in a vague IoT operational stage for metropolitan areas. As this work and related arrangements are not IoT-based, it intends to give a shrewd IoT-based arrangement that permits individuals to know ahead of the accessibility of free parking spot. This arrangement is a cutting edge arrangement. The IoT tends to the purchaser's equipment and the B2B endeavour. The appraisals on IoT are assessed at between $300 B and almost $2000 B in 2020, depending upon scientists (dependent upon what is addressed in IoT publicize). Gartner's attention on IoT included regard has been trademark. Each year since 2020, 10 trillion IoT protest shipments are normal and used as part of a wide range of uses. By a wide margin associated LED lights are the greatest. Gartner states that sensors play a fundamental role in IoT technology in terms of innovations. IoT design remains questionable; however, for all segments of the business sector there will not be a single engineering. The increased revenue estimated for the year 2020 will be received from governments by more

than 80% of the \$300B. Administration openings related to the organization and the management of the different things with which common buyers communicate, both in the home and during their daily businesses, will take place. Certainly, the esteem chain will be multi-layered for IoT gadgets and administrations, using information that uses examinational programming to draw up patterns useful for additional items and advertising. To change enormous volumes of information into some (before long coordinated by a human character) amount, IoT makes 'extreme information', which assessment should convey usable/eminent data [20].

3 Vision-Based Systems for Parking

The conventional vehicle leave framework can't give genuine guide to road occupants and drivers. The clog of leaving in urban communities is a significant issue, especially in non-industrial nations; many model leaving frameworks have been proposed by different researchers to address this. Various approaches to make the leaving framework cleverer, more solid and robust models were proposed [4]. Intelligent parking systems are technologically enabled environments with protection and proactive functions to help an inhabitant to manage his or her daily lives that are targeted at him or her. Sensors and actuators for detecting change of status and initiating beneficial interventions should be used as a standard smart parking implementation [5]. Two arrangements have vision-based leaving system. The first is to search for the measure of void regions for vehicle parking. The second is to check the parking spots of the private property.

4 Smart Street Parking System with IoT-Based Real Time

Webcams that regulate the car park have been the foundation of smart and intelligent parking systems so far. To check for space occupancy, the framework employs image preparation techniques. Tang et al. [21] offer a savvy street leaving framework situated in WSN, in which distant sensors are put in a parking spot vehicle with each parking spot fitted with a sensor centre that perceives and controls a parking spot. Smart street parking system with IoT-based real time is spurred by the utilization of an IoT smart road stopping framework, in the light of the fact that IoT can gather data with the expectation of complimentary parking spots by drone. In this undertaking, drone pictures acquired from a UAV framework, the most recent innovation for the Geo-spatial information obtaining, are utilized to plan free stopping highlights. This altogether lessens the expense and season of the venture. The robot video yield is put away in a data set and can be seen related to the Google 2D guide of the caught regions. To change these Rasters over to vector maps you utilize created outlines from the caught video that contain XY coordinates for each edge. In both Arc GIS worker and guide motor, vector maps created are put away. An extraordinary ID is put

away for each free parking area. For the last 15 years, the public authority has battled hard to diminish the unmanaged street leaving problem. The populace of vehicles is eating up alarmingly quick in their road. The staggered idea of stopping on the roads had fizzled. The proposed IoT-based SSPS model is partitioned into three modules, every one of which has four layers as per the IoT design. An IoT-based RTSSPS model was intended to identify the vehicle by coordinating with token numbers by utilizing IR sensors at each leaving space. In the event that a vehicle is before the street entrance, the vehicle administrator creates a special id token and through the primary design module, this one of a kind id passes to the leaving sensor worker and takes the data from the cloud module through the server farm module. Arduino UNO, the microcontroller, will get the sensor readings and break down them. The opening is viewed as involved when the sensor is perused from base too high for in excess of 5 s, yet is low than space empty. The outcome will be sent to the application that shows the vehicle leave. On the 16x2 LCD screen of the stopping region section, if an empty parking area can be found. The outcomes are presented. With the guide of token number, an engine constrained by L283D IC opens the entryway by which the vehicle enters the vehicle leave.

5 A Real-Time Smart Street Parking System with an IoT-Based Architecture

Smart street parking system (here after SSPS) design dependent on IoT gives the best way to deal with the drawbacks of the other old style and clever vehicle parks. It basically comprises sensors to distinguish the vehicle's essence out and about vehicle park. The SSPS dependent on IoT has two key shrewd engineering for road leaving and a keen calculation for road parking. Sensor junctions are relocated to the side of the road, and each crossroads is constituted in the midst of a parking lot and on a car parking lot. The world's alluring field is perceived by each sensor centre point, Wi-Fi and irregularly by sunlight and street light. The SMS is moved to the cell phone of the client when a centre point perceives a vehicle by changing to fundamental station. This primary station assembles the data from various centres. It additionally sends the distant worker to the presentation board, which can be mounted at the hour of beginning and leaving the roads. This examination attempts to utilize estimation on road stopping through inherited attributes or an unquenchable way to deal with improve on the examination on sensors to decrease the quantity of sensors and give them full coverage. The SSPS design dependent on IoT is parted into three segments: Smart brilliant road station module dependent on IoT, Smart Street shrewd server farm module and Cloud smart road stopping module dependent on IoT (Fig. 1).

Fig. 1 Real-time smart street parking system with an IoT-based architecture

6 IoT-Based WSN Centric Smart Street Parking Module

In the light of the Internet of Things, WSN centric smart street parking module is a combination of usage layer, board administration layer, organization and passage layer, availability sensors and organization layer for IoT design. The IoT-driven WSN assists with setting up sensor centre points easily. The vehicle leaves are shown through objects to customers. The application layer of this engineering is intended to sensor and associate vehicles, when the driver needs to leave in the vehicle leave and to gather data about free parking spots and ownership. Through the use of big data analysis, the management service layer offers QoS, data analytics, security control, system modelling, and control of devices (Fig. 2). Countless IoT pre-treatment and understanding volumes delivered by remote sensors and brilliant gadgets through distributed computing and a passage framework ought to have the option to get to doors, network layer and sensor availability layer. This layer is gathered by its points and the sort of information, for example, ecological, attractive, block and observation sensors and so on by shrewd gadgets and sensors. This stack produces WSN centric IoT preparing and gives the data to a focused on the spot for additional handling. To get an outcome or produce some yield from the keen IoT street stopping engineering, ongoing data is gathered and processed. In the light of the way that it would lessen the ideal chance to track down a free parking spot and furthermore decrease tainting and improve vehicular movement, it is a considered response to the topic of pausing and upgrade individual fulfilment of brilliant metropolitan networks.

Fig. 2 WSN centric smart street parking system module based on IoT
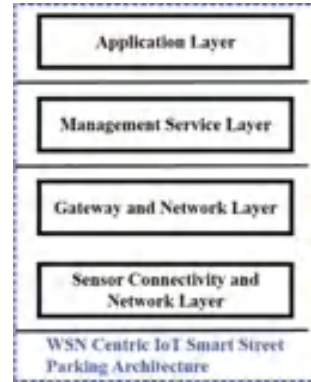
7 Smart Street Parking Module IoT-Based Data Centre

In the past consecutive years, the volume of data transferred via many digital devices and traffic reduction has seen an unrivalled explosion in this study, which requires strong communication techniques in order to transmit a high level of data and information. The world's most significant issue today is the huge amount of important Big-Data streaming between various organizations and trade data with more noteworthy pressure and security in existence for information move for information stock [22]. An enormous measure of data will be gathered and produced by IoT, which centres around information assortment, stockpiling and perception. To effectively misuse spatial time qualities of information arranged by drone, a summed up framework to gather information is required. It isn't trifling to remove full importance from crude information. In keen urban communities, flexibility and calculation power are typically important to analyse information for an enormous scope, in the light of the fact that outwardly screening is significant for client portrayal in smart cities. Information collected via application and WSN drives IoT to the database system, pre-processed and transmitted to the client via high layer AMOLED, LCD and LED applications (Fig. 3).

The following modules can be shared the WSN and cloud-centric databases in the data-centred IoT architecture:

- Data module and management of parking for free slot management.
- Using cloud computing, a security management module may detect those who are residing in the country illegally.
- Management sensor module for managing location and searching for optimization via the updated database.
- Server administration module, which includes managing servers that are linked to a database and reporting on the situation via SMS.

Fig. 3 Smart street parking
module based on IoT data
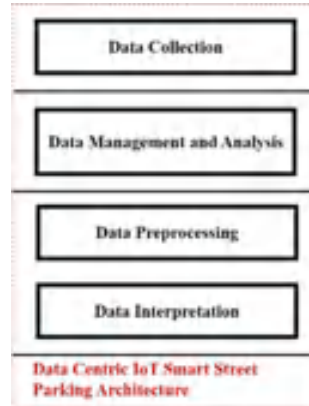
8 Smart Street Parking Module IoT-Based Cloud Centre

IoT-driven, insightful road stopping module is the focal point of the whole engi-
neering and uses investigation apparatuses to convey between IoT WSN focus and
IoT information centre.

Cloud computing gives programming to gather and deal with the framework,
administration and by and large conduct in the changing actual climate through the
proposed design (Fig. 4). The distributed computing climate is similarly helpful to
look and show the way to the stopping region on the LED show at the section, in the
light of the accessible space, with the proposed algorithm. The drivers may enter the
parking garages and have less stopping time.

Fig. 4 Module for an
IoT-based cloud-centric
smart street parking system

9 Smart Street Parking System Algorithms IoT-Based Real Time

Numerous regions have accepted a stopping and information framework to attempt to decrease the hole. This project creates a road data collection using a robot in our suggested framework and assigns an ID to each resident's parking place. When a driver or occupant enters the road for the vehicle park, this framework generates a number that combines the id number of the parking spot, the last two digits of the vehicle enrolment number and the last two digits of the occupant house number. As soon as the client holds the parking spot, the booking authority makes an impression on the client. At the vehicle leave, the client is related to a code checked by the authority, utilizing the leaving ticket or IoT. Notwithstanding, assuming the vehicle doesn't possess a vehicle, it has been approved to give consent by means of SMS to the proprietor of the vehicle and the position to produce the interesting distinguishing proof number. This computation has been provided to assess the problem of street leaving for the outside and the leaving proprietor in the case of this vehicle and vehicle event. Memory and figure power are limited in sensor centres.

10 Street Parking Algorithm Implementation

SSPS (PL, N), SSPS $\rightarrow$ intelligent street parking system.
PL $\rightarrow$ linear N slot array.
Let $S \rightarrow$ entrance to the Street Node $\rightarrow$ C-nodes, CP currently position of C_i and C_j position the succeeding node, D_{min} and D_{max} are the minimum and the maximum distance, PS is the parking area, and the DB is the streets database.
Entry $\rightarrow$ $S =$ CP, C, Db, PS, D_{min}, D_{max}, C_i, C_j, C_{next}
Leaving $\rightarrow$ L293D Signal

1. Set-up: CL Current Car Placement, when residents or drivers access the parking road and hunt for an empty place, the authority generates the only id for them.
2. If car park slots (PS) $= 0$ and car does not allow the road, L293D $= 0$.
    Else if L293D $= 1$ or more automobiles allows you to enter the road after receiving a unique identity if car parking (PS) $= 1$ or more than 1.
3. For the IoT/PARKING slot, there are two neighbouring tables for all C-nodes of C_i do D_{min} and D_{max}. This is the connection's weight (w) and distance (d) between C_i and C_j nodes.

    If the slot owner's w is smaller than D_{min} then D_{min} ¼w,
    Update $\rightarrow$ C_{next} ¼C_j
    Exit
    if Else $< D_{max}$ then for owner slot parking D_{max} ¼w Update $\rightarrow$ C_{next} ¼C_j

end for return C_{next}

4. Event for the vehicle

- The sensor shall detect and report the same to the microcontroller when the vehicle arrives at the gate.
- The space available in the parking lot is inspected by the microcontroller.
- This transmits the results to the displayed web application.
- A parking ticket, with a single id, opens the gate and enters the car, if space is provided.
- When a person parks the car according to the ticket number, the sensor detects their presence and communicates it to the microcontroller.
- The database is updated by the microcontroller.
- The door does not open if there is no parking space.

5. Event for car out

- When the customer exits the spot, the sensor will detect the free opening and notify the microcontroller.
- The web application is equally communicated by the microcontroller.
- The database has been updated.
- The position of the slot is now free of charge.
- The outcome is presented on the online application, and the user enters the gate.
- When the gate opens for the vehicle, the sensor input matches the ticket number and sends a SMS to the owner, who says yes.

6. End

IoT-based SSPS assessment: An SSPS prototype is developed for testing the work process. The SSPS IoT-based test was conducted on the basis of observational vacuums during the prototype test. When IoT-based SSPS was used, the ideal opportunity to look for a freestanding car park was considerably reduced. According to a telephone survey, the average speed taken by drivers is 20 min. It was found that the drivers could locate free parking spots using a clever token number with IoT-based SSPS, which is reduced to 1–2 min. Therefore, it is not necessary to walk and only time is used for the entry and parking of a car in the car park. The IoT-based SSPS aims at facilitating the identification of the drivers of an empty parking area. This requirement can be met by the web application. The application is authorized to manage the parking and installation in real time. The status and position of the parking space can be seen by customers. Accompanying condition defines the location of the car park opening:

If ((S = High) && (T [5)) then;
P_N is not equal to EMPTY
Else P_N is equal to EMPTY

PN and T is the nth vehicle leave space and time where S is the sensor signal. If the sensor signal distinguishes a presence, it turns out to be High and if for 5 s the spot

stays High, the position remains empty. As the ideal opportunity for discovering a vehicle leave is diminished, there will likewise be a decrease in the gridlock because of the drivers wandering around the inquiry offers. Consequently, the clog thickness of various business sectors utilizing customary leaving frameworks were contrasted and the SSPS dependent on IoT. The thickness of blockage in a road can be expressed the accompanying:

$$\text{Density of congestion} = \frac{(\text{Search} + \text{Tparking})\omega \text{Naverage}_{\text{cars}}}{\text{ECs}}$$

where ECS is a car equivalent area designated for convenient vehicle circulation and easy recovery. The ECS is considered equivalent to 23 m^2.

11 Conclusion

This work surveys the flow field examination and tries to foster an IoT-based SSPS that will be appropriate for agricultural nations and will improve execution by lessening the quantity of clients who don't discover a stopping space. This work depicts the model plan and design of the IoT-explicit SSPS. SSPS IoT-based engineering comprises three sections IoT-based keen street leaving, IoT-based shrewd road leaving module and IoT-based savvy road leaving with road leaving algorithms. The legitimate correspondence of every module, for example, the keen vehicle leaving module based IoT, information-driven brilliant road vehicle leaving module and the cloud-driven keen vehicle leaving module, has effectively embedded our proposed design and calculation in a genuine situation. The sensors ought to be fitted out and associated at the intersection with the regulator and speak with the keen road leaving module WSN-driven IoT. During execution, the calculation figures out that the normal time the client stands by to look through the vehicle leave is fundamentally diminished.

References

1. Ahvenniemi H, Huovila A, Pinto-Seppa I, Airaksinen M (2017) What are the differences between sustainable and smart cities? Cities 60:234–245
2. Pattnayak P, Jena OP (2021) Green IoT based technology for sustainable smart cities. In: International conference on innovative computing and communication (ICICC-2021), New Delhi, India, pp 1–5
3. Bibri SE, Krogstie J (2017) ICT of the new wave of computing for sustainable urban forms: their big data and context–aware augmented typologies and design concepts. Sustain Cities Soc 32:449–474
4. Horch A, Kubach M, Roßnagel H, Laufs U (2017) Why should only your home be smart? A vision for the office of tomorrow. In: IEEE international conference on smart cloud (SmartCloud), New York, NY, pp 52–59

5. Carrillo E, Benitez V, Mendoza C, Pacheco J (2015) IoT framework for smart buildings with cloud computing. In: IEEE first international smart cities conference (ISC2), Guadalajara, pp 1–6
6. Beloglazov A, AbawajyJ, Buyya R (2012) Energy–aware resource allocation heuristics for efficient management of data centers for cloud computing. Future Gener Comp Syst 28(5):755–768
7. Mohanty SP, Choppali U, Kougianos E (2016) Everything you wanted to know about smart cities: the internet of things is the backbone. IEEE Consum Electron Mag 5(3):60–67
8. Maksimovic M, Gavrilovic Z (2016) Connecting sciences in green: internet of things and economy. In: ENTECH'16/IV. International energy technologies conference, pp 173–182
9. Vidas-Bubanja M (2014) Implementation of green ICT for sustainable economic development, pp 1592–1597
10. Yasir M et al (2017) Internet of things based smart cities: recent advances and challenges. IEEE Commun Mag 55(9):16–24
11. Nuaimi HE, Neyadi M, Nader, Jaroodi J (2015) Applications of big data to smart cities. J Internet Serv Appl 6(25):1–15
12. Alam F et al (2017) Data fusion and IoT for smart ubiquitous environments: a survey. IEEE Access 5:9533–9554
13. Batty M et al (2012) Smart cities of the future. Euro Phys J 214:481–518
14. Pirbhulal S et al (2017) A novel secure IoT-based smart home automation system using WSN. Sensors 17(1):1–18
15. Medvedev A et al (2015) Waste management as an IoT-enabled service in smart cities. In: Lecture notes in computer science, vol 9247, pp 104–115
16. Kimbahune V, Deshpande AV, Mahalle PN (2017) Lightweight key management for adaptive addressing in next generation internet. Int J Amb Comput Intell 8(1):20
17. Sawlikar AP, Khan ZJ, Akojwar SG (2016) Efficient energy saving cryptographic techniques with software solution in wireless network. Int J Synthetic Emotions (IJSE) 7(2):19
18. Addo I, Ahamed SI, Yau SS, Buduru A (2014) A reference architecture for improving security and privacy in internet of things applications. In: IEEE international conference on mobile services, pp 108–115
19. Marquez MD, Lara RA, Gordillo R (2014) A new prototype of smart parking using wireless sensor networks. In: Colombian communications and computing (COLCOM), pp 1–6
20. Sharma S, Chhatarpal RH (2014) Survey on internet of things and design for a smart parking. Int J Inventive Eng Sci 2(9):11–16
21. Diaz Ogás MG, Fabregat R, Aciar S (2020) Survey of smart parking systems. Appl Sci 10(11)
22. Fahim A, Hasan M, Chowdhury MA (2021) Smart parking systems: comprehensive review based on various aspects. Heliyon 17(5):07050. https://doi.org/10.1016/j.heliyon

A Hybrid PSO-Fuzzy Trust Energy Aware DRP in Wireless Sensor Network

Y. Chittibabu, B. Spandana, Roja Gurrapu, and G. Venkata Hari Prasad

Abstract Minimization of consumption of energy, improving transmission of data in a secure manner, and enhancing the life span of the network are the main objective of wireless sensor network (WSN). Trust is the feature that is utilized in day to day life and helps in decision-making during complex issues. A trusted network can compute via applications of fuzzy logic in neighboring nodes. An important heuristic algorithm is computer program for Shuffled Frog Leaping (SFLA) that belongs to the category of swarm intelligence optimization and triggers novel heuristic research in finding an optimal solution for a mathematical function. The movement of swarms and social behavior of fishes and birds are mimicked in the population-based Particle Swarm Analysis (PSA).This work presents a hybrid PSO-Fuzzy Trust Energy Aware Directed Random Propagation (DRP) routing.

Keywords Wireless sensor network (WSN) · Trust · Fuzzy logic · Computer program for Shuffled Frog Leaping (SFLA) · Particle Swarm Analysis (PSA)

Y. Chittibabu (✉) · G. Venkata Hari Prasad
Department of Electronics and Communication Engineering, CMR College of Engineering and Technology, Hyderabad, Telangana, India
e-mail: ychittibabu@cmrcet.ac.in

G. Venkata Hari Prasad
e-mail: drgvenkatahariprasad@cmrcet.ac.in

B. Spandana · R. Gurrapu
Department of Computer Science and Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India
e-mail: spandana.d@gnits.ac.in

R. Gurrapu
e-mail: rojagurrapu@gnits.ac.in

1 Introduction

In general, WSN is considered as a group of wireless sensor nodes which are large in number, cost conservative, resource restricted, and battery powered in nature. The life span of every node of WSN depends upon the topology as well as the whole network due to constraint in resources, battery power, and similarly relevant factors; thus, power consumed can be considered as the major dependent factor of WSN research [1]. The WSNs are designed in such a way that their main aim is to utilize the available resources effectively and satisfy quality requisites such as reliability and real time. Moreover, energy can consume more due to increased reliability and low delay, which can reduce the service lifetime of WSN. Thus, efficiency of WSN can be improved depending upon how well the communication protocol is optimized for meeting the restrictions on quality of service in restricted energy conditions.

Solution for even complicated problems may be found with no human intervention through the application of fuzzy logic concepts [2] and its implementation in tiny simplistic systems (microcontrollers) to huge networked systems. With the use of imprecise and redundant inputs definite answers are got. Precise outcomes are got by the use of fuzzy variable which uses in accurate data input. Fuzzy rules are user friendly and simple rules such as "if then else" is used for nontechnical users [3].

The main aim of energy efficient clustering and routing algorithms is to optimize energy consumption, which can improve the lifespan of SN, total packet transmission, and accumulation of dead SNs because of draining of energy. Energy can be conserved in a network by balancing load among SNs.

The initial trust of the node assessed with the help of direct or indirect method, packet integrity, and energy level combine together to form the ultimate trust of a node. In a network, the node having greatest final trust value is selected as Certificate Authority and this node has the authority to issue trust certificates and the validity of these certificates are limited and has to be renewed on elapsing. During data transaction, nodes without certificates will not be included. Reliable and unreliable nodes will be classified by a fuzzy analyzer, and once the malicious node asks for certification, the other nodes are notified by the certifying authority through agitations [4].

Kennedy and Ebenhart, in the year 1995, proposed PSO as a global optimization metaheuristic method, which is an extremely widespread optimization method. In SFLA, every frog in the algorithm act as a meme carrier; through personal interaction between frogs, the meme evolution is reached. The same scale meme groups are formed by SFLA frog population and every meme group have their own search strategy. All frogs mix together after specific number of iteration, and SFLA reaches the next round of local search, where message exchanged between ethnic groups. The balance between local search and global information assures that global optimal solution can be reached by SFLA [5].

This study involves a fuzzy trust-based energy aware multipath routing protocol such as Directed Random Propagation (DRP) with PSO and SFLA. Relevant literature is dealt with in Sect. 2, method used in the study is explained in Sect. 3, outcomes of the experiments are dealt with Sects. 4, and 5 presents the conclusion to the work.

2 Literature Survey

Micro-electro Mechanical Systems (MEMS) was the brain child of the technological evolution in digital electronics. Drastic changes were brought about due to this evolution in the area of micro-sensors. Micro- sensors are popular because of its low cost and energy consumption, self-configurable, high reliability, and its capacity to work in an heterogeneous environment. Routing sensors can be considered as the backbone of sensor network due to efficient automated systems as these can provide better usability of a system in the most effective and efficient manner. Based on certain relevant applications, categories of routing protocols were studied by Mittal and Iwendi [6]. This survey concentrated on the applicability of sensor routing protocols in applications at both indoor and outdoor.

Efficient routes are selected a fuzzy-based delay and intelligent routing method which is energy aware. Quality of link, residual energy, available buffer size as well as proximity (distance) were taken into consideration by a fuzzy logic system by routing decisions was proposed by Mothku and Rout [7]. In a network, in a routing path, the next hop node is the one with increased remaining energy and free available buffer, decent link quality as well as close proximity. Moreover, analysis of network performance was done with a number of network states. Network simulator 3 is used for simulation based on the performance metrics including the quantity of retransmissions, delay, lifetime, power usage in addition network stability.

Singh and co. [8] suggested a fuzzy logic mechanism with fuzzy routing method and Dijkstra's routing algorithm that chooses the shortest route. Within sensors, shortest and best paths are computed by Dijkstra's algorithm. Throughput, energy consumption, and packet delivery ratio (PDR) are evaluated and are contrasted with current methods. It was found out that the proposed model was better than many other existing techniques.

A 2-stage optimized power consumption approach called AIFMDMCS was formulated by Lakshmi and Priyaa [9]. CHs are elected in this work with the state of indeterminacy in selecting factors using the help of fuzzy decision-making based on intuition. Data that is received from cluster nodes is collected and integrated by the CHs. Intuitionistic fuzzy inference is used to transfer the data packets that are integrated to BS for enhancing load balancing, when there is high traffic and to detect collision. The results simulated show that the method has greater efficacy in improving the network's life, as in the case of WSN the optimum shortest path is detected, and there is vagueness at the time of electing both CHs and consideration is given to degree of indeterminacy.

An improved SFLA was proposed by Dai et al. [5] because of the issues such as falling easily into local optimum and local optimization decision present in Shuffled Frog Leaping Algorithm (SFLA). The algorithm can be balanced and replaced in the original SFLA by an adaptive weighting factor so as to balance the algorithm and replace the update strategy of SFLA. Four benchmark functions and optimization problems of WSNs were optimized and results of simulation show that both optimization schemes were achieved effectively and also a reasonable $W \times N$ routing plan was put forth effectively.

To find a solution to the discrepancy service excellence and the duration of WSNs, a novel power consumption equilibrium. Zhou et al. suggest a prototype, [1] using SFLA. This prototype utilized a mathematical expression for consumption of power which the physical layer uses transmit power first, followed by received power, bandwidth of the signal, and energy consumption balance's optimization function which is built using total consumption of energy and power transmission of WSN. In the second phase, neural wavelet network was used to decrease the extensive dependent feature of the optimization function, and the signal is solved using SFLA. Ultimately, during simulation the efficiency of this algorithm compared with others is studied and the outcome of this algorithm is more advantageous in terms of error frame, quantity of nodes surviving and life span of the network.

3   Methodology

This section details on Directed Random Propagation, Fuzzy Trust Energy Aware DRP, Particle Swarm Analysis, computer program for Shuffled Frog Leaping, and Proposed PSA-Fuzzy Trust Energy Aware DRP.

Directed Random Propagation (DRP)

Two messages from the neighborhood areas used by DRP in enhancing the propagation efficiency. "A Last Hop Neighbor List" (LHNL) is added by DRP toward the header of every sharing beforehand being spreading to the subsequent the node LHNL field is propagated to the next node with the neighbor list. On reception, the node compares one node from its own list and the LHNL field to neighbor that is not a LHNL is chosen randomly. The TTL is then decreased; the shares are passed to the following hop, and so forth. A random neighbor is chosen whenever the LHNL overlaps finally with or has the neighbor list of the relaying node, similar to that of PRP scheme. This propagation technique states that the probability of forwarding a share to and fro is reduced by DRP by discarding this propagation technique within any two consequent steps. When PRP and DRP are compared, DRP tries to pus outwards from a sender, and hence an improved propagation efficacy leads to the specified TTL value [10].

Fuzzy Trust Energy Aware DRP

Through comparison of recommendations, trust on similar interests is assessed generally of other users with direct experiences and generally is a part of reputation system. The trust level of every node is assessed in the trust model and malicious nodes that are the root cause for modifications of data packet, there is detection of loss and misreporting. To begin with, the trust level of every node is assigned a 1, with the assumption that all nodes are trustworthy. On the basis of malicious behavior, the node's trust level changes from 1 to 0. If the trust level is lesser than a pre-established threshold value the node is then considered is an evil node and are found out and isolated from a network. An active node's energy consumption is WSN is depends on its power of receiving, processing and transmitting a message.

The amount of power needed for sending k bits across a distance of d, n the space is calculated by the following equation

$$E(k, d) = E_k + E_{kd}, \tag{1}$$

where E is the energy required to run the circuitry.

Particle Swarm Optimization (PSO)

Ebenhart and Kennedy propose a new technique which is based on evolutionary computation and is called PSO in the year 1995 [11], which takes its inspiration from the social behaviors of fish schools and bird flocks. The roots of PSO are based upon social psychology and synthetic life, and in engineering in addition to computer science. It makes use of a "population" of components which move around in a difficulty with hyperspace specific velocities. At every iteration, adjustment of the particles' velocities is done stochastically; adjustments are made as per the previous the particle's optimal position as well as finest location in the neighborhood. A user defined fitness function was formed which was the basis of derivation of particle best and neighborhood best. The evolution of particle movement is toward the optimum or a close-to-optimal solution. Here, the term "swarm" can be considered from erratic movement particle count in a, for instance a swarm of mosquitoes can be a better example than flocking bird or school of fish. Size or nonlinearity of a problem does not affect PSO which is a computation intelligence-based technique and can find solution to number of problems dependent upon convergence to optimal solution, where a number of analytical techniques fail [12].

For every function, fitness particle will be executed, and there will be computation and storage of best solution. Best is the current optimal fitness value; test is the optimal fitness function obtained in the particles among neighbors. The velocity as well as position undergo update in each generation through Eq. (2)

$$
\begin{aligned}
V_{pd}^{k+1} &= \omega V_{pd}^k + c_1 r_1 (\text{pbest}^k - X^k) \\
&\quad + c_2 r_2 (\text{gbest}^k - X_{id}^k) \\
X_{pd}^{k+1} &= X_{id}^k + V_{pd}^{k+1}
\end{aligned}
\tag{2}
$$

wherein ω refers to inertia weight, that ranges between 0.2 and 0.9; k denotes the Iteration number; V^k denotes the velocity found in the dth dimension of its pth particle.

X^k denotes the actual position of the dth dimension belonging to the pth particle; the pbest and the gbest denote the memory of the particle; c_1 and c_2 denote the cognition, as well as the social factor; r_1 and r_2 denote the random functions that are distributed uniformly in [0, 1].

Advantages:

(1) Implementation of PSO is easy as there are only few attributes which needs adjustment.
(2) PSO has a better memory capability as the best value of its own and its neighbor a reremembered every particle member.
(3) Diversity of swarm is maintained by PSO as every particle utilizes the information that are associated with the successful particle so as to enhance itself.

Algorithm for Shuffled Frog Leaping (SFLA)

Shuffled Frog Leaping Algorithm was presented by Eusuff and others in the year 2003 which is based on a convergence of both PSO algorithm and competitive evolutionary thought. With regard to SFLA, each frog is considered as a meme carrier, and through various frog personal interaction, evolution of meme is achieved. The frog population in SFLA is classified into same scale meme groups, and their own local search strategy is performed by every meme group. After specific local iteration, there is remixing of all frogs, and next round of local search is entered in SFLA which shows that exchange of information can take place between ethnic groups. Local search and global information exchange equilibrium strategy ensures that SFLA can converge to the global optimum [13].

A number of frogs are present in leapfrog algorithm with similar structure and each one of them represents a specific solution. The division of population into various sub-population are called as memeplexes. There are number of memeplexes with a new set of "frogs" with differing cultures and their basis is on local strategies of search. Every frog in the memeplex will have its own ideas and also will be affected by community also. Once the process of jumping and evolution are finished, the cultural ideas of the sub-population can be mixed at the moment of calculation. Both local search as well as its jumping are carried out till the moment when the convergence criteria are reached.

The steps to be followed in SFLA are stated here [14]:

(1) SFLA has population P with the potential solution established through a new set and digital frogs (n)
(2) These frogs are ranked decreasing order of exercise, then are formed in to memeplexes, which are subsets (m).
(3) Frog i is given by: $X_i = (X_{i1}, X_{i2}, \ldots, X_{is})$, where S indicates the quantity of elements.

(4) The most physically fit frogs reside within each memeplex is denoted as X_w while the frog with the worst fitness is denoted as X_b.

(5) The frog that has global best fitness is X_g.

(6) Frogs that the worst fitness will be enhanced through formulas 3 and 4.

$$D_i = \text{rand}() \, (X_b - X_w) \tag{3}$$

$$X_{w\_new} = X_{w\_old} + D_i(-D_{\max} \leq D_i \leq D_{\max}) \tag{4}$$

where in rand refers to an arbitrary number within [0, 1]. D_i refers to a frog's size in jumping steps, while $D_{\max}$ refers to the largest step permitting a shift in the location of the frog. In the event of the new the fitness level of is greater more than the present one, X_w is accepted.

If there is no enhancement, it is computed as per formulas 3 and 4 which are repeated with X_b which is substituted by X_g. In the event of no enhancement, a new X_w this arbitrarily created the updation operation is iterated for a specific number of times. After a pre-established set of mimetic evolution steps in every memeplex, the evolved solutions are substituted in the a fresh populace Shuffle is the term for this. It encourages a global sharing of information among the toads. Next, the populace is ranked in descending performance ranking, and top performers position of the frog is updated. A reorganization of the set of frogs to their associated memeplexes is done till the conversion criteria are arrived at.

PSO-Fuzzy Trust Energy Aware DRP Being Proposed

Frequent monitoring of energy levels of nodes are done. Each one is allotted with a score of energy level, cooperation in clustering, stability of the link and neighbor node information maintenance with different weight ages such as $W2$, $W3$, $W4$, $W5$, and $W6$, and calculation is done using arithmetic mean value of neighbor node values. Final score is the trust score for every node and storing is done in the routing tables of respective nodes. Propagation of trust scores is done to all neighbor nodes, and every update of value is also communicated to them.

The trust level of the sensor node is initialized with T_i as trustworthiness while U_i is untrustworthiness. T_1, T_2 refers to the trusted nodes. In reputation units, the node values are calculated through the formula as follows (5 and 6),

$$\text{Mini: } T_i = \min(T_1, T_2), \text{ Min: } U_i = \min(U_1, U_2) \tag{5}$$

$$\text{Max: } T_i = \max(T_1, T_2), \text{ Max: } U_i = \max(U_1, U_2) \tag{6}$$

The trust as well as un-trust value are got through the following formula (7 and 8):

$$T_i = \frac{\text{avg}(T_1, T_2)}{1 - (\text{avg}(T_1, U_2) + \text{avg}(T_2, U_1))} \tag{7}$$

$$U_i = \frac{\text{avg}(U_1, U_2)}{1 - (\text{avg}(T_1, U_2) + \text{avg}(T_2, U_1))} \tag{8}$$

Using the *Ti* and *Ui* the evaluation level of the sensor networks are calculated through (9).

$$\text{Evaluation value} = \frac{T_i}{T_i + U_i} \tag{9}$$

In the process of CH selection, consideration is done for fuzzy rules with precedence and rule with repudiation. The energy levels are monitored using active fuzzy rules that have increased priority win comparison to the other rules. The cluster member nodes are handled using fuzzy rules. This makes the decision of whether a node should join or leave a cluster based on certain rules that are handled with exception. Fuzzy systems are strong in nature with effortless implementation which stands a benefit of handling nonlinear systems. Fuzzy energy efficient routing protocol uses fuzzy logic system. The primary objective of fuzzy energy effective routing protocol is minimization of energy consumed and maximization of network life span. Beginning from SNs to BS, optimum path is found using fuzzy system. SNs help in collect of data from its neighboring nodes

The flowchart of Proposed Particle Swarm Optimization is shown in Fig. 1.

This work involves a trust aware secure routing protocol with consideration given to optimal energy utilization. There are four states in energy utilization—stage 1 is formation of cluster, and the clustering of clusters based on distances between nodes and k value I selected by the administrator of the computer network. In stage 2, the CH for every cluster is chosen by the use of trust scores, the distance, the speed of movement, and energy available. In the second stage, all clusters are also changed dynamically depending upon the mobile nodes movement by use of fuzzy rules that are generated with the help of existing fuzzy membership function. Trust scores are calculated in the third stage and analysis of various attacks in the network is identified, and also untrusted mobile nodes are identified. Reliable routing in network environment is done through untrusted mobile nodes that are isolated.

4   Results and Discussion

The number of nodes used for experiments are 100, 200, 300, 400, 500, and 600. Table 1 shows the parameter of PSO. Tables 2, 3, and 4 as well as Figs. 2, 3, and 4 reveal the likelihood of interception, typical end-to-end delay (sec) as well as Average Packet loss rate (%).

The interception probability for the SFLA-Fuzzy Trust Energy Aware DRP is given in Table 2 and picture 2. Outperforms by 28.6, 11.8, and 52.63% compared to Energy aware DRP with Fuzzy Trust at 100, 300, and 600 nodes, respectively. The likelihood of an interception for the SFLA-Fuzzy Trust Energy Aware DRP
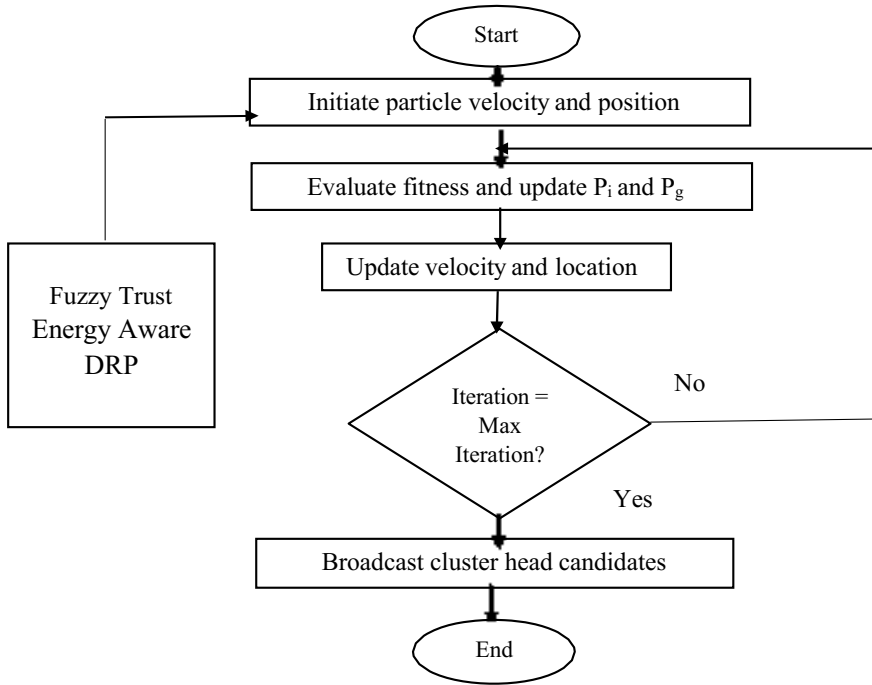
Fig. 1 Flowchart for proposed particle swarm optimization

Table 1 Parameters of PSO

PSO parameter	Value
Number of people	100
The most iterations allowed	20
The weight of inertia	W_{max} 0.9, W_{min} 0.4
Velocity is constant	$C1 - 2, C2 - 2$
Mistake gradient	$Ie - 06$

Table 2 For SFLA-Fuzzy Trust Energy Aware DRP, the likelihood of an interception

Amount of nodes	Energy aware DRP with fuzzy trust	Trust energy aware DRP PSO-fuzzy	Trust energy aware DRP SFLA-fuzzy
100	0.16	0.135	0.12
200	0.13	0.1	0.09
300	0.12	0.09	0.07
400	0.09	0.07	0.06
500	0.07	0.055	0.04
600	0.06	0.035	0.02

Table 3 End-to-end SFLA-Fuzzy Trust Energy Aware DRP average delay (s)

Amount of nodes	Energy aware DRP with fuzzy trust	Trust energy aware DRP PSO-fuzzy	Trust energy aware DRP SFLA-fuzzy
100	0.00161	0.001515	0.00152
200	0.00172	0.00165	0.00163
300	0.01784	0.016045	0.01648
400	0.03055	0.025335	0.02889
500	0.0619	0.05647	0.05849
600	0.06835	0.063815	0.06403

Table 4 Average packet loss rate (%) for SFLA-Fuzzy Trust Energy Aware DRP

Amount of nodes	Energy aware DRP with fuzzy trust	PSO-energy-conscious DRP with fuzzy trust	Trust energy aware DRP SFLA-fuzzy
100	6.91	6.615	6.41
200	10.74	10.31	10.17
300	10.57	10.175	9.94
400	16.04	15.465	15.29
500	20.34	19.68	19.49
600	24.71	23.445	23.29
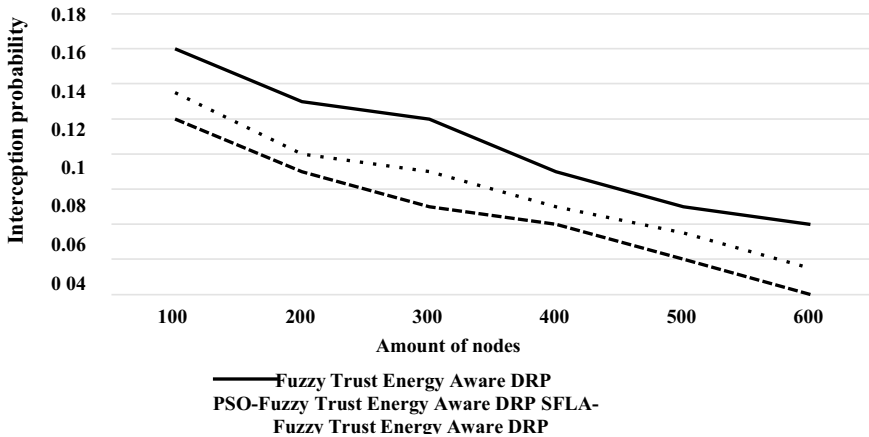
Fig. 2 Probability of interceptions for SFLA-Fuzzy Trust Energy Aware DRP

outperforms by 25, 100, and 54.55% compared to PSO-Fuzzy Trust Energy Aware DRP, respectively, at 100, 300, and 600 nodes.

It is noted in Table 3 and Fig. 3 that the average end-to-end delay (s) for SFLA-Fuzzy Trust Energy Aware DRP outperforms by 5.75%, 0.33%, and 7.93% compared
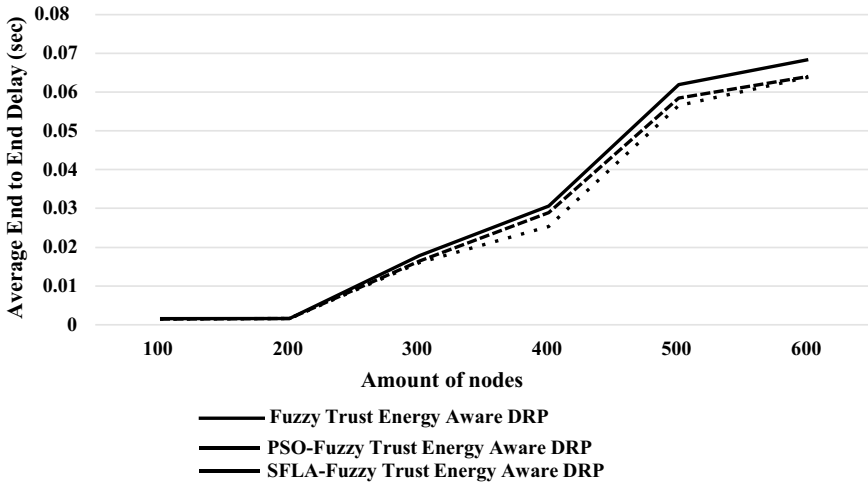
Fig. 3 Average end-to-end delay (s) for SFLA-Fuzzy Trust Energy Aware DRP
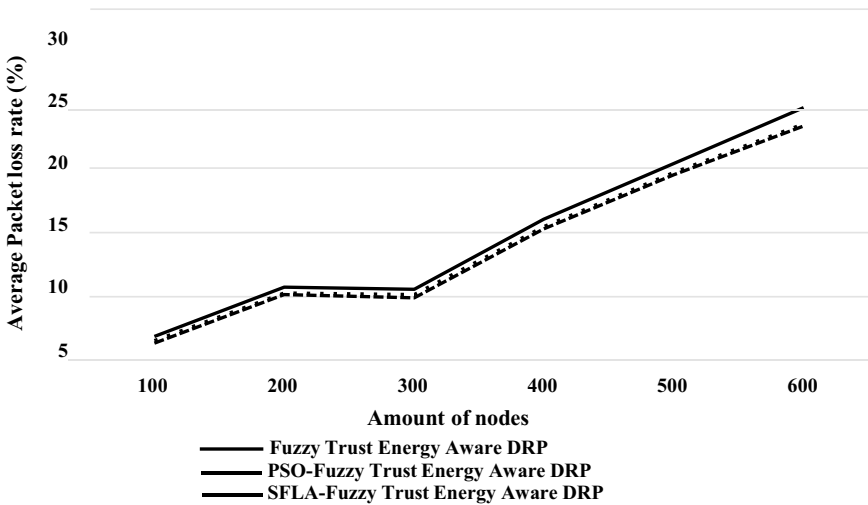
Fig. 4 For the SFLA-Fuzzy Trust Energy Aware DRP, the average packet loss rate is (%)

to Fuzzy Trust Energy Aware DRP at 100, 300, and 600 nodes, respectively. The average end-to-end delay (s) for SFLA-Fuzzy Trust Energy Aware DRP outperforms by 2.67%, 6.53%, and 0.34% compared to PSO—at 100, 300, and 600 nodes, respectively, Fuzzy Trust Energy Aware DRP.

Table 4 and Fig. 4 show that the SFLA-Fuzzy Trust Energy Aware DRP's Average Packet Loss Rate (%) outperforms by 7.51%, 3.15%, and 6.14% compared to Energy

aware DRP with Fuzzy Trust at 100, 300, and 600 nodes, respectively. The SFLA-Fuzzy Trust Energy Aware DRP average packet loss rate (%) outperforms by 2.34%, 5.92%, and 0.7% compared to PSO-Fuzzy Trust Energy Aware DRP at 100, 300, and 600 nodes, respectively.

5 Conclusion

In WSNs, a number of sensor nodes are collected and scattered through a large wireless environments. Results show the likelihood of an interception for the SFLA-Fuzzy Trust Energy Aware DRP outperforms by 28.6%, 11.8%, and 52.63% compared to Energy aware DRP with Fuzzy Trust at 100, 300, and 600 nodes, respectively. The likelihood of an interception for the SFLA-Fuzzy Trust Energy Aware DRP outperforms by 25%, 100%, and 54.55% compared to Trust Energy Aware DRP PSO-Fuzzy at 100, 300, and 600 nodes, respectively.

References

1. Zhou C, Wang M, Qu W, Lu Z (2018) A wireless sensor network model considering energy consumption balance. Math Probl Eng
2. Mohamed DYABI, Abdelmajid HAJAMI, Hakim ALLALI (2014) CATP: an enhanced MANETs clustering algorithm based on nodes trusts and performances. Int J Innov Technol Explor Eng 4(1):1–9
3. Chaudhary A,Tiwari VN, Kumar A (2016) A new intrusion detection system based on soft computing techniques using neuro-fuzzy classifier for packet dropping attack in MANETs. Int J Netw Secur 18(3):514–522
4. Manoj V (2012) A novel security framework using trust and fuzzy logic in MANET. Int J Distrib Parallel Syst 3(12):285–299
5. Dai YQ, Wang LG, Shi QH, Song L (2012) Performance analysis of improved SFLA and the application in economic dispatch of power system. Power Syst Prot Control 10
6. Mittal M, Iwendi C (2019) A survey on energy-aware wireless sensor routing protocols. EAI Endorsed Trans Energy Web 6(24)
7. Mothku SK, Rout RR (2019) Adaptive fuzzy-based energy and delay-aware routing protocol for a heterogeneous sensor network. J Comput Netw Commun
8. Singh M, Rai A, Kumar M (2019) Fuzzy logic trust based routing in wireless sensor network. Gyancity J Eng Technol 5(1):6–15
9. Lakshmi NG, Priyaa DS. Handling of indeterminacy for trust aware energy consumption using adaptive intuitionistic fuzzy environment in wireless sensor networks
10. Del Valle Y, Venayagamoorthy GK, Mohagheghi S, Hernandez JC, Harley RG (2008) Particle swarm optimization: basic concepts, variants and applications in power systems. IEEE Trans Evol Comput 12(2):171–195
11. Arjmand S, Adibnia F (2016) Job scheduling in cloud environment based on shuffled frog leaping algorithm. Int J Human Cult Stud (IJHCS) 1(1):290–302
12. Sahoo BM, Amgoth T, Pandey HM (2020) Particle swarm optimization based energy efficient clustering and sink mobility in heterogeneous wireless sensor network. Ad Hoc Netw 106:102237

13. Haseeb K, Islam N, Saba T, Rehman A, Mehmood Z (2020) LSDAR: a light-weight structure based data aggregation routing protocol with secure internet of things integrated next-generation sensor networks. Sustain Cities Soc 54:101995
14. Strumberger I, Bacanin N, Tuba M (2017, December) Hybridized elephant herding optimization algorithm for constrained optimization. In: International conference on hybrid intelligent systems. Springer, Cham, pp 158–166
15. Moussa N, El Belrhiti A, El Alaoui A (2021) An energy-efficient cluster-based routing protocol using unequal clustering and improved ACO techniques for WSNs. Peer-to-Peer Netw Appl 14(3):1334–1347
16. Sahoo BM, Amgoth T, Pandey HM (2020) Particle swarm optimization based energy efficient clustering and sink mobility in heterogeneous wireless sensor network. Ad Hoc Netw 106:102237
17. Mishra AK, Kumar R, Kumar V, Singh J (2017) Advances in computational intelligence. In: A grid-based approach to prolong lifetime of WSNs using fuzzy logic. Springer, Singapore, pp 11–22
18. Otoum S, Kantarci B, Mouftah HT (2017) Hierarchical trust-based black-hole detection in WSN-based smart grid monitoring. In: Proceedings of the 2017 IEEE international conference on communications (ICC), Paris, France, pp 1–6, 21–25 May 2017
19. Ayati M, Ghayyoumi MH, Keshavarz-Mohammadiyan A (2018) A fuzzy three level clustering method for lifetime improvement of wireless sensor networks. Ann Telecommun 73(7–8):535–546

Online Advertising Dataset Using ANN (Artificial Neural Networks) and LR (Linear Regression Techniques)

A. Srinivasulu, K. B. Chowdappa, M. Deena Babu, L. Venkateswara Reddy, and A. Vijay Kumar

Abstract Advanced displaying provides online sentiment digital advertisers with information into their sales. Sales are critical in increasing the business when sentimental digital marketing methods are used. Sentiment digital marketing is one approach for providing information about their company, gadgets, and products. Utilizing web-based advertising metrics like aggregation are designed to save you time when you need the most recent information for a show or report you're dealing with against a cutoff time. Business members can use our unique time span to expand their business. The current system has a high error rate, a poor business visualization strategy, and a high time complexity. AI technology has been employed in a variety of ways to improve the reach of target audiences in online targeted digital advertising. Recent study shows that improved computational force promotes capacity-focused granular crowds. This study explores and identifies several AI technologies used to improve specific web-based advertising. These three categories widely recognize and divide the word-of-mouth, client-driven, and content-driven promotion via radio, television, and newspapers methodologies that compose AI-based Internet designed promoting strategies. The proposed AI computation accurately predicts information at 94.50% using linear regression and neural network methods.

Keywords Online advertising · Newspaper · Digital marketing data · Linear regression · ANN · Machine learning

---

A. Srinivasulu (✉)
Data Science Research Laboratory, Blue Crest University, Monrovia 1000, Liberia
e-mail: srinu.asadi@gmail.com

K. B. Chowdappa
Department of CSE, GPREC, Kurnool, A.P., India

M. D. Babu
Department of IT, MREC, Hyderabad, TS, India

L. V. Reddy
Department of CSE, KGRCET, Hyderabad, TS, India

A. Vijay Kumar
Department of AI, VJIT, Hyderabad, TS, India

1   Introduction

Extensive research on public relations tactics and strategies has been undertaken to boost promotions' viability with various crowds of objective people [1] reasoning breakthroughs give Internet marketing an advantage above traditional approaches by allowing computerized adverts to be streamlined. AI (ML)-based algorithms increase focusing accuracy by anticipating the most significant advertisements for customers based on context-oriented or previous customer data [2]. These advancements in artificial intelligence and data-driven technologies serve to alleviate problems encountered by marketing specialists while, in general, improving client experience. Several experiments have been conducted to develop AI algorithms that improve the distribution of specific adverts [3]. However, the key AI strategies utilized to boost concentration and public awareness have not been defined to the best of the authors' knowledge. As a result, this study investigates and categorizes numerous uses of AI methods into 02 broad categories: content-driven and client-driven advertising approaches. Furthermore, unique for detecting click extortion for AI approaches are identified.

2   Literature Review

In an ever-confusing advanced environment, AI addresses the difficulty of accurately anticipating specific objective crowd members. Target advertising attempts to send the most critical marketing messages to customers [2], and it employs AI-based approaches to automate and progress cycles for prospective customers recognition, market, data extraction segmentation. Individually shared and consumed material is more important than fragment and land data alone in predicting objective crowds and their purchasing behaviour, hence AI applications to client and content-based approaches surpass traditional market division [4]. For example, abstract elements of client-created information on various electronic media platforms, such as Twitter, can be utilized to accurately anticipate and coordinate target swarms [5].

Furthermore, increases in personalization and a reduction in interference with marketing communications help to higher client retention, advertising effectiveness and speculating profits.

Conduct focusing (BT) is one approach for getting targeted marketing to the right people. BT selects the most relevant adverts for customers based on recorded customer activity, such as identifying clicked joins, pages viewed, look, and prior transactions from the client's browsing history [2]. Because of the importance of web indexes such as Google, online endeavours and web browsing have developed as two of the most well-known online habits. Web browsing behaviour assists advertisers in learning about the tastes of their clients and identifying population segments. Using actual customer online behaviour increases the importance and personalization of marketing communications to targeted customers [6]. Furthermore, the customer's search queries work in tandem with the sponsor's catchphrases to determine which

notices should be sent to the client. Another technique to improve targeted advertising is to anticipate customer behaviour in real time. A prominent method for estimating Internet ads is the cost-per-click (CPC) model, in which advertisers are paid when customers click on a notice. As a result, anticipating a customer's active clicking factor (CTR), or likelihood of tapping on an offer, is unquestionably significant [7]. Placement, separation, and order of alerts are all aspects that influence CTR's average revenue [8]. Supported chase (SC) is the game plan of scholarly adverts in light of the match between the client's request question and expressions perceived by the advertiser [9]. For example, when the customers search for sponsor-purchased terms, the relevant advertisement is displayed. Promoting computerized promotion selection improves the consumer experience [10].

A recommended framework known as user profiling is a conduct-based methodology that searches for helpful instances in a client's behaviour to discover what the client finds interesting and uninteresting [10]. It is critical to determine clients' interests in order to personalize adverts to their preferences. Customers can be profiled based on attributes and preset classifications if they expressly (client provided data) or verifiably (prior web searches, audits, and reading) exhibit a mix of interests [11]. Personalization of communications is critical to improving the customer experience; therefore, the capacity to distinguish between clients through behaviour focusing is essential.

3   System Methodology

3.1   Existing System

Deep learning [2] strategies can make use of two existing approaches, such as neural networks [5], and have the following computation constraints.

The drawbacks are as follows:

   i. Tiny precision
  ii. More complexity in time
 iii. More execution time
  iv. More errors.

NN constraints of NN calculation: the following are some of the drawbacks:

   i. Less accuracy
  ii. High time intricacy
 iii. High execution time
  iv. High blunder rate.

Fig. 1 Sentiment digital marketing advertising input dataset

3.2 Input Dataset

See Fig. 1.

3.3 Proposed System

Techniques for machine learning like artificial neural networks include two proposed techniques (Fig. 2).

	A	B	C	D
1	TV	Radio	Newspaper	Sales
2	230.1	37.8	69.2	22.1
3	44.5	39.3	45.1	10.4
4	17.2	45.9	69.3	12
5	151.5	41.3	58.5	16.5
6	180.8	10.8	58.4	17.9
7	8.7	48.9	75	7.2
8	57.5	32.8	23.5	11.8
9	120.2	19.6	11.6	13.2
10	8.6	2.1	1	4.8

Fig. 2 Sentiment digital marketing advertising input dataset

4 Experiment Results

Our framework approach and implementation are founded on the essential principle of guaranteeing that the CSV data marketing data utilized in research meets the initial expectations of the arrangement sections. As a result, this framework configuration is a method or strength for displaying the layout, components, modules, interfaces, and data required for a suitable construction to meet basic requirements. There is some dispersion and collaboration with the informational collections in terms of their construction evaluation, framework approach [15], and framework structure. To assess implementation or capability, the application-predictable yield is utilized. It has been observed that critical details have a major impact on their system's inspection that brings about a feasible building to a common structure, given the key qualities of the appropriate patient; which eventually conformed to our required condition. Furthermore, it strives to suit the needs of present system users to an exceptional degree.

4.1 ANN Algorithm

We are forced to expand ANN to an all-encompassing ANN (S) in order to achieve greater accuracy, efficiency, and time complexity.

1. Start Python with latest version.
2. Import libraries for Pandas, Numpy, Seaborne, and Matplotlib.
3. Loading the CSV dataset.
4. Construct visualization.
5. Perform the target values into the data frame.
6. Converting categorical data into numerical data.
7. Use only one column for the target value during initialization.
8. Call corer () on data frame producing the intensity map that is sns.
9. Heat map reducing the attributes in.
10. X data frame.
11. Apply ANN to the CSV data stage.
12. Compare the predicted values and model fitting apply disarray network.
13. Print the classification [4] report for CSV data at last.

Hence as a result, it has been discovered that the results of calculations are more precise and take less time to complete; defining the CSV data as part of the initial expectation.

4.2 Results

See Figs. 3, 4, 5, 6 and 7.

Fig. 3 Sentiment digital marketing data classification

Fig. 4 Linear regression to predict sales

4.3 Validation/Evaluation Metrics

We demonstrated and evaluated the effects of our suggested method on CNN and RNN using the methods listed below. In order to investigate the confusion matrix, initial definitions include actual positive (AP), false positive (UP), false negative (UN), and actual negative (AN) individually. The number of cases was accurately predicted based on the requirements because of OP. In addition, B measures led to an incorrect estimation of the required number of examples.

Fig. 5 Stabilized comparison between radio and newspaper
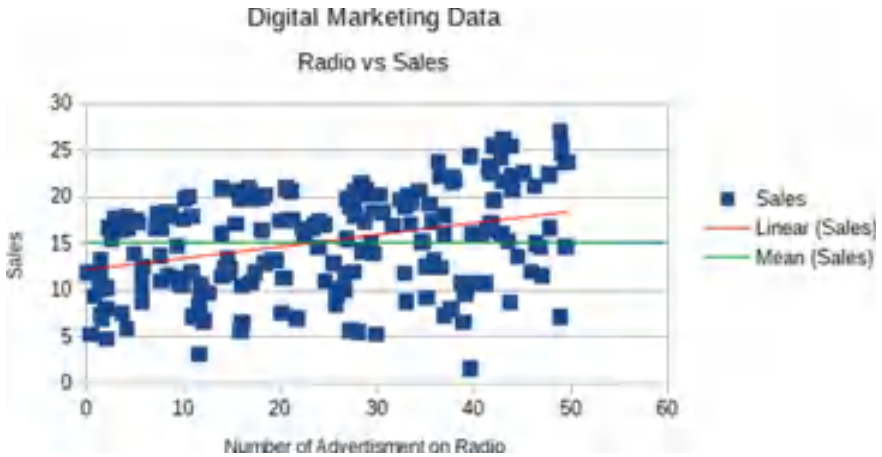
Fig. 6 Final conclusion radio versus sales

$$\text{Quality} = \frac{\text{BP} + \text{UN}}{\text{BP} + \text{UP} + \text{BN} + \text{UN}}$$

$$\text{Preciseness} = \frac{\text{BP}}{\text{BP} + \text{UP}}$$

$$\text{Callback} = \frac{\text{BP}}{\text{BP} + \text{UN}}$$

$$F\text{ - measure} = \frac{2 \times \text{Preciseness} \times \text{Callback}}{\text{Preciseness} + \text{Callback}}$$

Fig. 7 Digital marketing data medium versus sales

5 Conclusion

Several Telangana hospitals and the UCI, Kaggle and GitHub repository donated data for the study. This study investigated how to improve online objective publicizing by merging various AI approaches such as web-based linear regression and neural networks marketing strategies. The current study was prompted by increased income in order to reconcile information-driven and man-made reasoning approaches to computerized marketing. TV, newspaper, and online advertising methods, as well as client-driven and content-driven radio methods, are identified and classified as separate Internet public relations approaches and individual AI-based tactics. Furthermore, the article recognizes the need for snap extortion based on AI detection approach to protect people against malicious clicks. The design prepares for future study into artificial intelligence (AI) and human-made brainpower approaches. It also intends to improve the focus and sufficiency of online advertising systems, as well as to solve advertising security and protection concerns. Finally, this research paper provides a generally outstanding judgement on advanced special simulated intelligence processes with high precision, low time complexity, and high throughput.

References

1. Weerahandi H, Hochman KA, Simon E, Blaum C, Chodosh J, Duan E, Garry K, Kahan T, Karmen-Tuohy SL, Karpel HC et al (2021) Post-discharge health status and symptoms in patients with severe COVID-19. J Gen Intern Med 36(3):738–745
2. Choia J-A, Limb K (2020) Identifying machine learning techniques for classification of target advertising. The Korean Institute of Communications and Information Sciences (KICS), Elsevier B.V. CC BY-NC-ND license; ICT Express 6:175–180. Available online at http://creativecommons.org/licenses/by-nc-nd/4.0/, https://doi.org/10.1016/j.icte.2020.04.012, www.sciencedirect.com
3. Saura JR (2020) Using data sciences in sentiment digital marketing: framework, methods, and performance metrics. Department of Business Economics, Rey Juan Carlos University, Madrid, Spain, © 2020. Received 30 Mar 2020. Accepted 3 Aug 2020. Available online 15 Aug 2020; J Innov Knowl. Elsevier Espana, S.L.U. CC BY-NC-ND license. http://creativecommons.org/licenses/by-nc-nd/4.0/
4. Aladwani AM, Dwivedi YK (2018) Towards a theory of socio citizenry: quality anticipation, trust configuration, and approved adaptation of governmental social media. Int J Inf Manage 43:261–272
5. Alrifai R (2017) A data mining approach to evaluate stock-picking strategies. J Comput Sci Coll 32(5):148–155
6. Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. Manage Sci 57(8):1485–1509
7. Arias M, Arratia A, Xuriguera R (2014) Forecasting with twitter data. ACM Trans Intell Syst Technol (TIST) 5(1):1–24
8. Avery J, Steenburgh TJ, Deighton J, Caravella M (2012) Adding bricks to clicks: predicting the patterns of cross-channel elasticities over time. J Mark 76(3):96–111
9. Ballestar MT, Grau-Carles P, Sainz J (2018) Customer segmentation in e-commerce: applications to the cashback business model. J Bus Res 88:407–414
10. Choi JA, Lewis R (2017) Culture and the star-power strategy: comparing American and Korean response to celebrity-endorsed advertising. J Glob Mark 30(1):3–11
11. Jamshidi M, Lalbakhsh A, Talla J, Peroutka Z, Hadjilooei F, Lalbakhsh P, Jamshidi M, La Spada L, Mirmozafari M, Dehghani M et al (2020) Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. IEEE Access 8:109581–109595
12. Chen Y, Kapralov M, Canny J, Pavlov DY (2009) Factor modeling for advertisement targeting. In: Advances in neural information processing systems, pp 324–332
13. Capatina A, Kachour M, Lichy J, Micu A, Micu A, Codignola F (2020) Matching the future capabilities of an artificial intelligence-based software for social media marketing with potential users' expectations. Technol Forecast Soc Change 151:119794
14. Lo SL, Cornforth D, Chiong R (2015) Effects of training datasets on both the extreme learning machine and support vector machine for target audience identification on twitter. In: Proceedings of ELM-2014, vol 1. Springer, Cham, pp 417–434
15. Lo SL, Chiong R, Cornforth D (2015) Using support vector machine ensembles for target audience classification on Twitter. PLoS One 10(4)
16. Li H, Zhang D, Hu J, Zeng H, Chen Z (2007) Finding keyword from online broadcasting content for targeted advertising. In: Proceedings of the 1st international workshop on data mining and audience intelligence for advertising, pp 55–62
17. Ta A (2015) Factorization machines with follow-the-regularized-leader for CTR prediction in display advertising. In: Proceedings of 2015 IEEE international conference on big data (big data), IEEE, pp 2889–2891

Artificial Intelligence and Deep Learning-Based Agri and Food Quality and Safety Detection System

Amogh Shukla, Debangan Mandal, Radhey Shyam Meena, and V. Vijayarajan

Abstract In a number of fields, such as image recognition, speech recognition, facial identification, and others, deep learning (DL) has proven to be a successful way for analysing huge volumes of data. It has recently started to be used by researchers in the fields of mechanical engineering and food science. We have a long history of research, but we have never encountered a study that used food as the primary research medium. In addition to providing detailed descriptions of the structure of any common convolutional neural network (CNN) architecture, artificial intelligence (AI), and Internet of things (IoT) data training methodologies, this paper introduces deep learning (DL). Using deep learning as a computational technique, we went through a sizable number of research to identify solutions to food-related problems such food recognition, calorie calculation, fruit, potato, meat, aquatic product safety detection, food supply chain, and even food contamination. Fruit, potatoes, meat, aquatic product safety detection, the food supply chain, and even food contamination are some of these problems. Different datasets, preprocessing techniques, networks, and systems were examined in each study, and the outcomes were compared to those of other studies. We investigated how big data may be used to regulate food standards and discovered some surprising trends. Since our findings indicate that DL has the potential to be a new tool for the inspection of food safety, it outperforms several methods like manual attribute extractors and conventional machine learning algorithms.

Keywords Neural networks (CNN) · Artificial intelligence · Deep learning · Food safety · IoT · Big data

A. Shukla · D. Mandal · V. Vijayarajan (✉)
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
e-mail: virtual.viji@gmail.com

R. S. Meena
Department of Industrial and Management Engineering, Indian Institute of Technology
and Ministry of New Renewable Energy, Kanpur, India

1 Introduction

In order for everyone to lead a healthy and active lifestyle, safe, nutrient-dense food must be readily available, according to the World Food Submit's definition of "global food security" [1]. Maintaining food safety is a crucial aspect of the food industry's trade and commerce in the modern world (FI). It is also widely used because it is essential to human health in both developed and developing nations. An estimated 600 million cases of food-borne illnesses occur each year, causing more than 420,000 fatalities. It is anticipated that food-borne viral infections will cost the US economy more than 10 billion dollars annually. As a result, the increased demand for food around the world raises the stakes for food safety during the manufacturing and distribution processes. Regardless of how difficult it may be, ensuring food safety is critical. Because FI is so complex, concerns about food safety can arise at any time. The first stage of the FI is food production, which is also an activity that is critical to maintaining the integrity of the food supply. The food retail industry strives to reduce the amount of unsold food that ends up in landfills. These four stages can be thought of as a product's "life", and each one can have an impact on the quality of the food produced [2].

Primary goal of AI is to enable computers to "understand" the meanings of data. Significant examples have emerged in computer vision [3], language processing, and other fields in recent years. One of the most obvious advantages of AI is its ability to automatically extract information from large amounts of data. As a result, massive amounts of data are required for AI. Big data approaches are mostly used for gathering, storing, and querying massive amounts of data, allowing for basic processing. There are several redundancies in the data obtained. As a result, artificial intelligence and big data could be used to extract relevant information from redundant data.
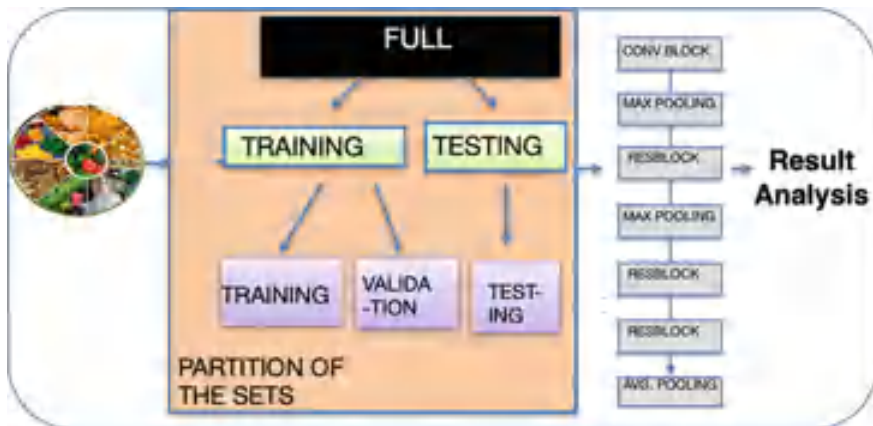
Fig. 1 Proposed model of this work

Food safety has gotten a lot of attention from the scientific community because of its importance in the FI, especially in the fields of food science and even information science. Figure 1 clearly shows that people's interest in food safety has grown over the last decade, as evidenced by the increasing number of Google searches for terms like "blockchain" , "big data", and "AI". In recent years, there has been a surge in research on distributed ledger technology (blockchain), artificial intelligence (AI), and big data [4–6]. The researchers presented a number of examples to substantiate their concerns and potential as part of their investigation into the role that big data plays in ensuring food safety. Vision, text, and functional-based solutions are just a few of the artificial intelligence (AI) applications in the food industry. A paper published earlier in 2021 [7] presented new supply chain solutions that use blockchain technology or its applications. These works investigated a wide range of applications. Food safety systems, on the other hand, are frequently complex systems that use AI, big data, and blockchain technology to achieve their goals. As a result, the research topic necessitates a thorough examination of food safety technologies.

A deep learning is a set of machine learning algorithms that include:

- To extract and manipulate information, they employ a massive number of several layers of nonlinear processing units.
- They can learn supervised (e.g. categorization) or unsupervised.
- They learn several layers of play that correspond to different levels of abstraction; these levels comprise a set of concepts.

This paper investigates instances where AI learning data was used, as well as cybersecurity attack strategies to improve AI dependability. The significance of controlling learning data prior to machine learning is discussed in this context. This is accomplished by investigating instances in which data was incorrectly used. Another AI solution that preserves data is blockchain-based learning data ecosystem models [8]. These models are used to ensure that no tampering has occurred with the data.

The remaining sections of the paper are organised as follows: Sect. 2 presents the most recent study, while Sect. 3 discusses the recommended strategy. In the following section, we will go over the results of the tests we ran to compare the performance of our system to that of other systems and provide a summary of our findings. In terms of Sect. 5, we have reached the end of the road.

2 Literature Review

According to [9], using the technique presented in this study to evaluate the aesthetic quality and differentiate the quality of the requirement can result in a very accurate rate, which must have a positive impact on peanut output and industry growth. In the table, 100 (97 + 2A + 1P) denotes the presence of 100 grains in this grade, 97 of which are the same as the number obtained by hand, + 2A denotes the presence of

two worm-eaten grains labelled as D, and +1H denotes the presence of one normal seed labelled as P. The study investigates the image processing method used to evaluate the quality of peanut kernels; using national standard data, multiple amplitudes of the output designs are developed. The data in [10] could come from agriculture, food production, supply chain, traceability, or consumers. Sensors are data collection points on the Internet of things, but consumer opinions expressed on social media sites are the data source. Data processing is now typically performed remotely using high-performance computers, a practice known as "cloud computing". The process information gathered can be used to make decisions about how to improve the performance of the activities or to make relevant recommendations.

In the study [11], neural networks trained with acoustic frequency responses are used to detect dairy products that contain or do not contain non-dairy additives (NDA) and to differentiate organic from non-organic food items. Although the hypothesis is tested using butter samples, the method could also be applied to other dairy products. While trial and error would be time consuming, an artificial intelligence (AI) algorithm could be used to extract high-level features of a material's behaviour at various frequencies and compare them across different materials [4, 12]. The study [13] focuses on the application of 4.0 industrial revolution technologies in agriculture and the food industry, such as computer vision and artificial intelligence. The current research, in particular, provides a thorough understanding of computer vision and intelligence approaches used in a wide range of agricultural applications, including food production, agriculture-based apps, farming, plant data analysis, smart irrigation, and so on. In addition, the report emphasises the critical importance of employing environmentally friendly 4 IR technologies to ensure that humanity has enough food by 2050. Relevant sources and use cases were used to address the importance of the agriculture industry, as well as investments in AI and other vision technologies.

In [14], it is suggested that a cloud-hosted virtualization platform be used in the food manufacturing process. The data from NIR spectrometry on samples is fed into a set of artificial intelligence algorithms, which analyse the organoleptic properties of the samples. For example, the use of virtualization technology has increased the efficiency of cheese production. ICatador is a cloud-based platform that allows a wide range of stakeholders to interact and share information from a variety of locations. The report incorporates instrumental data on a consistent basis [15]. In [16], instead of manual grading, use a computer vision-based system that includes both hardware and software to assess product quality. Hardware, such as a camera, conveyor system belt, sensing devices, and variable speed sensors, is required based on the goods we grade. The collected photos' properties are retrieved, and image processing methods are used to preprocess the shots. A variety of industries are gradually developing grading techniques for evaluating product quality [17]. Finally, the appropriate artificial intelligence methodology is established to categorise and evaluate quality detection. Agriculture is the one of the most common application of artificial intelligence [18–23] (Table 1).

Table 1 Comparative analysis for food safety [24]

Fruits	Features	Technique	Accuracy
Mango	Change in colour in images	Sensor of IR vision and model of Gaussian mixture	Not defined
Harumani mangoes	Size, hue, and proportion	Model for separation based on Fourier	90%
Cashew	Hue, texture, dimensions, and form of cashews	Multiresolution wavelet transform and SVM and BPNN classifiers	95%
Cherry tomato	Colour, texture, shape	SVM and KNN classifier techniques	Not defined
Peanut	Hue, texture, and dimensions	BPNN techniques in AI	Not defined
Apple	Hue and texture	FNN and SVM techniques	89%
Mango	Dimensions, shape, weight, and defects of surfaces on mangoes	SVR, MADM, and FIL AI techniques	87%

3 Proposed Methodology

Machine learning has proven to be a useful technique for data analysis in a wide range of industries. Because of the difficulties in analysing raw natural data, traditional machine learning systems always include a step involving human function extraction. A computer may use representation learning to generate features from raw data for purposes of identification, grouping, or regression. This is an important theory because the method of convolution is critical to the operation of a convolutional neural network. Convolution is the first layer of a convolutional neural network, and it uses a variety of kernels to extract new information from a picture. Similarly, it is up to the individual to determine the number of completely linked layers. Several kernels are used within the convolution layer to extract more characteristics from the images and text [25]. The max pooling technique significantly reduces both the size and number of network parameters. The information produced by this layer is sent on to the complete connection network layer after being transformed into a vector with only one dimension. At this level, the approaches that are most commonly used in neural networks are implemented. This layer can be repeated indefinitely to create a network with a higher level of complexity. Similarly, the user has control over the number of fully linked layers. Convolutional neural networks (CNNs), a subclass of deep neural networks, are commonly used in computer vision, and speech recognition provides an example of a typical CNN structure that could be used for image categorization (Figs. 2 and 3).
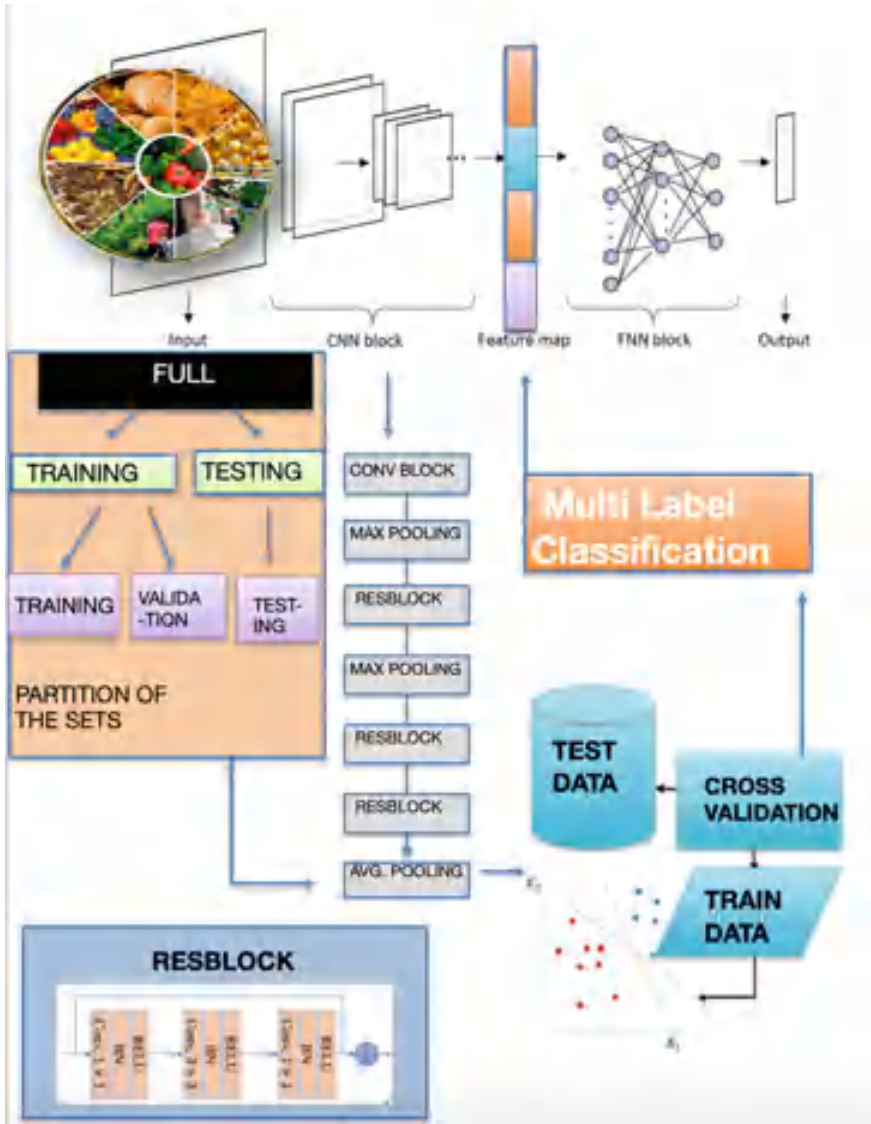
Fig. 2 Proposed model using CNN for food safety

The intelligent refrigeration is integrated into the ageing algorithm, voice indication, and picture processor to function. The input represents the material that is processed and inspected using the data inputs, and the output is displayed by the microprocessor in the form of signals. The result shows the age number as well as the ages of the items (primarily used for vegetables). The shelf life count is kept for a maximum of 30 days. Intelligent refrigeration has up to 96.55% dependability [26,
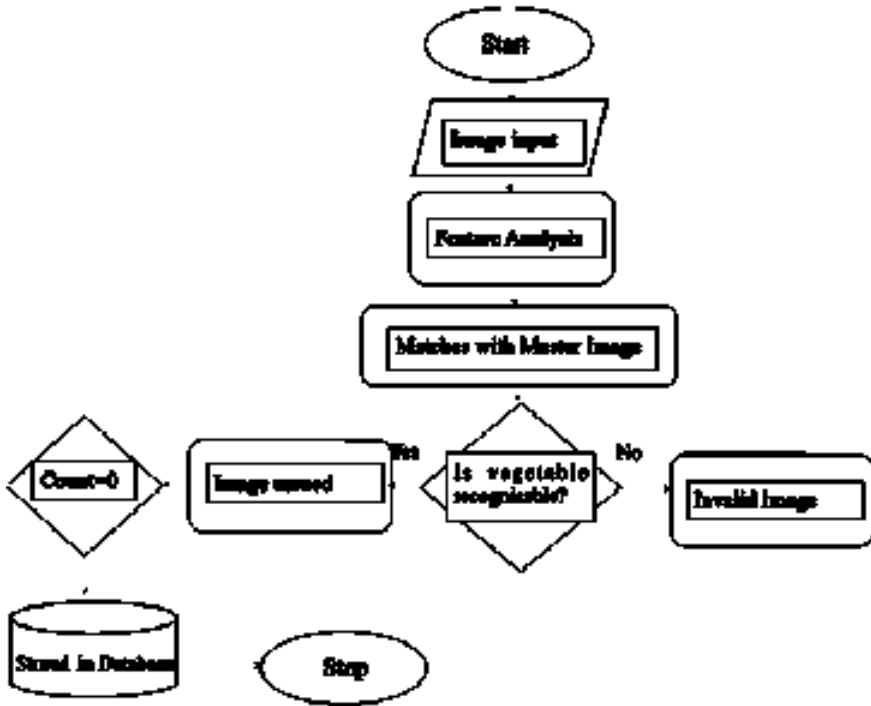
Fig. 3 Image Processing flow chart

30]. Fruit and vegetable intensive sorting entails sorting and grading them based on ripeness, weight, size, density flaws, and other factors. It was designed to sort ten belts of fruits at a rate of 15 fruits per second using infrared and ultraviolet images. The basic components of automatic visual inspection are (1) core control unit; (2) interfaces panel and storage centre; (3) weight sensor; (4) light sensors; and (5) output unit. The authors provided an update on the state of authentication in the Internet of things. A report examined the grading method of various farm goods using AI, which used a machine vision technique and comprised independent hardware and software components, for effective grading.

4   Results and Discussion

In order to compare the aforementioned classification efforts, researchers created a confusion matrix for each round of cross-validation to extract the counts for true positive (TP), false positive (FP), true negative (TN), and false negative (FN) instances [27] (Fig. 4; Tables 2 and 3).

$$FP = \sum_{j-1}^{M} C_{ij} - C_{ii} \tag{1}$$

$$TP = C_{ii} \tag{2}$$

$$FN = \sum_{j-1}^{M} C_{ji} - C_{ii} \tag{3}$$

$$TN = \sum_{i=1}^{M} \sum_{j=1}^{M} C_{ij} - (FP + TP + FN) \tag{4}$$

Researchers evaluated each strategy's performance using three metrics: I accuracy, II sensitivity, and III precision.

More than 125 distinct characteristics could be discerned from the high-quality images for example. The features were retrieved in each phase using methodical approaches to feature selection. The total number of possible characteristics when three different feature sets are combined is 119. In terms of feature sets, only seven elements are shared by both of these websites. This is due in part to the fact that the elytra of different beetle species and genera can appear very different from one another [28, 31]. As a result, statistical enrichment processes could not always highlight the relevant common features, because some of the recovered characteristics were discovered to be connected with one another and repeated.
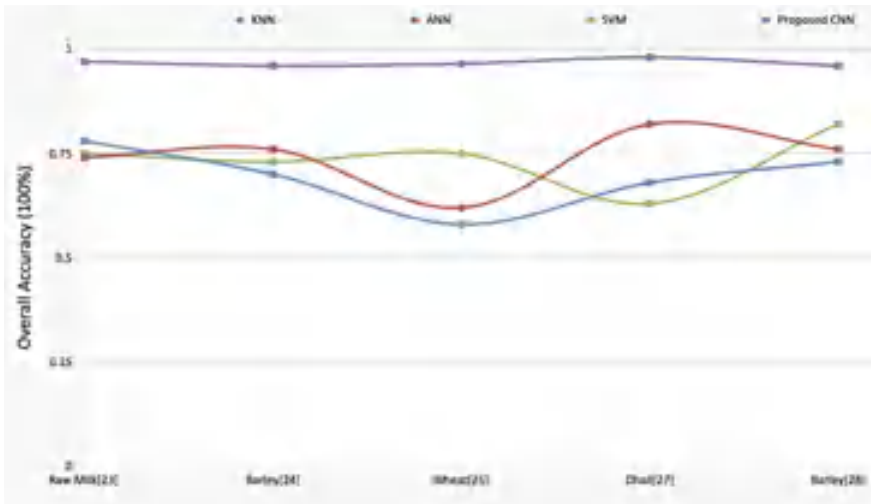
Fig. 4 Graph of overall accuracy for existing and proposed algorithms

Table 2 Result analysis of existing and proposed algorithms

S. No.	Indicators	Algorithms	Accuracy (%)	Sensitivity (%)	Precision (%)
1	Raw milk	KNN [17]	84.65	83.47	81.28
		ANN [28, 31]	85.60	85.40	87.61
		SVM [15, 20]	90.47	89.09	86.5
		Proposed CNN	98.25	97.21	96.9
2	Feed barley	KNN [17]	89.57	84.82	86.50
		ANN [28, 31]	92.62	91.52	90.91
		SVM [15, 20]	94.51	93.92	92.42
		Proposed CNN	97.83	96.62	95.98
3	Feed wheat	KNN [17]	70.62	69.61	68.83
		ANN [28, 31]	82.03	81.74	80.90
		SVM [15, 20]	90.40	89.30	88.74
		Proposed CNN	98.51	97.71	96.83
4	Feed rice	KNN [17]	86.15	85.38	84.15
		ANN [28, 31]	91.24	90.61	89.94
		SVM [15, 20]	95.28	94.75	93.69
		Proposed CNN	98.71	97.09	96.48
5	Feed dhall	KNN [17]	79.26	78.95	77.72
		ANN [28, 31]	85.49	84.47	83.27
		SVM [15, 20]	90.10	89.20	88.79
		Proposed CNN	97.19	96.43	95.86
6	Feed barley	KNN [17]	84.63	83.62	82.52
		ANN [28, 31]	89.32	87.31	86.24
		SVM [15, 20]	93.48	92.12	91.30
		Proposed CNN	98.72	97.14	96.42

Table 3 Food data selected and descriptions

S. No.	Indicators	Data sources	Countries
1	Raw milk	India commodity dashboard	India
2	Feed barley	India commodity dashboard	India
3	Feed wheat	India commodity dashboard	India
4	Feed rice	Food and Agriculture Organization	India
5	Feed dhall	Food and Agriculture Organization	India

5 Conclusion

This article demonstrates how the food industry can transition from an antiquated method to a more modern method that makes use of automation. Despite the rapid introduction of new technologies in the food industry, AI and BDA have success-

fully established dependable platforms for interacting with cutting-edge technology. Artificial intelligence (AI) guides the activities of various trans-disciplinary systems such as artificial neural networks (ANNs), machine learning, artificial sensing, computer vision, fuzzy logic approaches, and robotics to investigate aspects such as quality, colour, and texture. This one-of-a-kind method included data pattern analysis and workflow modification in order to provide a result that was accurate as well as trustworthy, used fewer human resources, was efficient, and assisted the user in predicting future occurrences over time. The food industry, which is currently dealing with an increase in the number of quality issues, may benefit from the use of these approaches. It was only a matter of time before drones became a significant component in food supply chain management. Furthermore, sensors are an important part of the food preservation process. Artificial intelligence (AI) and large amounts of data have enabled the food industry to produce superior results that are also more optimum and efficient.

References

1. Belasco WJ (2007) Appetite for change: how the counterculture took on the food industry. Cornell University Press, Ithaca, NY, USA
2. Langer J, Alfirevic N, Pavicic J (2017) Organizational change in transition societies. Routledge
3. Kryvoplias-Volodina L, Gubenia O (2018) Processes, equipment and control systems of food production. In: Section 2, 84 international scientific conference of young scientists and students. Youth scientific achievements to the 21st century nutrition problem solution
4. Rosenzweig ES, Brock JH, Lu P, Kumamaru H, Salegio EA et al (2018) Restorative effects of human neural stem cell grafts on the primate spinal cord. Nat med 24(4):484
5. Ortony A, University of Illinois, Urbana-Champaign, Clore GL, Collins A (1990) Cambridge University Press
6. Feigenbaum EA (1977) The art of artificial intelligence: themes and case studies of knowledge engineering. IJCAI
7. Araújo SO, Peres RS, Barata J, Lidon F, Ramalho JC et al (2021) Characterising the agriculture 4.0 landscape-emerging trends, challenges and opportunities. Agronomy 11(4):667
8. Miranda J, Ponce P, Molina A, Wright P (2019) Sensing, smart and sustainable technologies for agri-food 4.0. Comput Industr 108(2019):21–36
9. Sunil Chauhan N, Singh J, Chandra S, Chaudhary V et al (2018) Nonthermal techniques: application in food industries. A review. J Pharmacognosy Phytochem 7(5):1507–1518
10. Panda SK, Shetty PH (2018) Innovations in technologies for fermented food and beverage industries. Springer, Cham
11. Schwendicke F, Samek W, Krois J (2020) Artificial intelligence in dentistry: chances and challenges. J Dent Res 99(7):769–774
12. McCorduck P (2004) Machines who think: a personal inquiry into the history and prospects of artificial intelligence. CRC Press, USA
13. Bunker S (2018) Artificial intelligence: building smarter machines. Children's Book Media Rev 39(5):1–2
14. Ferreira PMM (2018) Artificial intelligence: an exploratory study about the impact on service
15. Rana A, Dhiman Y, Anand R (2022, Jan) Cough detection system using TinyML. In: 2022 international conference on computing, communication and power technology (IC3P). IEEE, pp 119–122
16. Crowther-Heyck H (2006) Patrons of the revolution. Ideals and institutions in postwar behavioral science. Isis 97(3):420–446

17. Singh H, Ramya D, Saravanakumar R, Sateesh N, Anand R, Singh S, Neelakandan S (2022) Artificial intelligence based quality of transmission predictive model for cognitive optical networks. Optik 257:168789
18. Tsoumakas G (2019) A survey of machine learning techniques for food sales prediction. Artif Intell Rev 52(1):441–447
19. Garre A, Ruiz MC, Hontoria E (2020) Application of machine learning to support production planning of a food industry in the context of waste generation under uncertainty. Oper Res Perspect 7:100147
20. Milczarski P, Zieliński B, Stawska Z, Hłobaż A, Maślanka P et al (2020) Machine learning application in energy consumption calculation and assessment in food processing industry. In: Rutkowski L, Scherer R, Korytkowski M, Pedrycz W, Tadeusiewicz R, Zurada JM. (eds) Artificial intelligence and soft computing. ICAISC 2020. Lecture notes in computer science, Springer, Cham, pp 369–379
21. Moshelion M, Altman A (2008) Current challenges and future perspectives of plant and agricultural biotechnology. Trends Biotechnol 33(6): 337–342. https://doi.org/10.1016/jtibtech.2015.03.001
22. Vilkhu K, Mawsona R, Simonsa L, Batesb D (2008) Applications and opportunities for ultrasound assisted extraction in the food industry—a review. Innov Food Sci Emerg Technol 9(2):161–169
23. Klyuchka E, Dmitri K, Vitaly D, Lukyanov A, Gaponov V (2019) New methods of seeds functional state and activity control for the development of the biotechnical feedback concept. AIP Conf Proc 2188(1):030015. https://doi.org/10.1063/1.5138408
24. Elakkiya N, Karthikeyan S, Ravi T (2018, March) Survey of grading process for agricultural foods by using artificial intelligence technique. In 2018 Second international conference on electronics, communication and aerospace technology (ICECA). IEEE, pp 1834-1838
25. Floridi L (2016) Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions. Philosophical Trans R Soc Math Phys Eng Sci 374(2083):20160112. https://doi.org/10.1098/rsta.2016.0112
26. Sharma S, Gahlawat VK, Rahul K, Mor RS, Malik M (2021) Sustainable innovations in the food industry through artificial intelligence and big data analytics. Logistics, 5(4), pp 66
27. Tesauro G, Touretzky D, Leen T (1997) Advances in neural information processing systems. MIT Press, Cambridge, MA, London, pp 1–9
28. Singh SK, Thakur RK, Kumar S, Anand R (2022, March) Deep learning and machine learning based facial emotion detection using CNN. In: 2022 9th international conference on computing for sustainable global development (INDIACom). IEEE, pp 530–535
29. Grunert KG, Jeppesen LF, Jespersen KR, Sonne AM, Hansen K et al (2005) Market orientation of value chains: a conceptual framework based on four case studies from the food industry. Eur J Mark 39(5/6):428–455
30. Velnath Ravanan R, Subasri MG, Vimal Kumar KT, Dhivya PS, Kumar, Roobini K (2021) Next generation smart garbage level indication and monitoring system using IoT. Smart Technol Commun Robot (STCR) 2021:1–4. https://doi.org/10.1109/STCR51658.2021.9588961
31. Michael MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349(6245):255–260

Personalized E-learning System Using Linear Regression for Intelligent Tutoring Systems

Anupama Vijaykumar and Vindhya P. Malagi

Abstract An e-learning system allows the learners to attend courses online from any geographical location, at any time using the Internet. The standard e-learning system does not give learners an individualistic model as it is not personalized and is inappropriate for all users. Hence, the satisfaction level of learning a course online is low for many users. This problem can be solved by applying an adaptive learning model for e-learning systems. This type of learning combines intelligent teaching with machine learning techniques to personalize the learners' learning experience. In adaptive e-learning, the learning content is delivered based on the learner's knowledge level, experience level related to the course, interests, and background. This kind of learning favours an effective way to deal with the self-paced learning. A framework for developing an adaptive system is explored here, based on the core concepts of adaptive/personalized e-learning systems or the so called intelligent tutoring systems (ITS).

Keywords Personalized learning · E-learning · Active learning · Adaptive learning · Machine learning · Linear regression · Intelligent tutoring systems (ITS) architecture

A. Vijaykumar (✉)
Jyothy Institute of Technology, Bangalore, India
e-mail: anupamarevadi@gmail.com

V. P. Malagi
Dayananda Sagar College of Engineering, Bangalore, India
e-mail: hod-ai@dayanandasagar.edu

1 Introduction

Technology has had a significant impact on all aspects of people's lives, especially in the education sector. The COVID-19 pandemic also helped accelerate these advancements in education. E-learning is a common term for online learning which is an Internet-based method of delivering learning materials or assessing the students' knowledge. Many e-learning courses (contents and assessments) are available on the Internet, however they are common to all learners and are not tailored to suit the learners' need or pace of learning and many merely supply HTML pages as course content which leaves some leaners unsatisfied with the e-learning system in use. The common content and assessment given on these systems cannot satisfy the needs of all its users in the same way. This paved way for an adaptive e-learning system that solves the shortcomings of the traditional e-learning system.

1.1 *Adaptive E-learning*

Adaptive e-learning is based on the premise that identifying a user's learning style, particularly in online learning, is a critical feature to be considered for improving ones experience in online learning. The user expects a unique teaching style that aids in the achievement of the learning goal. The ability to adjust to the user's learning style and educational background based on the user's ongoing activity during the learning process is the most important characteristic of adaptive e-learning. ITS are fast picking up as a popular method of education delivery considering the self-paced/personalized learning.

1.2 *Intelligent Tutoring Systems*

Assessments make up an integral part of the ITS system, the aim of the assessment is to quantify students' comprehension and performance, as well as to help both the educational system and the learners to gain a better understanding of their knowledge level and the gaps. With the help of systematic assessment and marking procedures, educational institutions are becoming increasingly adept at measuring students' knowledge levels. Assessment techniques are critical in such systems, since they help the instructors to determine how well the students have grasped the key concepts and also to track students' progress and performance in classroom learning. Accurate assessment can lead to tutoring that is more tailored to each student's needs, resulting in more successful learning. The sheer amount a teacher takes to grade would take a lot of time and would lead to human errors. With automated assessments, this problem is solved and it also gives the user an option to take as many tests as they want to make them confident with the learning material. It would also

reduce the workload on teachers and help them use their time productively to make better teaching plans for their students.

This paper contributes to the topic by analyzing the following methods used in AIEd and determining their effectiveness in the same. The paper presents nine other papers discussing this topic and gives an overview of their work in this field.

2 Related Work

a. In this paper by Muangprathub [1] the main aim is to develop a learning recommendation component in an intelligent tutoring system (ITS). This paper applied FCA as an intelligent method based on rule-based reasoning that dynamically predicted and adapted to a learner's style to recommend topics to the learner. The FCA was utilized to work on the information base supporting adaptive learning, which can be accomplished by reasonable information development. This was also used with students in a real teaching/learning environment to evaluate the performance of the proposed teaching approach. The results show that the proposed system can be used to help learners to improve their learning experience.

b. An artificial intelligence subject is discussed in [2] by Grivokostopoulou F., an educational system that supports students' learning and instructors' teaching search algorithms. Knowledge representation and reasoning, search algorithms, constraint satisfaction, and planning are the four key disciplines of the system. Intelligent tutoring systems (ITSs) are a type of computer-based educational system that incorporates intelligence to improve the efficacy of instruction. This is mainly achieved by employing artificial intelligence approaches to reflect their pedagogical judgements and knowledge of the domains they teach, learning activities, student characteristics, and evaluation. Furthermore, the evaluation system's performance was compared with that of professional (human) teachers to measure its effectiveness.

c. The PHP intelligent tutoring system (PHP ITS) by Weragama et al. [3] was created to make it easier for novices to learn the PHP language and create dynamic web pages. The PHP ITS works by giving students activities to complete and then offering feedback based on their answers. The PHP ITS accomplishes this by using artificial intelligence (AI) theories to model the subject content taught by the system, such as classical and hierarchical planning first-order predicate logic. The results revealed that it could accurately assess over 96% of student-supplied exercise responses.

d. Wang et al. [4] study intend to provide an intelligent software system that optimizes and enhances the self-learning process of computer engineering students during their courses. The system's assessment process is based on a hybrid artificial intelligence method that includes an artificial neural network (ANN) and the vortex optimization technique, an optimization algorithm (VOA). The

outcome was tested with students, and positive results were found in optimizing and enhancing self-learning.

e. Latham et al. [5] propose to develop a novel conversational intelligent tutoring system (CITS) that leads a tutoring conversation and dynamically predicts and adapts to student learning style. It mimics a human tutor by implicitly modelling the learning style during tutoring and personalizing the tutorial. Oscar is helpful, and their learning gain increased by 13%. Moreover, to establish a deeper understanding of the topic, learners can intuitively explore and discuss topics in natural language.

f. This research by Alshammari et al. [6] compares adaptive e-learning systems that consider learning style from three perspectives: learner model, domain model, and adaptation model. In addition, a set of relevant criteria was proposed to provide insight into the spectrum of approaches and procedures utilized in a representative collection of AES.

g. The goal of this review by Hlioui et al. [7] is to provide an overview of learner modelling in adaptive e-learning systems and the primary approaches for extracting personalization factors. It is inferred from this research that the straightforward method of collecting learner data is not dependable or absolute. At the level of tracking and assessing the learner's performances throughout a learning situation, the implicit derivation of the learner's model based on behavioural indications appears to be a more gratifying strategy. However, the behavioural markers provided in the literature are inadequate to model all learners' personalities.

h. Gurupur et al. [8] describe how to employ concept maps and Markov chain analysis to evaluate student learning outcomes. The primary goal of this tool is to promote the usage of artificial intelligence techniques by assessing a student's grasp of a particular field of study using idea maps and Markov chains. The tool's mechanism for performing the required evaluation uses XML parsing. Based on the experiment results and observations, we can infer that our application and its approach will assist instructors in identifying and evaluating their ability to induce strong knowledge of topics and concepts.

3 Methodology

3.1 Architecture and Design

In this section, the proposed ITS system is presented (Fig. 1).

The proposed system consists of four main modules: (a) the user profile and preferences module, (b) pre-questions module, (c) ML assessment module and (d)
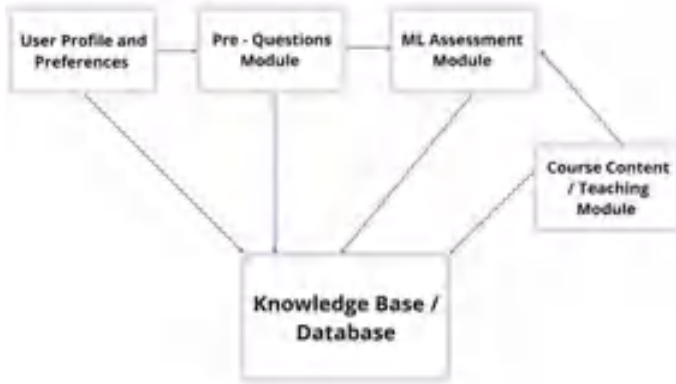
Fig. 1 Overview of model in the proposed system

course content/teaching module, the model also includes a knowledge base/database to store the information generated from these individual modules.

a. The user preferences module stores and collects the user's data and preferences; it keeps track of the progress of each individual and can be further used to generate a progress report. The user inputs some of this data; the information is the output of the other modules.

b. The pre-questions module is a module that evaluates the users' understanding on a particular course to determine a level. This module helps us determine from where to start teaching. The user answers a set of questions as soon as they sign-up; this questionnaire consists of a wide range of topics, from general questions like your name, age, etc., to technical questions on the subject. A simple adaptive algorithm evaluates these questions, and the algorithm's output is the level from where the user begins the learning journey.

   Below is the adaptive algorithm:

```
Step 1: Input: Age of the student
Step 2: If age > 5 and age<=10
                score=1
     elif age>10 and age<=15
                score =2
     elif age>15 and age<=20
                score =3
     else:
                score=4
  Step 3: If 'Yes' for choice
                if age>5 and age<=10
                    score+= 1
                elif age>10 and age<=15;
                    score+=1
                elif (age>15 and age<=20) or age>20
                    null
            Else 'No' for choice
                if age>5 and age<=10
                    null
                elif age>10 and age<=15
                    null
                elif (age>15 and age<=20) or age>20
                    score-=1
  Step 4: If right option
                score +=0.5

  Step 5: If right answer
                score +=0.5
        else
                null
  Step 6: If choice = option 1
                if score>3: score=2
                else: score=1
        if choice = option 2
                if score>3: score=2
        else: score = 2
        if choice = option 3
                if score>3: score=3
        else: score=2
 Step 7: Null
 Step 8: Print score
 Step 9: End
```

The course content is curated for five different levels. These levels differ in the content and the difficulty level of the questions in the assessment test. The five levels are:

Level 1: Beginner with no experience in coding.
Level 2: Beginner in this programming language but has experience with a different language.

Level 3: Basic.
Level 4: Intermediate.
Level 5: Advanced.

Depending on the evaluation of the level, the course content is provided to the learner.

c. The assessment module is a machine learning model that employs a linear regression model to give learners an adaptive exam experience. Linear regression is a machine learning approach that uses supervised learning. It performs a regression operation. Regression models the desired prediction value based on independent variables. It is mainly used to forecast and determine the relationship between variables. Linear regression is a mathematical technique for predicting the value of a dependent variable (y) based on the value of an independent variable (x). A linear link between x (input) and y (output) is identified as a result of this regression technique (outcome). The term "linear regression" was coined as a result. Linear regression hypothesis function: h: XY (Fig. 2).

The dataset we use is a set of questions which are then represented in binary, the correct answer is 0, and the wrong answer is denoted by 1. There is a column that is labelled, level change that is used to either move up to the next level or move down a level (Fig. 3).

The assessment algorithm starts the test on the level the user is at; it includes the questions taught and discussed in the course content. The algorithm is designed so that, for every three questions, there is a change in the level of difficulty of the questions. This is done to ensure the users answer questions ranging from easy to difficult on the course content. If the user answers a question wrong, then the algorithm evaluates the other answers given by the user and decides if there should be a decrease in the difficulty of questions. In the same way, if the user answers most

Fig. 2 Linear regression flowchart
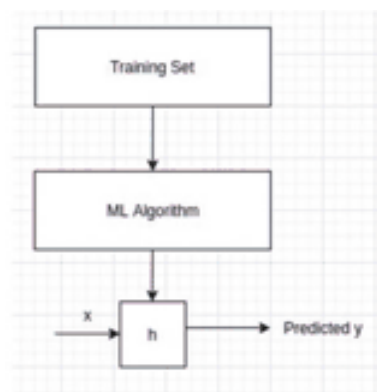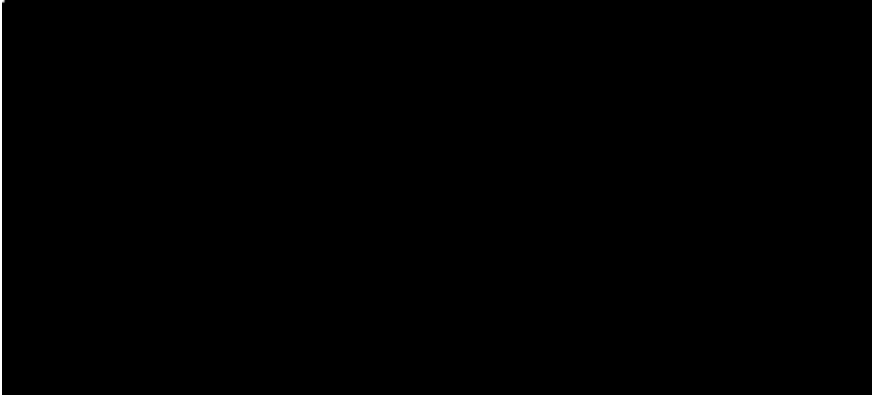
Fig. 3 Dataset for the linear regression model

or all questions correctly, then the difficulty level increases. The input for the level change is given in a boolean form, as shown above.

d. The course content module is a comprehensive collection of PPTs, videos, and other learning materials for the learners' disposal. This content is curated with ease of learning, prior knowledge, and the quality of content in mind. This module also houses the questions that are used in the assessment module. However, this module's primary use is to present the learning material in a presentable form.

The knowledge base is the place that stores all the course content and assessment questions. NoSQL format was used and Firebase was chosen as the database. The updated information from all algorithms is stored in the database.

4   Implementation

The implementations of the above-mentioned modules are done as follows:

The application is built using React, Typescript, and CSS on the front end. In the back end, Node.js, Python, and Firebase are used. A connection between the front end and the database is established with Firebase. In the same way, a connection between the back end and Firebase is done to retrieve data faster. Node.js is used for the connectivity of all these and the smooth running of the application. Firebase is a NoSQL data store backed by Google. It also offers the easy implementation of multiple APIs. In addition, Firebase also authenticates users and provides an option for Google Sign-up and Sign-in, which are implemented here (Figs. 4, 5, 6 and 7).
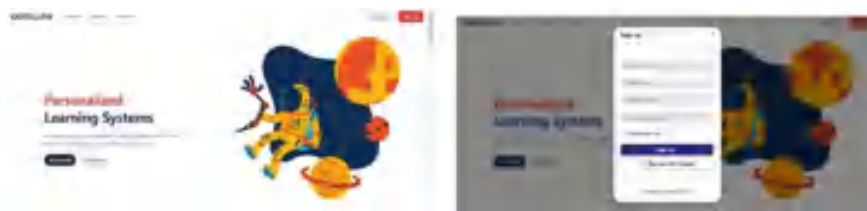
Fig. 4 Hero page and sign-up modal
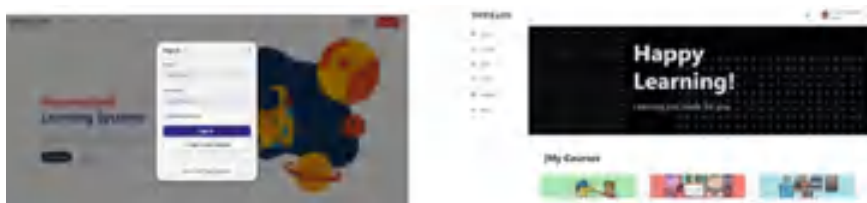
Fig. 5 Sign in modal and dashboard

Fig. 6 Course catalogue and quiz page

Fig. 7 Test catalogue

5 Conclusion

Personalized learning is a modification of the traditional e-learning system that makes it more interactive and user-centric. The main goal of this work is to focus on the adaptive e-learning system and its benefits, the proposal of an applied model in

light of adaptive learning's basic standard is showcased here. This research has presented personalized e-learning systems that integrate learning styles from four perspectives: (a) the user profile and preferences module, (b) pre-questions module, (c) ML assessment module and (d) course content/teaching module, the model also includes a knowledge base/database to store the information generated from these individual modules. ITS's are adaptive systems that employ intelligent technology to customize learning based on the characteristics of each learner. Using smartphone and web apps, this research generated learning suggestions in ITS that dynamically forecast and adapt to a learner's style. The assessment module is a machine learning algorithm that employs the linear regression model to provide users with adaptive test insight. Based on the web application, it built a primary programming language, Python programming language teacher, for learners. Pre-question module was utilized to evaluate the degree of course knowledge for adjusting to a learner's style to illustrate the suggested adaptive algorithm. All of the course and evaluation questions are kept in the knowledge base. It can be inferred from the experiments and observations that the tool and the approach linked with it will assist learners in recognizing and analyzing their capacity to induce strong knowledge of topics and enhance their learning methods.

6 Future Work

Future work for this research will be:

- Implementing the assessment module using the formal concept analysis.
- Development of language translation that helps the students learn better in their native language.
- Developing an automated report for the learners to give the insight of their performance.
- Providing the learners with automatic feedback to enable or disallow the insertion of unrelated topics.

7 Funding Details

References

1. Muangprathuba J, Boonjing V, Chamnongthai K (2020) Learning recommendation with formal concept analysis for intelligent tutoring system. 2405–8440/© 2020 The Author(s). Published by Elsevier Ltd.
2. Grivokostopoulou F et al (2017) An educational system for learning search algorithms and automatically assessing student performance. Int J Artif Intell Educ 27:207–240
3. Weragama D, Reye J (2014) Analysing student programs in the PHP intelligent tutoring system. Int J Artif Intell Educ 24:162–188
4. Wang D, Han H, Zhan Z, Xu J, Liu Q, Ren G (2015) A problem solving oriented intelligent tutoring system to improve students' acquisition of basic computer skills. Comput Educ 81
5. Lathama A, Crocketta K, McLeana D, Edmonds B. A conversational intelligent tutoring system to automatically predict learning styles
6. Alshammari M, Anane R, Hendley RJ (2014) Adaptivity in E-learning systems. In: 2014 eighth international conference on complex, intelligent and software intensive systems
7. Hlioui F, Alioui N, Gargouri F. A survey on learner models in adaptive e-learning systems. Multimedia Information System and Advanced Computing Laboratory University of Sfax
8. Gurupur VP, Pankaj Jain G, Rudraraju R (2014) Expert systems with application, 23 Dec 2014

Nucleus Segmentation Using K-Means Clustering for Analysis of Microscopy Images

Sourabh Singh, Vikrant Bhateja, Sparshi Gupta, Siddharth Verma, Shabana Urooj, and Dac-Nhuong Le

Abstract Analysis of microscopy images involves scanning of information related to color and texture. The color levels of nucleus and its neighbors especially in blood smear images helps to identify malignancy of the cells. The motive of this work is to perform color-based segmentation on the blood smear images. Primarily, these images are subjected to contrast stretching using dark contrast algorithm (DCA). This enhancement of the image increases the contrast of the nucleus and its adjoining components. To perform color-based segmentation on this enhanced microscopic image, k-means clustering is used. It is a type of iterative hard clustering where the segmentation criteria is dependent upon the similarity error which depends on Euclidean distance and position of centroids. The segmented image obtained in the form of color-based clusters is evaluated using overlap ratio (OR) which is beneficial in analysis of microscopic images.

Keywords Blood smear images · Centroid · DCA · Iterative hard clustering · K-means · OR · Similarity error

S. Singh · S. Gupta · S. Verma
Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM), Faizabad Road, Lucknow, U.P. 226028, India

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

V. Bhateja (✉)
Department of Electronics Engineering, Faculty of Engineering & Technology (UNSIET), Veer Bahadur Singh Purvanchal University, Shahganj Road, Jaunpur, Uttar Pradesh 222003, India
e-mail: bhateja.vikrant@gmail.com; bhateja.vikrant@ieee.org

S. Urooj
Department of Electrical Engineering, College of Engineering, Princess Nourah Bint Abdulrahman University, P. O. Box 84428, Riyadh 11671, Saudi Arabia

D.-N. Le
Faculty of Information Technology, Haiphong University, Haiphong, Vietnam
e-mail: nhuongld@hus.edu.vn

1 Introduction

Cell anatomy is based on study of blood cells. This anatomy is carried out using various features which are helpful in detecting malignancy in leukocytes (commonly known as white blood cells). In these features apart from their morphology their color variations also play a major part. The color features not only give us the information about healthy and malignant cells but also helps to differentiate between different types of WBCs. The conventional methods to identify types of cells is through manual observations, in these approaches also color parameters play a major role, but these approaches are subjected to inaccuracy due to parallax error [1]. Computer-based analysis helps us to identify and calculate parameters and variations which are negligible to the human eye [2]. There are several color variations and variations in randomness which can only be calculated using computer aided analysis [3, 4]. A contrast stretching technique proposed by Harun et al. [5] segments nucleus with a fusion of noise removal through median filter and dark contrast algorithm. Segmentation done through clustering helps in analyzing color-based features. A technique proposed by Su et al. [6] for segmentation bone marrow smear images uses k-means clustering. The segmentation was done through k-means clustering and hidden Markov random field (HMRF). This method of segmentation was based on color-based segmentation and was successful in categorizing WBC types. Another example of segmentation through k-means clustering to segment texts discussed in [7] shows us binary segmentation using k-means clustering. Another color-based segmentation approach used by V. Acharya et al. [8] for segmentation of blood smear images was through k-medoid clustering. The results of this clustering were quite satisfactory in color-based segmentation to detect blood disorders. A thresholding approach through multi-Otsu thresholding by Dasariraju et al. [9] was able to perform color-based segmentation but it did not store the initial color of nucleus. Apart from it, the results of cell classification had a good accuracy through this approach. A deep learning approach for detection of mutation in bone marrow smear images proposed by Eckardt et al. [10] was done using faster region-based convolutional neural net (FRCNN) and an image annotator tool for segmentation and boundary detection. Experimentally, k-means clustering is used as an effective technique in color-based segmentation. The approach selected for this paper primarily uses DCA for contrast stretching of the image and k-means clustering for color-based segmentation. The organization of the paper in the following parts is as: Sect. 2 contains the proposed approach of contrast stretching and color-based segmentation, Sect. 3 contains experimental results and analysis and Sect. 4 consists of conclusion.

2 Proposed Approach of Contrast Stretching and Color-based Segmentation

The proposed approach of performing color-based segmentation on blood smear images is carried out by contrast stretching preceded by segmentation. To perform contrast stretching of these blood smear images the technique used is dark contrast algorithm (DCA). This stretches the contrast of this RGB image making it experimentally perfect for segmentation. Further, the color-based segmentation is carried out by k-means clustering.

2.1 Contrast Stretching Using Dark Contrast Algorithm (DCA)

Dark contrast algorithm (DCA) is a contrast stretching technique which stretches the contrast of darker regions. This further helps to identify color pigmentation of blood cells and their components. The rate of how much the pixel intensity will be stretched is dependent on the parameters of DCA, namely threshold value (*TH*) and stretching factor (*NTH*). For dark stretching process *TH* should be smaller than *NTH* [11]. The procedural steps on which DCA works is given in Algorithm 1 [12].

Algorithm 1: Procedural Steps for the Proposed Approach of Contrast Stretching

　　Begin

　　　　　　Step 1:　　*Input* image as a

　　　　　　Step 2:　　*Initialize* DCA variables

　　　　　　　　　　　　TH←100

　　　　　　　　　　　　NTH←150

　　　　　　Step 3:　　*Read* variables of image p

　　　　　　　　　　　　N_r ← Number of rows

　　　　　　　　　　　　N_c ← Number of columns

　　　　　　　　　　　　min_a ← Minimum Intensity

　　　　　　　　　　　　max_a ←Maximum Intensity

　　　　　　Step 4:　　For $i = 1$ to N_r

　　　　　　Step 5:　　For $j = 1$ to Nc

　　　　　　Step 6:　　If $a\,(i, j) <$ TH

　　　　　　　　　　　$a\,(i, j) = [((a\,(i, j) - \text{min\_}a)/(\text{TH} - \text{min\_}a)) * \text{NTH}]$

　　　　　　　　　　　Else

　　　　　　　　　　　$a\,(i, j) = [((a\,(i, j) - \text{TH})/(\text{max\_}a - \text{TH})) * (255 - \text{NTH}) + \text{NTH}$

<div style="text-align:center">End If</div>
<div style="text-align:center">End For</div>
<div style="text-align:center">End For</div>

Step 7:   *Display* Contrast Stretched Image *a*

End

These output contrast stretched images are further assessed to check their enhancement using parameters like measure of enhancement using entropy (EMEE), and measure of enhancement (EME) [13–16].

2.2 Color-Based Segmentation Using K-Means Clustering

For segmentation of these enhanced blood smear images on the basis of color, k-means clustering is used. K-means clustering is a type of iterative hard clustering. Hard clustering refers to the clustering technique in which a data item from the input data set belongs to only one cluster [6]. K-means is iterative in nature because the centroid value is updated in each step to perform clustering. A pictorial illustration of how an input data set is divided into two clusters based on colors and centroid is shown in Fig. 1.

In K-means clustering, K represents the number of clusters which are defined as inputs. This clustering technique divides the data into K clusters. The clustering depends on square of Euclidean distance which is the distance between a data point and the centroid of cluster given by Eq. (1).

$$D = (x_1 - c_1)^2 + (x_2 - c_2)^2 \tag{1}$$

where, D is the squared Euclidean distance between data points x_1 and x_2 and their centroids c_1 and c_2. This further gives rise to clustering error given in Eq. (2) [18].

$$E = \sum_{i=1}^{N} \sum_{j=1}^{K} (x_i - c_K)^2 \tag{2}$$

Fig. 1 Illustration showing **a** Input data and **b** Two clusters of data divided on the basis of color and centroid [17]

(a)   (b)
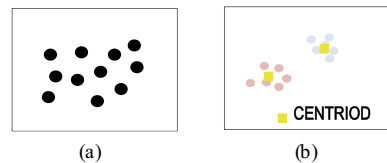
Here, K is the number of clusters, C_K is the centroid, x_i is the input data set and N and K are the number of elements in dataset and number of clusters, respectively. With the help of Euclidean distance and clustering error given in Eqs. (1) and Eq. (2) respectively, the centroids are updated and the dataset is divided into clusters of similar data. The proposed approach of color-based clustering divides the data into clusters of similar colors. The number of clusters are set to two to segment only the most prominent color which is nucleus in the case of these blood smear images. The nucleus hence obtained from this segmentation can be used to gather color information for healthy and malignant nucleus. This segmented image is further binarized to assess the performance of segmentation. The segmentation is evaluated using overlap ratio (OR) [15] with reference to the ground truth. The procedural steps for the proposed approach of color-based segmentation of enhanced blood smear images is depicted in Algorithm 2.

Algorithm 2: Procedural Steps for Color-Based Segmentation Using K-Means Clustering

Begin

Step 1:	*Input* Contrast Enhanced Image as *a*	
Step 2:	*Initialize*	
	Number of Clusters, $K \leftarrow 2$	
	$x_{(i, j)} \leftarrow a$	
Step 3:	*Read centroid* c_i	
Step 4:	*Calculate* Euclidean Distance between x and c using Eq. (1)	
Step 5:	*Calculate* error using Eq. (2)	
Step 6:	*Calculate* new centroid c_l	
Step 7:	If $c_i = c_l$	
	Display Segmented Cluster	
	Else	
	Repeat from Step 3	
	End If	

End

3   Experimental Results and Analysis

3.1   Dataset

The dataset to perform this experiment on blood smear images, the dataset is obtained from the online repository of ASH [8], for multicell images and cancer imaging archive [9], for single cell images. These images are both of cancerous and non-cancerous blood smears. These images acquired by a high-resolution camera are of different dimensions, i.e., 400×400 and 720×960, respectively. To perform color-based segmentation, one image was chosen from each dataset. The proposed methodology was applied on both the images and the results were recorded.

3.2   Simulation Results and Analysis

Two images chosen earlier from the dataset were used to perform contrast stretching using DCA and color-based segmentation using K-means clustering algorithm [5, 18]. The efficacy of contrast stretched image using DCA was evaluated using image quality assessment (IQA) metrics like EME and EMEE. While the efficacy of segmentation technique is evaluated using overlap ratio (OR).

The difference between the intensity of images (a) and (b) is visually perceptible as shown in Fig. 2. As we can see in images (b) the contrast of nucleus is increased and this color of nucleus is an important parameter used by hematologists to check for malignancy. Further, the color segmented images in (c) gives us the results of successful color-based segmentation of nucleus. These enhanced and segmented images are also evaluated qualitatively. The image quality of contrast stretched images are given in Table 1.

On observing the values of both EME [13] and EMEE [14] for both the input and contrast enhanced image, we observe a significant increase in the IQA metrics. This increase shows that the enhanced or the contrast stretched image is of high quality in terms of texture and contrast when compared with the input image. Likewise, the color segmented image is done evaluated using overlap ratio (OR) [15] given in Table 2.

The value of OR reaches to 1 in both the images which indicates perfect overlapping with the ground truth image [15]. On comparing the results of color-based segmented image with the enhanced image visually, we observe that the nucleus is segmented successfully. This result is also verified with the OR. Hence, visually and qualitatively the performance of segmentation technique is evaluated and is found to be successful. Therefore, this approach of color-based segmentation of blood smear images gave favorable outcomes.
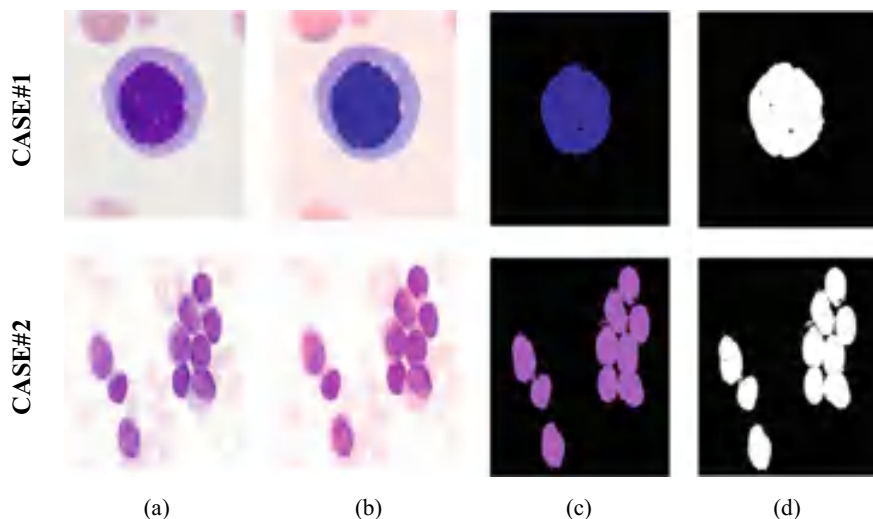
Fig. 2 Experimental results of images showing: **a** Input blood smear images, **b** Contrast stretched/contrast enhanced image using DCA, **c** Segmented image with RGB mask and **d** Segmented image with binary mask

Table 1 IQA metrics for input and enhanced microscopic images

	Single cell image (CASE#1)		Multi cell image (CASE#2)	
Metric	Input image	Enhanced microscopic image	Input image	Enhanced microscopic image
EME	2.309	3.558	3.99	4.523
EMEE	0.0305	0.4590	0.250	0.599

Table 2 Performance evaluation of segmentation technique

	Single cell image (CASE#1)	Multicell image (CASE#2)
Metric	**Nucleus**	**Nucleus**
OR	0.991	0.903

4 Conclusion

This paper presents an approach to perform color-based segmentation on microscopy images. In the first place, DCA was used to stretch the intensity of the input image. This helped to increase the contrast of the cell components, as we can see the contrast of nucleus is enhanced. These images were evaluated using EME and EMEE, both showing significant increase in their values. Further, for color-based segmentation k-means clustering was used. In this approach only the nucleus is segmented so the

number of clusters were initialized to 2. The number of clusters can be increased depending on how many segmented images are required. The resultant cluster or the segmented image gives the RGB mask of the nucleus. The results of segmentation were thereafter appraised using OR, showing successful color-based segmentation. This segmented image can further give information regarding the color variations in different regions of nucleus.

References

1. Jagadev P, Virani HG (2018) Detection of Leukemia and its types using image processing and machine learning. In: Proceedings of international conference on trends in electronics and informatics (ICTEI 2017). IEEE, Tirunelveli, India, pp 522–526
2. Rezatofighi SH, Zadeh HS (2011) Automatic recognition of five types of white blood cells in peripheral blood. Comput Med Imaging Graph 35(4):333–343
3. Bhateja V, Urooj S, Mehrotra R, Verma R, Lay-Ekuakilli A, Verma VD (2013) A composite wavelets and morphology approach for ECG noise filtering. In: Maji P, Ghosh A, Murty MN, Ghosh K, Pal SK (eds) Pattern recognition and machine intelligence. International conference on pattern recognition and machine intelligence, vol 8251. Springer, Heidelberg, pp 361–366
4. Raj A, Alankrita SA, Bhateja V (2011) Computer aided detection of brain tumor in magnetic resonance images. Int J Eng Technol 3(5):523–532
5. Harun NH, Bakar JA, Hambali HA, Khair NM, Mashor MY, Hassan R (2018) Fusion noise—removal technique with modified algorithm for robust segmentation of acute Leukemia cell images. Int J Adv Intell Inform 4(3):202–211
6. Su J, Liu S, Song J (2017) A segmentation method based on HMRF for the aided diagnosis of acute myeloid Leukemia. Comput Methods Progr Biomed 152(7):115–123
7. Basavaraju HT, Aradhya VNM, Pavithra MS, Guru DS, Bhateja V (2021) Arbitrary oriented multilingual text detection and segmentation using level set and Gaussian mixture model. Evol Intel 14:881–894
8. Acharya V, Ravi V, Pham TD, Chakraborty C (2021) Peripheral blood smear analysis using automated computer-aided diagnosis system to identify acute myeloid Leukemia. IEEE Trans Eng Manag: 1–14
9. Dasariraju S, Huo M, McCalla S (2020) Detection and classification of immature leukocytes for diagnosis of acute myeloid Leukemia using random forest algorithm. Bioengineering 7(4):120–131
10. Eckardt JN, Middeke JM et al (2022) Deep learning detects acute myeloid Leukemia and predicts NPM1 mutation status from bone marrow smears. Leukemia 36:111–118
11. Gupta S et al (2022) Analysis of blood smear images using dark contrast algorithm and morphological filters. In: 10th international conference on frontiers of intelligent computing: theory and applications. Springer, India
12. Verma S et al (2022) Segmentation of blood smear images using dark contrast algorithm and K-medoid clustering. In: 7th international conference on microelectronics, electromagnetics and telecommunications. Springer, India
13. Trivedi M, Jaiswal A, Bhateja V (2013) A no-reference image quality index for contrast and sharpness measurement. In: 3rd IEEE international advance computing conference (IACC). IEEE, India, pp 1234–1239
14. Prajapati P, Narmawala Z, Darji NP, Moorthi SM, Ramakrishnan R (2015) Evaluation of perceptual contrast and sharpness measures for meteorological satellite images. In: Soni AK, Lobiyal DK (eds) 3rd international conference on recent trends in computing (ICRTC), Procedia computer science, vol 57. Springer, India, pp 17–24

15. Kumar SN, Lenin Fred A, Ajay Kumar H, Sebastin Varghese P (2018) Performance metric evaluation of segmentation algorithms for gold standard medical images. In: Sa P, Bakshi S, Hatzilygeroudis I, Sahoo M (eds) Recent findings in intelligent computing techniques. Advances in intelligent systems and computing, vol 709. Springer, Singapore
16. Gupta P, Tripathi N, Bhateja V (2013) Multiple distortion pooling image quality assessment. Int J Converg Comput 1(1):60–72
17. Rokaha B, Ghale DP, Gautam BP (2018) Enhancement of supermarket business and market plan by using hierarchical clustering and association mining technique. In: Proceedings of international conference on networking and network applications. IEEE, China, pp 384–389
18. Aristidis L, Vlassis N, Veerbeek JJ (2003) The global k-means clustering algorithm. Pattern Recogn 36(2):451–461

A Novel Dynamic Latch Comparator Design and Analysis for ADCs

Kasi Bandla (ID)**, Atharva Dinakar, and Dipankar Pal** (ID)

Abstract Comparators are used in data converters, sense amplifiers, RFID and data receivers. This paper presents a novel comparator topology for ADC design, in which cascode transistors are stacked on the top of differential input section. In terms of speed, offset voltage, power dissipation and kickback noise, this design improves the comparator's overall performance. The design and simulations are carried out on standard UMC 180 nm technology, for 100 MHz clock, at 1.8 V supply using Cadence Virtuoso EDA tool for the sake of reasonable comparison.

Keywords Dynamic latch · Power delay product (PDP) · Kickback noise · Offset voltage · Low power

1 Introduction

In recent years, an analog-to-digital converter (ADC) has become a critical component in the integrated circuit (IC) design industry. For portable electronic systems including wireless communication devices, consumer electronics and medical equipment, high-speed, low-power circuits are essential. ADCs, on the other hand, offer a number of characteristics that make them more suitable for today's IC industries [1], including smaller transistor sizes, lower power dissipation and high speed. A new ADC with a lower supply defined by transistor size is necessary for the aforementioned features.

K. Bandla (✉) · A. Dinakar · D. Pal
Department of Electrical and Electronics Engineering, BITS-Pilani, KK Birla Goa Campus, South Goa, Goa 403726, India
e-mail: p20170429@goa.bits-pilani.ac.in

A. Dinakar
e-mail: f20170218g@alumni.bits-pilani.ac.in

D. Pal
e-mail: dipankarp@goa.bits-pilani.ac.in

In the design of the ADCs, the comparator is a key component, since comparator affects accuracy and performance of ADC [2]. Radio-frequency identification (RFID), data receivers, memory circuits, switching power regulators and other applications, all are required high-speed, low-power and high-resolution comparators [3, 4]. A low input voltage must be converted to a high output value using high-performance comparators. High gain and bandwidth are therefore required for a faster and more accurate comparator [5, 6]. A few examples of high-speed comparators are the multistage open-loop comparator, the regenerative latch comparator and the preamplifier latch comparator [7]. CMOS dynamic latch comparators are particularly well-liked in many applications due to their rapid speed, low power consumption, high input impedance and full-swing output. Additionally, latch-type comparators with a positive feedback system might offer a higher gain in regeneration mode. However, constructing a latch comparator for low-voltage operations can reduce dynamic input ranges and associated differential processes as well as increase power indulgences [2, 8], especially in rail-to-rail operations. On the other hand, random offset voltages brought on by device mismatches and random noise may reduce the accuracy of a latch-type comparator. Therefore, reducing offset voltages is one of the most important design factors for the dynamic latched comparator [9, 10]. Preamplifiers are typically used before regenerative latch stages to reduce offset voltage because they have the potential to amplify weak input signals into strong output signals, which can help them to overcome kickback noise and latch offset voltage [6, 11]. It does, however, dissipate a lot of static power. Consequently, a dynamic latch comparator without a preamplifier is highly desirable for low-power designs. Energy conservation is accomplished via the Charge Sharing Dynamic Latch Comparator (CSDLC) [8]. But the average power consumption of CSDLC is a little bit higher due to both output nodes transition at positive and negative clock (CLK) edges. The CSDLC cannot generate swing in the rail output [1, 11]. A Modified Strong-Arm Dynamic Latch Comparator (MSADLC) is being created to address the problems with CSDLC. However, there is an issue with offset voltage and speed. These problems are overcome by the SR Latch SADLC. However, it does it with a high real estate and greater power dissipation [12]. Therefore, in this paper, a novel comparator design is proposed to address these concerns and attain optimal performance.

The remaining sections are arranged as follows: The design methodology, principle of operation, and parameter analysis are described in Sect. 2. The simulation results of the proposed design are presented in Sect. 3 along with comparisons to other potential designs from existing literature. Section 4 concludes the paper.

Fig. 1 Basic symbol of comparator
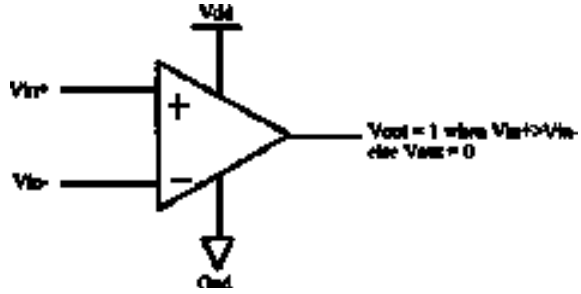
2 Design Methodology

2.1 Basic Operation

Two instantaneous analog voltages are compared by a comparator, which outputs a digital value, "1" or "0," to indicate a difference in the input's polarity. A typical comparator symbol is depicted in Fig. 1.

2.2 Proposed Comparator Topology

A novel dynamic latch comparator (Fig. 2) consists of a differential pair input stage and cascode transistors, which are stacked and sandwiched in between the inverters latch. In comparison to traditional dynamic latch comparators, the suggested design may provide high-speed, low-power dissipation and low offset at low supply voltages.

The schematic shown in Fig. 2 is having following features as compared to MSADLC.

1. Cascode transistors (M12 and M13) stacked on top of the input differential section.
2. The differential amplifier is sandwiched between the inverter latch pairs, along with cascode transistors.

As a result, the cascode transistors (M12 and M13) increase the comparator's gain, which leads to improve speed. So, the output response settles faster to either logic "0" or "1" based on the differential input voltages.
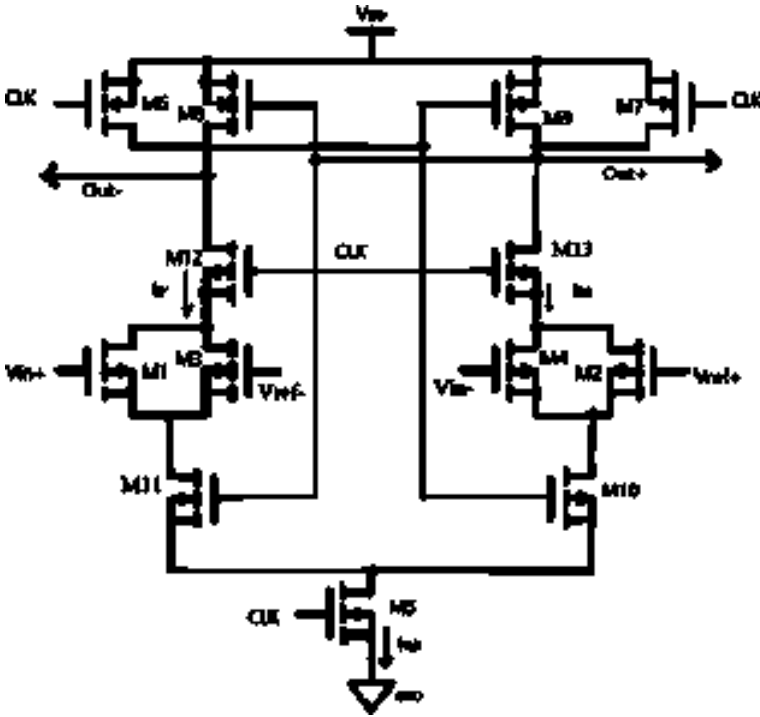
Fig. 2 Schematic of proposed novel comparator

2.3 Principle of Operation

Figure 2 operates in two phases:

1. **Pre-Charge Phase**: M12, M13 and M5 are turned OFF when the CLK signal gets close to zero. The current path is so blocked. So, the transistors M6 and M7 gets ON and sets both the output differential nodes, out+ and out− to VDD.
2. **Comparison Phase**: The cascode transistors M12, M13 and the tail transistor M5 are turn ON when the CLK signal is high. M6 and M7 are OFF and makes the differential output nodes out+ and out− to be separated from VDD. The cross-coupled inverter pairs M8–M9, M10–M11 re-generatively amplify the difference between V_{in} and V_{ref} and choose which output goes to VDD and which to GND based on that difference.

2.4 Transistor Sizing

The transistors needed to simulate the suggested circuit and compare it to other competing circuits for benchmarking should be as little as possible in order to meet

Table 1 Aspect ratios of transistors

Name of transistor	W/L
M1–M4	720 nm/180 nm
M5	720 nm/180 nm
M6–M9	1.13 µm/180 nm
M10–M11	720 nm/180 nm
M12–M13	720 nm/180 nm

the requirements for quick speed and little parasitic capacitance. The suggested circuit is therefore built utilizing digital scaling techniques. In Table 1, the optimal transistor sizes for the UMC 180 nm model library are presented.

2.5 Design Parameters

The following parameters are used to determine the comparator's performance:

(a) **Delay Analysis**

Two components, $t0$ and $t$latch, make up the comparator delay, which is depicted in Fig. 2. The initial term, $t0$, represents the period of time needed to discharge the output node voltage (i.e., load capacitance CL) before the first pMOS transistor switches ON. The tail transistor M5 and cascode transistors M12 and M13 are ON if CLK is equal to 1. If $|V_{in+}| > |V_{in}|$ and CLK $= 1$, transistors M1, M3 cause faster discharge of Vout− by turning ON M12. On this foundation, the following formula can be used to calculate the delay:

$$t_0 = \frac{C_L.V_{\text{thp}}}{I_{D1}} \cong 2.\frac{C_L V_{\text{thp}}}{I_{\text{tail}}}. \tag{1}$$

In (1), one can assume that the drain current is constant for small input differential voltage (ΔV_{in}) and its tail current proportion would be halved.

The second term t_{latch} refers to two cross-coupled inverters' combined latching delay, which aids in achieving absolute rail-to-rail voltage at the output. As a result, the latch evaluation time is calculated as follows:

$$t_{\text{latch}} = \frac{C_L}{g_{m(\text{eff})}} . \ln\left(\frac{\Delta V_{\text{out}}}{\Delta V_o}\right) \cong \frac{C_L}{g_{m(\text{eff})}} . \ln\left(\frac{V_{DD}/2}{\Delta V_o}\right), \tag{2}$$

where $g_{m(\text{eff})}$ is cross-coupled inverters' effective trans-conductance [1]. The initial difference voltage at the commencement of the regeneration phase (i.e., at $t = t_0$) affects the t_{latch} in a logarithmic way. Using (1), the initial output voltage difference ΔV_o can be determined as follows:

$$\Delta V_o = \left| V_{\text{out}(t=to)}^+ - V_{\text{out}(t=t0)}^- \right| \Rightarrow \Delta V_o = \left| V_{\text{thp}} \right| - \frac{I_{D2}.t_o}{C_L}. \tag{3}$$

Replacing ΔV_o [from (3)] into (2), the total delay can be determined by using the value of from (1) and the formula presented below, as seen in (4).

$$t_{\text{total}} = t_0 + t_{\text{latch}} \Rightarrow t_{\text{total}} = 2.\frac{C_L V_{\text{thp}}}{I_{\text{tail}}} + \frac{C_L}{g_{m(\text{eff})}}.\ln\left(\frac{V_{DD}/2}{\Delta V_o}\right). \tag{4}$$

(b) **Power Dissipation**

The primary contributors to the typical power dissipation by the supply voltage over a comparable time period are

$$P_{\text{Avg}} = \frac{1}{T}\int_0^T V_{DD} I_D \mathrm{d}t = f_{\text{clk}} V_{DD}\int_0^T I_{D1,2}\mathrm{d}t, \tag{5}$$

where $I_{D1,2}$ is the drain current flows from the supply voltage (V_{DD}) and f_{clk} is clock frequency of the comparator.

(c) **Offset Voltage**

The offset is the input error range that the comparator must fall inside in order to detect the aforementioned, exceedingly small voltage difference. As a result, the resolution and speed of the comparator are limited [13]. There are no offset canceling strategies discussed in this study. Due to MOS device mismatches, however, there is a trade-off between speed and accuracy. Offset's impact can be reduced, but not entirely avoided. Total offset voltage is dependent on threshold voltage mismatch ΔV_T, load resistance ΔR_L and transistor size $\Delta\beta$ (i.e., W/L ratios), and for the corresponding values (V_T, R_L and β), it is given by (6)

$$V_{\text{os}} = \Delta V_T + \frac{V_{\text{gs}} - V_T}{2}\left[\frac{\Delta R_L}{R} + \frac{\Delta\beta}{\beta}\right]. \tag{6}$$

In Eq. (6) the V_{os} is driven by $\Delta\beta$, which is the discrepancy between the overdrive voltage ($V_{GS} - V_T$) and transistor size. Equation (6) shows that the offset voltage reduces along with the common mode voltage ($V_{GS\text{-Low}}$).

(d) **Kickback Noise**

The analog input signal is converted back into a full-scale digital level via a positive feedback mechanism. Large voltage changes in the internal nodes are connected to the input, causing the input voltage to fluctuate; this phenomenon is typically referred to as kickback noise [14, 15].

3 Results and Discussion

The proposed design and simulations are carried out on the Cadence Virtuoso platform with UMC 180 nm Technology at a supply VDD of 1.8 V. The clock speed is set to 100 MHz during the process. For an accurate comparison with the optimal transistor sizes, all topologies are simulated on the same platform. The suggested design's typical screen images of simulated waveforms are illustrated in Figs. 3, 4 and 5. To prevent redundancy and for brevity, the same is omitted for the other topologies. Table 2 illustrates the performance investigations, which is shown graphically in Fig. 6.
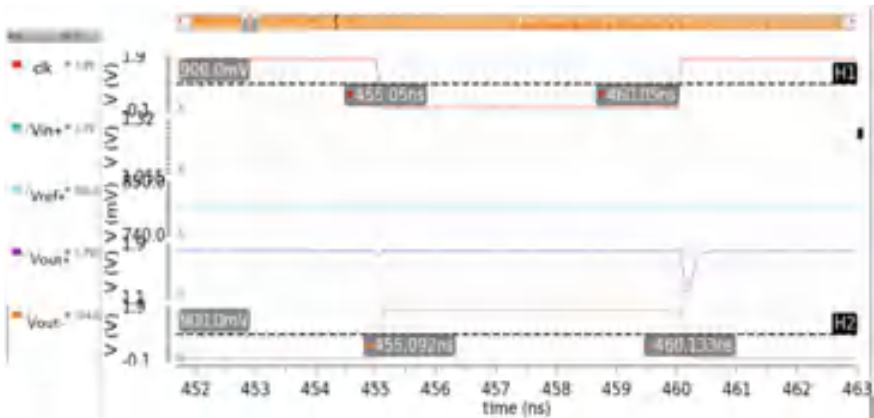
Fig. 3 Delay measurement
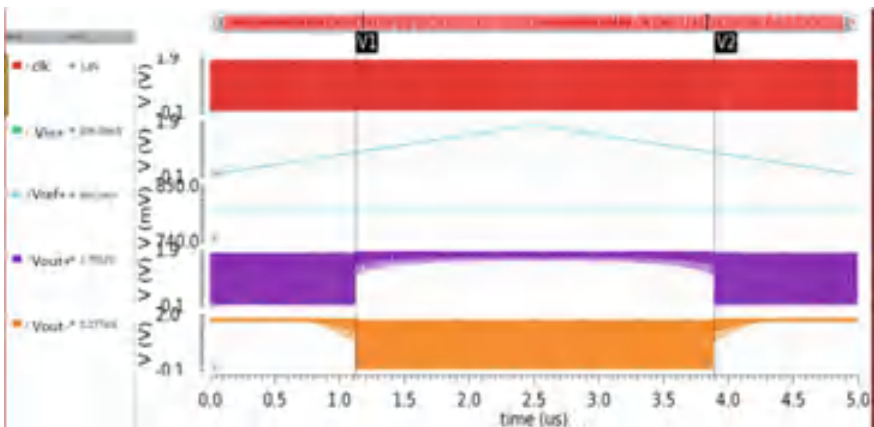
Fig. 4 Transient response to ramp input
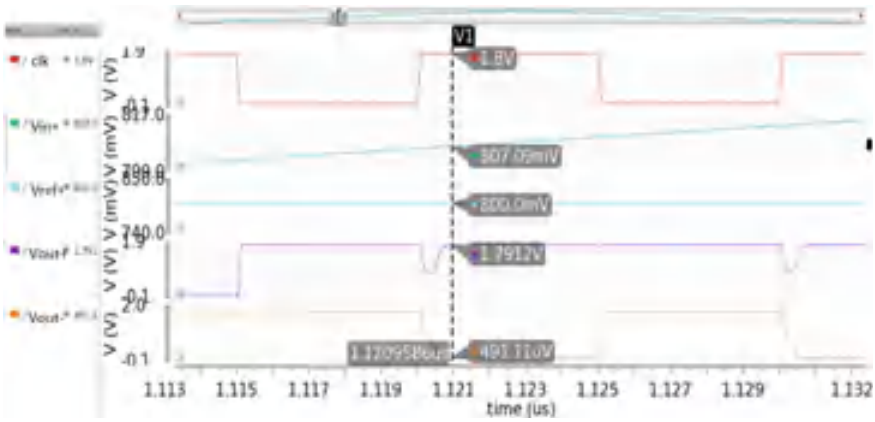
Fig. 5 Measurement of offset as input rises

Table 2 Performance comparison

Design topology	Delay (pS)	Power (μW)	PDP (fJ)	Offset voltage (mV)	Kickback noise (mV)
CSDLC [9]	178.1	18	3.2	63	21.62
SADLC [13]	92.75	4.82	0.45	6	18.27
MSADLC [1]	93.4	4.72	0.44	6	18.41
SR-SADLC [12]	72	18	1.29	3.06	11.3
Proposed Design	62.82	4.08	0.25	2.70	7.58
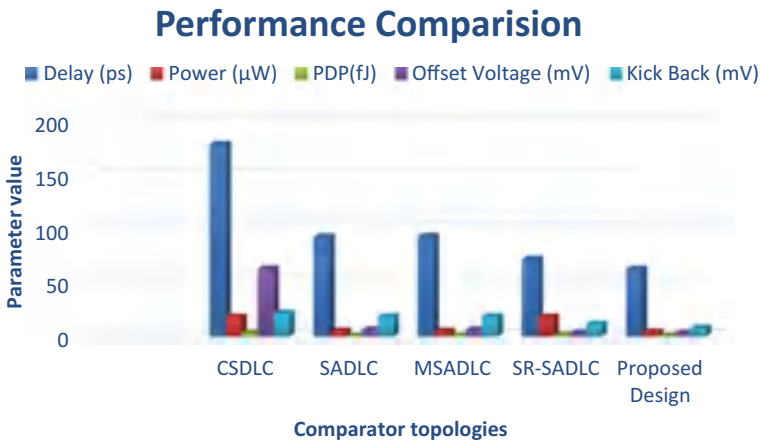
Fig. 6 Comparison in graphical representation

3.1 Parameter Measurement

The reference voltage V_{ref+} is set at 800 mV, CLK = 100 MHz and V_{DD} = 1.8 V.

- Power: The V_{in+} input voltage is varied from 0 to 1.8 V and back again. The power dissipation is averaged in this stage to get average power.
- Average Delay: The positive input, V_{in+}, remained at 1 V. The time to charge and discharge the output nodes is calculated with respect to CLK signal change. By averaging these two, the average delay is obtained.
- Kickback Noise: The input is scaled up to 1 V after being stalled at 600 mV for a while. The input's source resistance is 1 kΩ, and the excess voltage collected across this resistance is calculated. Similarly, the negative input is evaluated and averaged to obtain kickback noise.
- Offset Voltage: The input voltage V_{in+} is alternated between 0 and 1.8 V. The variance between V_{in+} and V_{ref+} is the offset voltage, which appears as the input rises at the time the output switching is measured.

4 Conclusion

On the basis of SADLC, this paper presents a novel comparator design and analysis. The proposed design is implemented using Cadence Virtuoso EDA tool using UMC 180 nm standard CMOS process at 1.8 V supply. In terms of speed, kickback noise and offset voltage, adding a cascode transistor proved to be beneficial. It achieves significant improvement in speed (33%), PDP (43% improvement) as compared to MSADLC [1]. Even though offset cancelation techniques are not used, it offers a very low offset (55% reduction) and also kickback noise (33% reduction). It is evident from Table 2 and Fig. 6 that the proposed topology outperforms the conventional topologies. In compared to SR-SADLC [12], the suggested comparator not only outperforms but also area efficient. As a result, the proposed topology is recommended for ADC systems that require high-speed, high accuracy and low power consumption. Next, the proposed design performance will be evaluated by doing post-layout simulation followed by DRC, LVS and parasitic extraction for the design of ADC.

References

1. Bandla K, Harikrishnan A, Pal D (2020) Design of low power, high speed, low offset and area efficient dynamic-latch comparator for SAR-ADC. In: 2020 international conference on innovative trends in communication and computer engineering (ITCE), Aswan, Egypt, pp 299–302. https://doi.org/10.1109/ITCE48509.2020.9047792
2. Goll B, Zimmermann H (1892) A comparator with reduced delay time in 65-nm CMOS for supply voltages down to 0.65. IEEE Trans Circuits Syst II Exp Briefs 56(11):810–814; Clerk

Maxwell J (1892) A treatise on electricity and magnetism, 3rd ed., vol 2. Clarendon, Oxford, pp 68–73

3. Khosrov DS (2011) An improved low offset latch comparator for high-speed ADCs. Analog Integr Circuit Signal Process 66:205–212

4. Razavi B, Wooley BA (1992) Design techniques for high-speed, high-resolution comparators. IEEE J Solid-State Circuits 27(12):1916–1926

5. Sandeep K, Arya N (2011) A comparative study of dynamic latch comparator. In: Proceedings of the international conference on advanced computing. Communication and networks, 1248–1252

6. Hong HC, Lee GM (2007) A 65-fJ/conversion-step 0.9-V 200-kS/s rail-to-rail 8-bit successive approximation ADC. IEEE J Sol-State Circuits 42(10):2161–2168

7. HeungJun J, Yong-Bin K (2012) A novel low-power, low-offset, and high-speed CMOS dynamic latched comparator. Analog Integr Circuit Signal Process 70:337–346

8. Bekal A, Joshi R et al (2015) An improved dynamic latch based comparator for 8-bit asynchronous SAR ADC. In: 2015 IEEE computer society annual symposium on VLSI, pp 178–182

9. Halim ISA (2011) Low power CMOS charge sharing dynamic latch comparator using 0.18 μm technology. IEEE-RSM 2011

10. Razavi B (2005) The strong ARM latch—a circuit for all seasons. IEEE Sol-State Circuits Mag 7(2):12–17. https://doi.org/10.1109/mssc.2015.2418155

11. Halim ISA, Abidin NANBZ, Rahim AAA (2011) Low power CMOS charge sharing dynamic latch comparator using 0.18 μm technology. In: 2011 IEEE regional symposium on micro and nano electronics, Kota Kinabalu, pp 156–160.https://doi.org/10.1109/RSM.2011.6088314

12. Bandla K, Harikrishnan A, Sethi S, Pal D (2020) Design of high speed and low offset SR latch based dynamic comparator. In: 2020 IEEE region 10 symposium (TENSYMP), Dhaka, Bangladesh, 5–7 June 2020, pp 52–55.https://doi.org/10.1109/TENSYMP50017.2020.9230938

13. He J, Zhan S, Chen D (2009) Analyses of static and dynamic random offset voltages in dynamic comparators. IEEE Trans Circuits Syst-I Regular Pap 56(5)

14. Figueiredo PM, Vital JC (2006) Kickback noise reduction techniques for CMOS latched comparator. IEEE Trans Circuits Syst 53(7):541–545

15. Almansouri A, Alturki A, Alshehri A, Al-Attar T, Fariborzi H (2017) Improved strong ARM latch comparator: design, analysis and performance evaluation. In: 2017 13th conference on Ph.D. research in microelectronics and electronics (PRIME). Available https://doi.org/10.1109/PRIME.2017.7974114

A Dual Stopband Frequency Selective Surface at 2.57 and 5.40 GHz

Athira G. Krishna and **Lalithendra Kurra**

Abstract A dual stopband single layer Frequency Selective Surface (FSS), covering the Wi-Fi frequency bands for mobile communication applications, is proposed in this paper. It is meant to stop the desired Wireless Fidelity frequencies and pass signals that lie out of the Wi-Fi bands. It employs concentric loops of square-shaped patches for the unit cell. The presented FSS occupies—20 dB bandwidths of 378 MHz (2380–2758 MHz) and 474 MHz (5178–5652 MHz), at frequencies 2.57 and 5.4 GHz, correspondingly. The FSS unit cell structure has a size of 14.7 mm × 14.7 mm (0.126 λo × 0.126 λo), where λo signifies the free-space wavelength of the first resonant frequency at 2.57 GHz. The purported design can be tweaked into the preferred resonance frequencies by varying the dimensions of each square ring resonator. All simulation parts are modeled by using Computer Simulation Technology (CST) Microwave Studio. Key factors for the proposed design are ease of manufacturing, a low reliance on the angle of incidence of plane wave, and the polarization direction.

Keywords Frequency Selective Surface · FSS · Dual band · Stopband

1 Introduction

As wireless networks have become widely fanned out, and as they are broadcasting networks, they can cause electromagnetic interference. Very sensitive devices used in airports or military camps and even in hospitals can interfere with the signals used. This can be bad for security reasons as well as can degrade the system's performance [1]. Though a radiation-free environment can be made possible by employing techniques like jammers and shielding paints, they are less preferred. Jammers need power supply while shielding paints, though environmentally friendly,

A. G. Krishna (✉) · L. Kurra
Department of Electronics and Communication Engineering, CVR College of Engineering, Hyderabad, Telangana, India
e-mail: athirashad2005@gmail.com

can block all other microwave signals required for other types of communications [2, 3]. Wi-Fi standard 802.11n works in both the 2.40 GHz (2401–2495 MHz) and 5 GHz (5180–5825 MHz) Wi-Fi frequency bands. Most routers have made the switch from using single band to dual band to choose between the Wi-Fi frequency bands. In the 2.4 GHz band, the lower frequencies that are transmitted can easily penetrate through walls, doors, and windows which are the main areas that cause signal leakage. Controlling the signal coverage of Wi-Fi is an effective way to avoid eavesdropping. Reinforced walls or walls with metallic shielding can be used to lessen the Wi-Fi coverage. However, these cannot single out specific frequency bands.

Remodeling the existing walls of buildings with band stop Frequency Selective Surface (FSS) can be an effective answer for such concerns. FSS can block or allow specific frequency bands as it can be well-thought-out as man-made electric wall or magnetic wall that depends on frequency. These structures find great applications as radomes, spatial selective filters, reflectors and absorbers, as well as electromagnetic band gap materials, electromagnetic shielding [4]. Structurally, FSS is a periodic surface consisting of 2D arrays of identical elements arranged on a dielectric substrate.

A plane wave incident on the FSS surface will be allowed to pass through completely or partially (passband) or will be reflected completely or partially (stopband). This behavior will be decided by the nature of the array elements, and this happens when the frequency of resonance of the FSS is completely compatible with the frequency of the incident EM plane wave. The abrupt expansion of multifunctional modern communication systems has given rise to a vast requirement for FSS with multiband performance.

The proposed design in this paper mainly concentrates on dual-band FSS as a part of the research on multiband FSS. A multitude of structures has been suggested in previous literature by others for different applications. Distinct shapes of the elements (or the complementary apertures) produce various resonant characteristics. Loop elements are promising candidates when designing FSS structures due to their band separation, angular stability, impedance bandwidth and low cross polarization [4, 5]. Similar application has been implemented using double-loop elements as the FSS surface [6, 7].

In [6], the author recommends a unit cell that is anchor-shaped. The structure works as a compact surface with two-band stop behavior at frequencies 2.4 and 5.0 GHz. The resulting bandwidths, according to their work, have been found to be best to prevent interferences to the WLAN. The material of the substrate is F4B-2 with a thickness of 1.5 mm. They were able to achieve -42 dB at the rejection frequency of 2.4 GHz and -40 dB at the rejection frequency of 5.0 GHz. It is claimed to have the distinction of having two resonance frequencies that can be adjusted individually in a certain frequency range. This has been done by varying the values of a single parameter 'g' which denotes the gap between two adjacent anchor-loaded loops.

A FSS structure is proposed in [7], with dual passbands for WLAN application. The design of the FSS structure consists of a rectangle shaped loop wire on the top metal plane. Its complementary is etched on the bottom plane. The paper indicates how, according to the principle of duality, a complementary structure provides a

performance that is exactly the opposite to the original one. The original band stop frequency response will be replaced by a bandpass one. A meandered structure is utilized to miniaturize the FSS. Meandering will increase the equivalent inductance while it will bring down the equivalent capacitance at the same time. The operating frequency is also reduced with the element size.

In [8], M.R. Chaharmir, J. Ethier, J. Shaker point out that the designs put forward in [5, 6] are wideband. This blocks part of the band outside Wi-Fi such as the LTE band (2300–2400 MHz for 4G-LTE). The FSS design ascertains how well the Wi-Fi signals can be blocked and at the same time allows other portions of the RF spectrum to pass undisturbed. This is specified for the lower band (2.45 GHz) because it is very close to the LTE mobile band. The FSS design in their work consists of a novel narrowband FSS element with a combination of open meandered loops and closed loops. A 2.45 GHz narrow band resonance is achieved by using the meandered open loop. The cell elements have been placed on the surface of a 50-μ PET substrate. In one of the most recent works [9], on square-loop FSS for band stop characteristics, the FSS structure works to eliminate unwanted radiation signal at 5.8 GHz.

A comparison between the band stop performance of square, octogen, and hexagon loops is carried out. The result of this work highlights the fact that the square-loop FSS structure has the better attenuation compared to the other two structures. The square-loop structure uses a FR-4 dielectric substrate with a thickness or height of 1.6 mm. The proposed work uses the advantage of square loop to get stop band characteristics at two bands respectively, for outer loop and the inner loop with good attenuation.

The rest of this paper is structured as follows. Section 2 presents the Frequency Selective Structure basic design considerations and describes the reason for selecting the square-loop element as the basic unit cell shape. Section 3 describes the proposed design parameters of the unit cell. Simulated results and the parametric analysis study of the proposed FSS and its discussion are given in Sect. 4. Section 5 provides concluding remarks.

2 FSS Basic Design Considerations

In this paper, the double stopband characteristic is achieved by using a double square metal loop for the multi-resonance. The square-loop configuration is chosen as the basic unit cell shape of the proposed design due to its symmetry. It has been researched and investigated by Brian Monacelli, Ben A. Munk, and others in [10] that the symmetry of the square loop is a boon in fabrication, as straight lines can be reproduced with high fidelity lithographically. The occurrence of grating lobes in the spectral response can be avoided due to the elements being tightly packed in this symmetry. A square loop, in addition to being simple in structure, and easy to fabricate, is kind of oblivious to polarization in the vertical/horizontal directions. Moreover, a good element for FSS design should be small in terms of wavelength. Leading candidates with this feature are seen to be members of group 2, namely, the

simple circular, square, and hexagon rings [4, 5]. Generally, the bandwidth of any FSS can be altered considerably by adjusting the spacing 's' between the periodic elements. Loop-type elements allow controlling the bandwidth by varying the FSS element geometry parameters without the problem of grating lobes setting in.

FSSs are fundamentally designed to operate as series/parallel LC circuits. A series equivalent circuit gives band stop filter action while a parallel equivalent circuit gives bandpass filter performance. The resonance frequency (f_r) of FSS is given by $f_r = 1/2\pi\sqrt{LC}$. Therefore, reducing the resonance frequency is equivalent to increasing the L or C of the unit cell. When perimeter of the square metal loop becomes an integer multiple of the incident wavelength, the resonance takes place. For a square-loop unit cell, resonance will occur for perimeter values nearly equal to integer multiples of the incident wavelength. By bending or meandering the square element further, the effective physical size can be increased in the limited area, without altering the dimensions of the original unit cell. This leads to miniaturization of the structure as well as reduction of the resonance frequency [11]. Electromagnetic Interference Shielding is a prime topic of research in many related works experimenting with hexagon and cross shapes, yielding filters with one to more than five stopbands that have been investigated for their unit cell element types, dimensions, angular stability as well as polarisation independence [12–15].

The proposed FSS unit cell has a circuit analog since it has the substrate layer and the element structure. The incident radiation will motivate flow of current in the wire which creates inductance. Gaps of 0.9 mm (between the outer and inner metallic loops) and 0.5 mm (within the loops) create capacitance. The inductance and capacitance together are equivalent to a resonant series LC circuit. Such a resonant FSS can be treated as a band stop filter (an electric wall). The equivalent impedance, reflection, and transmission coefficients of the equivalent circuit in [11] are given as,

$$Z = j\omega L + 1/j\omega C = j(\omega L - 1/\omega C), \tag{1}$$

$$S_{11} = -1/2[Z/Z_0], \tag{2}$$

$$S_{21} = 2[Z/Z_0]/(2[Z/Z_0] + 1). \tag{3}$$

From the above equations, it can be assumed that the impedance of FSS will have a zero value at some certain frequency. The current will reduce to zero with zero impedance. This gives zero current, and thus, the wave will be incapable of transmitting through the surface. This gives the structure its band stop features.

3 Proposed Design Parameters

The measurement parameters of the proposed FSS unit cell are shown in Fig. 1. A FR-4 lossy substrate ($\varepsilon r = 4.3$) houses the double ring element. The substrate height is $h = 0.8$ mm. During the design of FSS, it is decisive to take into consideration the substrate dielectric properties. This is because the electri6cal characteristics and the dielectric material together can influence the shaping and steadying of the spectral curves. The resonance can be slightly disrupted in the presence of a dielectric substrate. In such circumstances, length of the loops is adjusted to correct the resonance. $\varepsilon_{\mathrm{reff}} = 0.5(\varepsilon_r + 1)$ gives the effective permittivity of the wall substrate [4, 16]. The resonance frequency reduces with the increase in the εr. This is due to the loading effect of the dielectric [16].

The next design parameter to reflect on is the square ring sizes. The rings are designed with material annealed copper. The external and internal square rings individually contribute 2.65 GHz, with attenuation of 38.85 dB, and 5.57 GHz at an attenuation of 34.83 dB, respectively. The outer square-loop element length and breadth are decided at 57.2 mm and 0.5 mm, respectively. This sets the fr at 2.65 GHz. The inner square-loop equivalent dimensions of 46 and 0.5 mm provision a fr at 5.57 GHz. When both outer and inner loops are combined, with separation between the outer and inner loops at $s2 = 0.9$ and the separation of the outer loop with the dielectric boundary $s1 = 0.2$, the resonance frequencies get adjusted to 2.57 GHz for an attenuation of 36.5 dB and 5.4 GHz at an attenuation of 34.18 db.

This can be traced back to effect of mutual coupling between the rings. The design delivers sufficient attenuation at the Wi-Fi operating bands from 2380 to 2758 MHz and 5178 MHz to 5652 MHz, with the bandwidth of 378 MHz and 474 MHz, respectively. Wi-Fi standard 802.11n (also known as Wi-Fi 4) operates in both the 2.4 GHz (2401–2495 MHz) and 5 GHz (5180–5825 MHz) Wi-Fi frequency
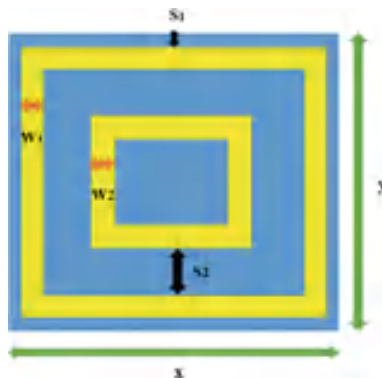
Fig. 1 FSS unit cell geometry ($x = 14.7$ mm, $y = 14.7$ mm, $S1 = 0.2$ mm, $S2 = 0.9$ mm, $w1 = 0.5$ mm, $w2 = 0.5$ mm)

bands. The band of frequencies from 2380 to 2758 MHz, and the frequency band from 5178 to 5652 MHz very appreciably includes the Wi-Fi4 2.4 and 5 GHz band.

4 Simulation Results

CST Microwave Studio software is employed to simulate the double square-loop FSS. Solver for frequency domain is preferred for imitating the repetitive resonant structure. As the FSS can be realized as an infinite periodic structure, it is considered good enough to simulate a single unit cell. A maximum attenuation of 36.5 dB and 34.18 dB is achieved at 2.57 GHz and 5.4 GHz, respectively.

The spatial filter unit cell was initially simulated with only the outer square loop as shown in Fig. 2a. The L and $w1$ of the outer ring element are adjusted to 57.2 and 0.5 mm, to adjust the f_r to 2.65 GHz. An attenuation of 38.9 dB was achieved with zero transmission characteristics as shown in Fig. 2b.

The FSS was then simulated with only the inner square loop as shown in Fig. 3a. The inner square loop when improved to a $L = 46$ mm and $w2 = 0.5$ mm promotes a high frequency of 5.57 GHz. An attenuation of 34.8 dB was achieved with a transmission as low as 0. 1 dB as shown in Fig. 3b.

When both outer and inner loops are combined in a single unit cell as shown in Fig. 4a, with the separation outer inner loops at $s2 = 0.9$ and the loop the dielectric boundary $s1 = 0.2$, the resonance frequencies get adjusted to 2.57 GHz for an attenuation of 36.5 dB and 5.4 GHz at an attenuation of 34.18 dB. This is by large due to mutual coupling between both loop elements.

(a)   (b)
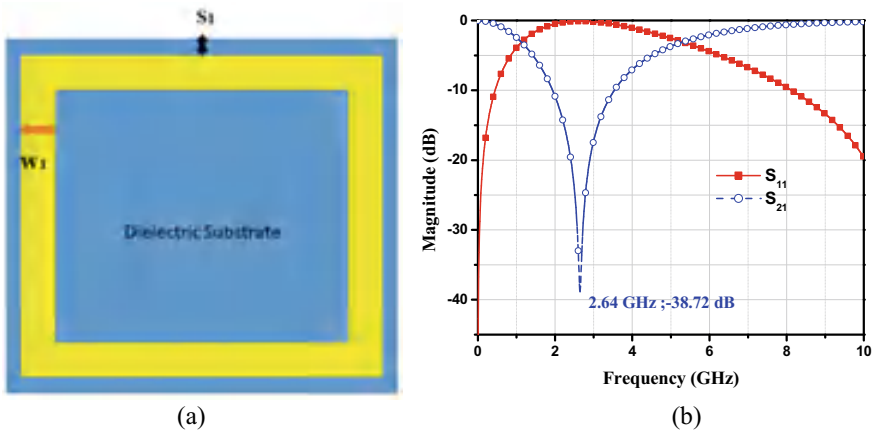
Fig. 2 **a** Unit cell with only outer ring, **b** FSS characteristics with only outer ring

(a)   (b)
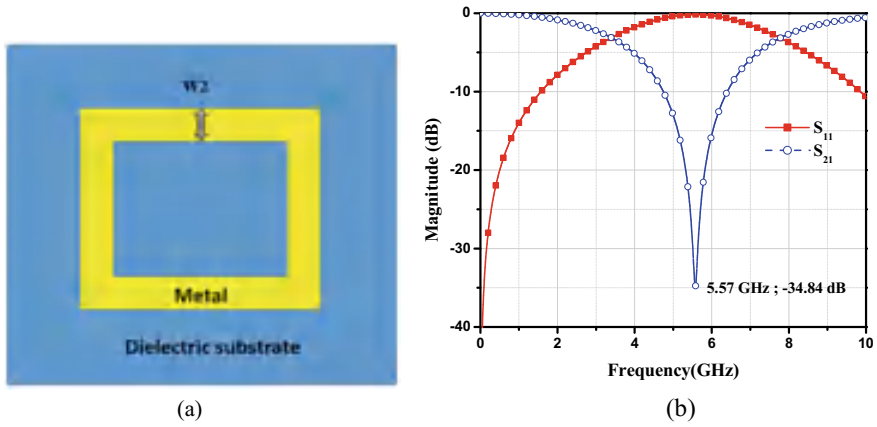
Fig. 3 **a** Unit cell with only inner ring. **b** FSS characteristics with only inner ring

(a)   (b)
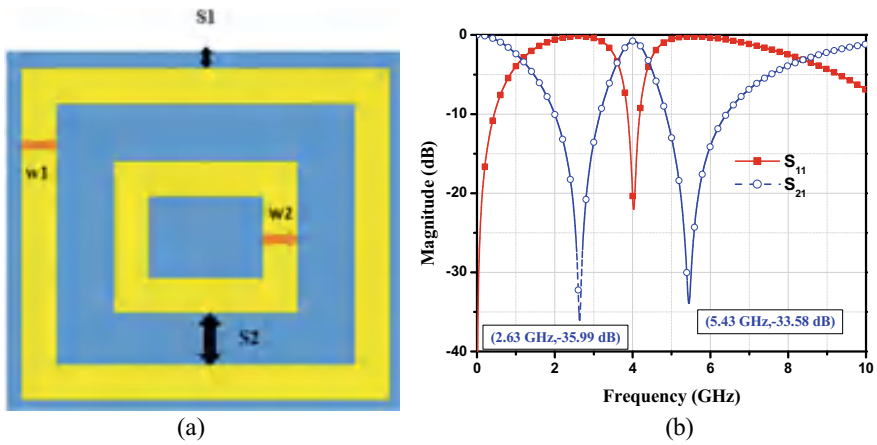
Fig. 4 **a** FSS unit cell with inner and outer rings, **b** FSS characteristics with inner and outer rings

4.1 Parametric Study

This section concentrates on various parameters of the unit cell. It examines the influence of these parametric variations on the stop band characteristics of the FSS. The proposed design of FSS is very accommodating, where any resonant frequency can be achieved by changing the parameters of the unit cell.

Parametric Variation of s1. The effect of change in the parameter 's1', the spacing between the dielectric wall and the outer loop wall, was studied by performing parametric sweep analysis on 's1', for values 0.1–2.0 mm having step width 0.1. Its FSS characteristics are shown in Fig. 5.

Fig. 5 FSS characteristics with variation in *s*1 values
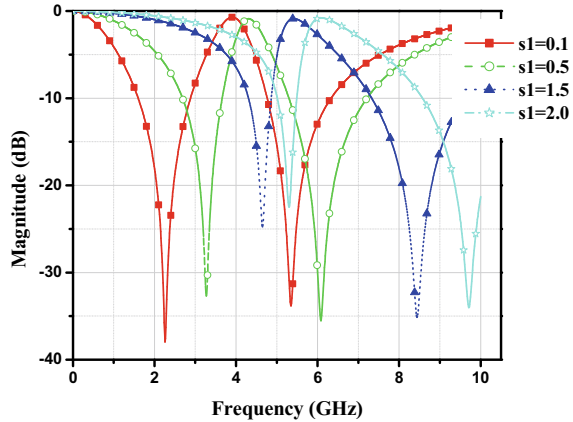
At a value of $s1 = 0.1$ mm, the resonant frequencies f_{r1} (due to outer loop) and f_{r2} (due to inner loop) have values of 2.2 GHz and 5.2 GHz, respectively. At a mid-value of $s1 = 0.5$ mm, f_{r1} and f_{r2} are 3.26 and 5.97 GHz. At the test value of $s1 = 1$ mm, the resonant values of outer and inner loops are, respectively, 3.94 and 7 GHz. At the maximum test value of $s1 = 2$ mm, the values of resonant frequencies f_{r1} and f_{r2} are 5.25 and 9.56 GHz. This indicates that a variation of the single parameter '*s*1' gives the flexibility to change both the resonant frequencies. An increase in value of '*s*1' results in increasing values of both the resonant frequencies. For the proposed design, the value of *s*1 has been fixed at 0.2 mm which provides 2.57 and 5.4 GHz for outer and inner square loops.

Parametric variation of s2. The influence of the spacing between the two loops was also studied with parametric sweep analysis by varying values of '*s*2' from 0.1 to 2.0 mm for a step width of 0.1 and its FSS characteristics are shown in Fig. 6. It was observed that by increasing the value of '*s*2' (the spacing between the outer loop and inner loop), only the higher resonant frequency f_{r2} was noticeably affected, while not much variation was seen in the f_{r1}. The higher resonant frequency due to the inner square loop increased from 3.88 GHz for a value of $s2 = 0.1$ mm to a frequency of 7.53 GHz for $s2 = 2.0$ mm.

The lower resonant frequency, however, stays put at around 2.57–2.6 GHz which indicates no influence of *s*2 on outer loop resonant frequency. Thus, the spacing between the two loops can be flexibly used to control the variation of the higher resonance frequencies. The proposed design uses $s2 = 0.9$ mm to obtain a higher resonant frequency at 5.4 GHz and an unaffected lower resonant frequency at 2.57 GHz.

Parametric variation of w1. The next parameter analyzed for parametric sweep is *w*1, i.e., the breadth of outer square ring. Width *w*1 is varied from 0.1 to 1.0 mm in steps of 0.1 and its FSS characteristics are shown in Fig. 7. The increase in *w*1 of the outer loop in the absence of the inner ring showed an increase in the lower resonance frequency (due to the outer square loop) from a frequency of 2.44 GHz

Fig. 6 FSS characteristics
with variation in *s*2 values
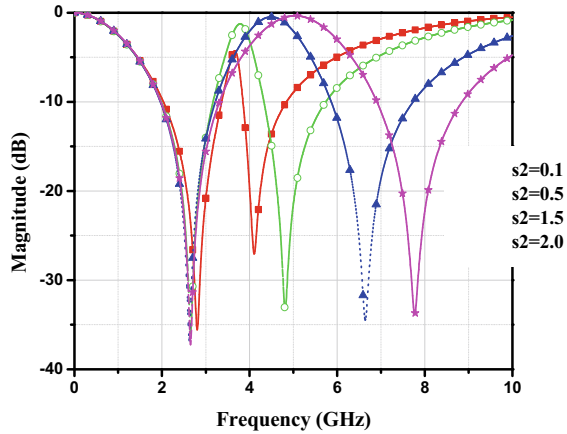
at $w1 = 0.1$ mm to a frequency of 3 GHz at $w1 = 1.0$ mm. To analyze any effects of varying the width of the outer square loop, on the inner loop, both the loops were studied together. It was observed that a variation of the width of the outer loop, which would change the spacing between the two loops, caused the lower resonant frequency to increase slightly from 2.42 to 2.91 GHz as in earlier case, but the higher resonance frequency (the inner loop) increased considerably high from 4.81 to 6.1 GHz. This implies that the outer loop width affects both the lower resonant frequencies (to a lower extent) frequencies. Design proposed uses $w1 = 0.5$ mm to adjust the resonant frequencies at 2.57 and 5.4 GHz.

Parametric variation of w2. The results of variation of the inner loop width '*w2*' are interesting. With increase in the values of inner loop width '*w2*' from 0.1 to 1.0 mm, the lower resonance frequencies were relatively unaffected (staying close around 2.5 GHz), while the higher resonance frequencies dependent on the inner loop width increased from 4.76 GHz at $w1 = 0.1$ mm to 6.2 GHz at $w1 = 1.0$ mm as shown

Fig. 7 FSS characteristics
with variation in *w*1 values
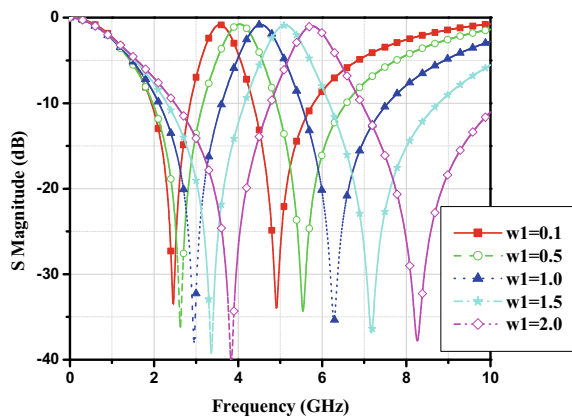
Fig. 8 FSS characteristics with variation in *w*2 values
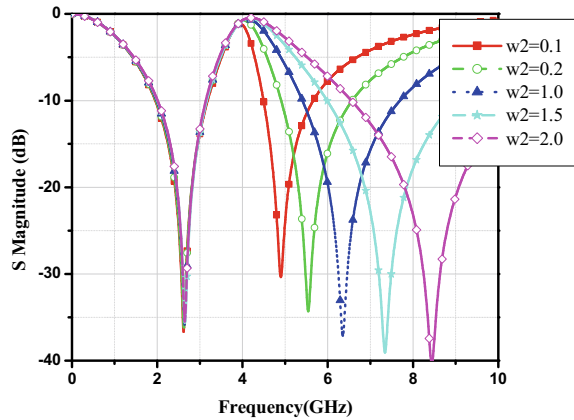
in Fig. 8. This indicates that the higher resonance frequencies can be individually controlled by varying the width of the inner square loop alone. The proposed design uses $w2 = 0.5$ mm to realize a higher resonant frequency of 5.4 GHz.

5　Conclusion

In this paper, an effective dual stopband spatial filter with resonant frequencies of 2.57 and 5.4 GHz has been proposed for Wi-Fi applications. It consists of moderately miniatured unit cells, in which separate adjustments of the frequencies of resonance have been made possible. It vouches as a desirable candidate for out Wi-Fi frequencies in the 2.57 and 5.4 GHz band of frequencies to avoid the egress of Wi-Fi signal outside the building. The FSS is advantageous in the sheer simplicity of structure to be fabricated, while at the same time allowing effective band stop characteristics that are comparable to the same effective shielding as in the works done previously by others.

References

1. Sung GH et al (2006) A frequency-selective wall for interference reduction in wireless indoor environments. IEEE Antennas Propag Mag 48(5):29–37
2. Mishra NK (2009) Development of GSM—900 mobile jammer: an approach to overcome existing limitation of jammer. In: Proceedings of the fifth IEEE conference wireless communication and sensor networks (WCSN), pp 1–4
3. Y-shield EMR-protection, shielding paints (online). http://www.yshield.com/shielding-paints.html
4. Munk BA (2000) Frequency selective surfaces: theory and design, vol 29. Wiley Online Library, Hoboken, NJ, USA
5. Wu TK (1995) Frequency selective surface and grid array. Wiley, New York

6. Yan M, Qu S, Wang J, Zhang J, Zhou H, Chen H, Zheng L (2014) Miniaturized dual-band FSS with stable resonance frequencies of 2.4 GHz/5 GHz for Wi-Fi applications. IEEE Antennas Wirel Propag Lett 13:895–898
7. Hu X, Zhou X, Wu L, Zhou L, Yin W (2009) A novel dual-band frequency selective surface (FSS). In: Asia Pacific microwave conference (APMC), pp 1227–1230
8. Chaharmir MR, Ethier J, Shanker. Design of a novel dual-band 2.4/5.5 GHz frequency selective surface using loop elements. J. Communications Research Center, Canada 3701 Carling Ave., Ottawa
9. Hong TY, Lee YS, Wee FH, You YK, Karim MNA, Ramli NH, Gan H-S, Jamlos MA (2019) Study of 5.8 GHz band-stop frequency selective surface (FSS). Int J Integr Eng 11(4):244–251
10. Monacelli B, Pryor JB, Munk BA, Kotter D, Boreman GD (2005) Infrared frequency selective surface based on circuit-analog square loop design. IEEE Trans Antennas Propag 53(2):745–752
11. Dong J, Ma Y, Li Z, Mo JA (2021) Miniaturized quad-stopband frequency selective surface with convoluted and interdigitated stripe based on equivalent circuit model analysis. J Micromachines 12:1027
12. Birwal A, Singh S, Kanaujia BK (2020) A novel design of ultra-wide stopband single-layer frequency selective surface using square-loop and cross. Int J Microw Wirel Technol: 1–10
13. Kiermeier W, Biebl E (2007) New dual-band frequency selective surfaces for GSM frequency shielding. In: Proceedings. 37th European microwave conference, pp 222–225
14. Farooq U, Iftikhar A, Farhan Shafique M, Junaid Mughal M (2019) A miniaturized and polarization insensitive FSS and CFSS for dual band WLAN applications. AEU-Int J Electron C 105:124–134
15. Yadav S, Jain CP, Sharma MM (2019) Frequency shielding with penta-band stop FSS for security and electromagnetic health applications. IEEE Trans Electromagn Compatibility Smartphone 61(3):887–892
16. Cheng DK (1989) Field and wave electromagnetics. Addison-Wesley, Reading, MA

Post-COVID-Efficient and Reliable Cardiovascular Disease Prediction Using Random Forest and GA with KNN

R. Suresh and Nagaratna Parameshwar Hegde

Abstract Machine learning uses data mining techniques to efficiently predict a cardiovascular disease based on training provided from health records of various hospitals. COVID predominantly a lung disease, in some cases the heart is also affected. Corona causes swelling and lymphatic substance in air sacs of lungs, which lessens the oxygen to reach blood stream as a result heart need to work more to pump blood through the body, effects people with preexisting heart disease. Due to insufficient oxygen, the heart attack can be caused by overwork. Previous research on heart disease prediction was based on a dataset obtained by the UCI repository. The model has validity issues that resulted in risk over estimation or risk under estimation because data are collected from a limited set of the population in a country. The present research overcomes the drawback by considering locally generated data and the dataset from UCI repository. The medical records are from various hospitals with prior consultation of experts in cardiology. The medical records are converted and inserted into a dataset to localize the prediction of CVD for the people in Telangana. To use feature-selected attributes from Random Forest and Genetic algorithm, as an input dataset to KNN algorithm as there is a need to dynamically add medical test attributes without effect on the accuracy of prediction. The proposed research concluded that Random Forest and Genetic algorithm used for features election increased the accuracy of KNN algorithm by 37.56%.

Keywords Random Forest · K-nearest neighbor algorithm · Cardiovascular disease

R. Suresh
Department of CSD, ACE Engineering College, Hyderabad, India

N. P. Hegde (✉)
Department of CSE, Vasavi College of Engineering (Autonomous), Hyderabad, India
e-mail: nagaratnaph@staff.vce.ac.in

1 Introduction

Heart attack is more frequent with COVID. A heart attack is because of high stress on heart, low oxygen or anemia, reported in patients with corona virus. In blood tests of COVID patients of some patients reported a increase troponin values in the blood, ECG deviations and chest pain."

The human body retaliates to corona virus contagion by liberation of cytokines a kind of protein released in the blood stream, for body cells to communicate for protecting body from assailant. A preventive action is exhibited in few persons, because of Genetic variation, which are more endangered to a cytokine storm. When a cytokine storm occurs, the body responds to infection which is overturned, resulting injury to kidneys, liver, and heart. A cytokine storm is difficult to survive. There are a number of unanswered questions brought on by the unanticipated COVID-19 epidemic that must be addressed [1]. Machine learning a subsection of artificial intelligence can be used to detect COVID implied complications after survival. Most of the current research is in progress to identify COVID-19 complications after recovery.

Increased values of troponin indicate heart tissue injury. Sometimes, this is from a heart attack. This is less commonly seen after COVID. Increased troponin levels during and after COVID, associated with abnormal EKG, are linked to mortality. The present research for CVD prediction aims to uncover hidden knowledge in identifying relevant attributes from a dataset collected from various hospitals in Telangana and UCI repository. Each attribute in the dataset is analyzed based on Random Forest and GA for feature selection based on the algorithm that the attributes are chosen. Most corporate hospitals rely on invasive-based diagnosis reports obtained at patients' admission in a hospital. Hence a chance of early prevention and treatment of CVD is not possible. So, there is a need for a system that would reliably and efficiently detect probable CVD which might occur in the future based on historical data available at various hospitals for timely treatment and prevention. To use feature-selected attributes from Random Forest and Genetic algorithm as an input dataset to KNN algorithm [2]. KNN is considered because there is a need to add medical test attributes without affecting the prediction accuracy dynamically. The research produces reports for evaluations with parameters like Precision, F-score, Recall, Average of Precision, and Accuracy for classification algorithm Random Forest and Genetic algorithm for feature selection and KNN algorithm for efficient CVD prediction.

2 Literature Review

COVID is reported first in Wuhan Province of China, in the month of December 2019. A new virus strain which was not been identified previously in humans which caused illness called COVID-19, and 2.7 million deaths are reported globally.

Myocardial injury is common for patients of COVID-19 at the time of admission in hospital. Older patients with injury had more cardiovascular mortalities. The troponin test was done on COVID-19 patients indicated increased troponin levels which caused a need for ventilator support and higher mortality [3]. The risk of myocardial injury and heart stroke is reported with COVID patients. Cardiac injury, in patients with COVID, had potential long-term cardiovascular effects. It is unclear whether myocardial injury and risk of cardiovascular disease are with COVID patients. Interval-based cardiac checkups of COVID patients are needed to understand health impacts. Many prediction systems have been developed for the diagnosis of various diseases using various methodologies. Breast cancer, diabetics, CVD, flu, colds, uterine fibroid illnesses, and others are examples of these diseases. For over two decades, data scientists have been predicting CVD. The majority of the papers used arrange of approaches with varying degrees of accuracy, including DT, NB, NN, and SVM. Others used data from nearby local hospitals, while some user data were from the UCI repository.

In [4], authors proposed a methodology for analyzing an individual's life habits and cultures parameters with a compactable neuron fuzzy deduction model, which acts as a predictive system for the doctors to predict the occurrence of CVD and assists doctors to characterize the scale of CVD. CVD can be prevented by changing the patient's lifestyle medications, taking a healthy diet, and exercise.

In [5], authors proposed a model that is used to accelerate operations while also increasing thoroughness and, most importantly for identifying illnesses in the quickest time possible, resulting in better execution than traditional diagnostic methods.

In [6], authors proposed a PSO and feedforward neural networks. The model divides the orders into two groups' of sick and healthy persons in the initial stage. The PSO method is used for all subsets in the third stage to find the best subset in terms of time, thoroughness, less cost, and accessibility. By using the PSO algorithm, the subset incorporated eight features sex, slope, oldpeak, exang, trestbp, fbs, chol, and age.

In [7], authors used a no. of algorithms to forecast cardiac diseases, including the DT model, J48 Model, LMT technique, NB model, KNN model, and SVM. They presented a new methodology for improving DT accuracies in the prediction of heart-related diseases. The WEKA and UCI datasets were used in the construction. The NB outperformed others, with the SVM, DT, and KNN following closely behind.

In [8], authors proposed Relief *F* and Rough Set approaches which were integrated into a hybrid classification system. The system comprises two subcategories: the RFRS feature drawing model and a classification model with multiple classifiers. Data preprocessing, feature extraction, and the Relief *F* algorithm are the three stages of the first system. Multiple classifiers' model is developed for the second system. Statlog data were obtained from the UCI repository. The accuracy of 92.59% was obtained.

3   Implementation of the Existing and Proposed Models

Libraries like Sklearn, Pandas, NumPy, etc., and CVD dataset are installed. The dataset consists of 10,082 patient records for the proposed model. Then, split the datasets in a way 80% for training purpose and 20% to test. *K* value is chosen based on the error curve, optimal variance, and biased error. The accuracy obtained by the existing model KNN algorithm model is used to predict the accuracy obtained which is 62.3%. The accuracy obtained by the proposed model Random Forest and Genetic with KNN algorithm is 99.86%.

3.1   Performance Evaluation of Existing and Proposed Models

Various parameters like Accuracy, Precision, Recall, and *F*-measure are calculated based on obtained Confusion Matrix.

Table 1 is the analysis of classification report obtained by KNN algorithm considering the dataset from UCI repository total of 303 records considered. The performances are compared with various valuation parameters.

Figure 1 is obtained an accuracy of 62.3% by considering the dataset which is taken in the split ratio of 20% for testing and 80% for training from 303 CVD dataset. Various accuracy calculating parameters are analyzed.

Table 2 is the analysis of classification report obtained by Random Forest and Genetic with KNN algorithm considering the dataset from various hospitals in Telangana and UCI repository total of 10,082 records considered. The performances are compared with various evaluation parameters.

Figure 2 is obtained an accuracy of 99.86% by considering dataset which is taken in the split ratio of 20:80 for training and testing from a total of 10,082 CVD dataset. Various accuracy calculating parameters are considered (Fig. 3).

ROC graph shows the classification model KNN algorithm performance at various thresholds of training and testing. The receiver operator curve was displayed with Recall on *X*-axis and with Precision on *Y*-axis. The AP value obtained by KNN algorithm is 0.62 (Fig. 4).

Table 1   Classification report obtained by KNN algorithm

Classification report				
Actual	Precision	Recall	*F*1-score	Support
0	0.58	0.56	0.57	27
1	0.66	0.68	0.67	34
Macro avg.	0.62	0.62	0.62	61
Weighted avg.	0.62	0.62	0.62	61

Fig. 1 CM obtained by KNN algorithm
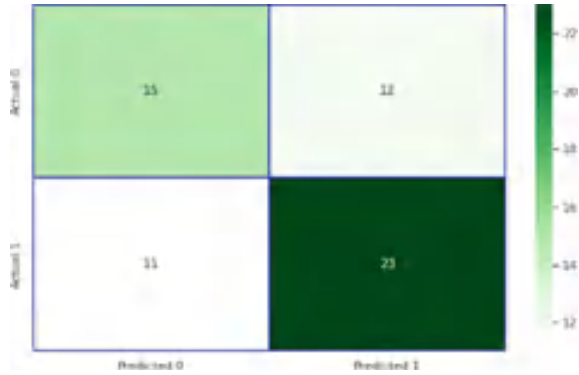
Table 2 Classification report obtained by Random Forest and Genetic with KNN algorithm

Classification report				
Actual	Precision	Recall	F1-score	Support
0	0.99	0.99	0.99	860
1	0.99	0.98	0.99	1157
Macro avg.	0.99	0.99	0.99	2017
Weighted avg.	0.99	0.99	0.99	2017

Fig. 2 CM obtained by Random Forest and GA with KNN
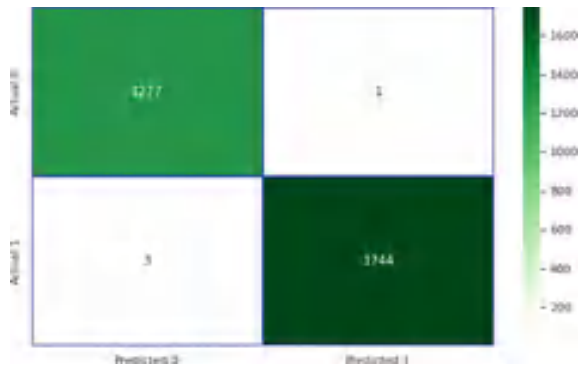
ROC graph shows the classification model Random Forest and Genetic and KNN algorithm performance at various thresholds of training and testing. The Receiver Operator curve was displayed with Recall on *X*-axis and with Precision on *Y*-axis. The AP value obtained by the SVM algorithm is 1.0.

The proposed KNN using Random Forest and Genetic algorithm for feature selection obtained an accuracy of 99.86%. Various performance evaluation parameters are considered in comparing the result obtained in an existing model and proposed model
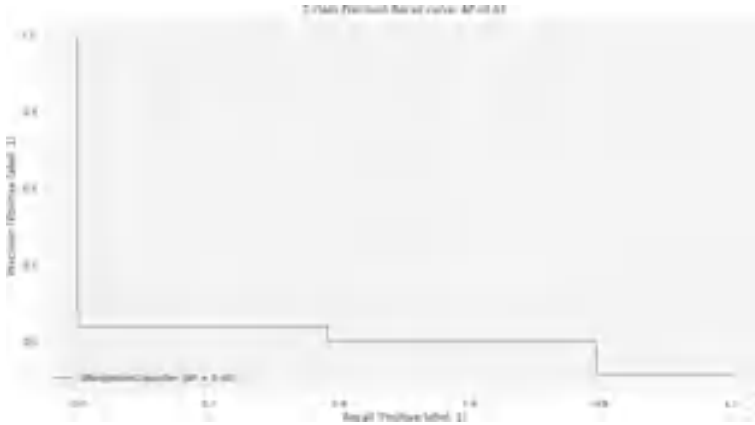
Fig. 3 Two-class precision and recall curve of KNN algorithm

Fig. 4 Two-class precision and recall curve of Random Forest and GA with KNN
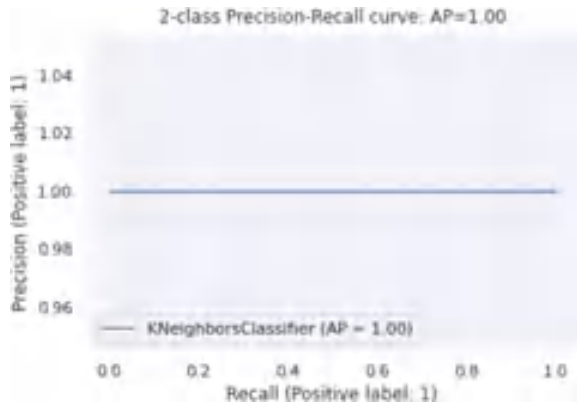
that is Random Forest and Genetic model with KNN. Accuracy increased by 37.56% considering Random Forest and Genetic algorithm for feature selection.

Table 3 is the performance comparisons of the existing KNN algorithm with the proposed Random Forest and Genetic with KNN algorithm obtained, analyzed the efficiency with various parameters.

Table 3 Performance details of KNN and Random Forest and GA with KNN

Algorithm used	Class	Precision	Recall	F-score	AP	Accuracy
KNN	0	0.58	0.56	0.57	0.6	62.3
	1	0.66	0.68	0.67		
Random Forest and GA with KNN	0	0.99	0.98	0.99	1.0	99.8
	1	0.99	0.98	0.99		

4 Conclusion

Data for the study were provided by several hospitals in Telangana and UCI repository. The data are of patients with corona virus and CVD. The dataset of 10,082 samples and 16 attributes out of 13 numeric input attributes such as tropt, bnp, age, sex, cp type, chol, fbs, restecg, thalach, exang, oldpeakvalue, slope, ca, thal, and output attribute known as CVD. Output is divided into two categories without CVD denoted by $CVD = 0$ and with CVD denoted by $CVD = 1$.The present paper feature has extracted a two new attributes Troponin and B-type Natriuretic peptide which are very much crucial in efficient prediction of heart disease of post-COVID patients. The proposed research concluded that Random Forest and Genetic algorithm used for feature selection increased the accuracy of KNN algorithm by 37.5%. Different feature selection strategies can be used in the future to handle high-dimensional data. Additionally, it is intended to put the suggested technique into practice to develop fresh frameworks for anticipating a variety of post-COVID-19 difficulties which might occur because of mutation.

References

1. Oh MS, Jeong MH (2020) Sex differences in cardiovascular disease risk factors among Korean adults. Korean J Med 95(4):266–275
2. Jamshidi M, Lalbakhsh A, Talla J, Peroutka Z, Hadjilooei F, Lalbakhsh P, Jamshidi M, La Spada L, Mirmozafari M, Dehghani M et al (2020) Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. IEEE Access 8:109581–109595
3. Kassania SH, Kassanib PH, Wesolowskic MJ, Schneidera KA, Detersa R (2021) Automatic detection of coronavirus disease (COVID-19) in X-ray and ct images: a machine learning based approach. Biocybern Biomed Eng 41(3):867–879
4. Adadi A, Lahmer M, Nasiri S (2021) Artificial intelligence and covid-19: a systematic umbrella review and roads ahead. J King Saud Univ Comput Inf Sci
5. Trevisan C, Sergi G, Maggi SJB, Dynamics H (2020) Gender differences in brain-heart connection in brain and heart dynamics. Springer, Cham, Switzerland, p 937
6. Naudé W (2020) Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. AI Soc 35(3):761–765
7. Karthik R, Menaka R, Hariharan M, Kathiresan G (2021) AI for COVID-19 detection from radiographs: incisive analysis of state of the art techniques, key challenges and future directions. IRBM
8. Surianarayanan C, Chelliah PR (2021) Leveraging artificial intelligence (AI) capabilities for COVID-19 containment. New generation computing, pp 1–25
9. Weerahandi H, Hochman KA, Simon E, Blaum C, Chodosh J, Duan E, Garry K, Kahan T, Karmen-Tuohy SL, Karpel HC et al (2021) Post-discharge health status and symptoms in patients with severe COVID-19. J Gen Intern Med 36(3):738–745
10. Kamal M, Abo Omirah M, Hussein A, Saeed H (2021) Assessment and characterization of post-COVID-19 manifestations. Int J Clin Pract. https://doi.org/10.1111/IJCP.13746
11. Wang Y, Hu M, Li Q, Zhang XP, Zhai G, Yao N (2020) Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. arXiv:2002.05534

12. Apostolopoulos ID, Mpesiana TA (2020) COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43(2):635–640

13. Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT (2020) New machine learning method for image-based diagnosis of COVID-19. PLoS ONE 15(6):e0235187

14. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X (2020) Artificial intelligence and machine learning to fight COVID-19

15. Shahid O, Nasajpour M, Pouriyeh S, Parizi RM, Han M, Valero M, Li F, Aledhari M, Sheng QZ (2021) Machine learning research towards combating COVID-19: virus detection, spread prevention, and medical assistance. J Biomed Inform 117:103751

16. Yadav DC, Pal S (2020) Prediction of heart disease using feature selection and random forest ensemble method. Int J Pharmaceutical Res 12(4)

17. World Health Organization, Dostupno J (2016) Cardiovascular diseases: key facts, vol 13, no 2016, p 6

18. Uyarand K, Ilhan A (2017) Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. Procedia Comput Sci 120:588–593

19. Pouriyeh S, Vahid S, Sannino G, De Pietro G, Arabnia H, Gutierrez J (2017) A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In: Proceedings of IEEE symposium computing communication (ISCC), July 2017, pp 204–207

20. Mourao-Miranda J, Bokde ALW, Born C, Hampel H, Stetter M (2005) Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. Neuroimage 28(4):980–995

21. Ghwanmeh S, Mohammad A, Al-Ibrahim A (2013) Innovative artificial neural networks-based decision support system for heart diseases diagnosis. J Intell Learn Syst Appl 5(3):176–183

22. Al-Shayea QK (2011) Artificial neural networks in medical diagnosis. Int J Comput Sci 8(2):150–154

23. Shamrat FMJM, Raihan MA, Rahman AKMS, Mahmud I, Akter R (2020) 'An analysis on breastdisease prediction using machine learning approaches.' Int J Sci Technol Res 9(2):2450–2455

24. Amin MS, Chiam YK, Varathan KD (2019) Identification of significant features and data mining techniques in predicting heart disease. Telemat Informat 36:82–93

25. Kausar N, Palaniappan S, Samir BB, Abdullah A, Dey N (2016) Systematic analysis of applied datamining based optimization algorithms in clinical attribute extraction and classification for diagnosis of cardiac patients. In: Applications of intelligent optimization in biology and medicine. Springer, Cham, Switzerland, pp 217–231

26. Mackay J, Mensah GA (2004) The atlas of heart disease and stroke. Tech. Rep. World Health Organization, Geneva, Switzerland

27. Ashraf M, Ahmad SM, Ganai NA, Shah RA, Zaman M, Khan SA, Shah AA. Prediction of cardiovascular disease through cutting-edge deep learning

28. Hastie T, Tibshirani R, Friedman J (2020) The elements of statistical learning. Data mining, inference, and prediction. Springer, Cham, Switzerland

Estimation of Doubly Selective Channel in FBMC-OQAM and OFDM Systems

K. Pranathi and A. Naveena

Abstract The multicarrier modulation (MCM) purpose is to facilitate the transmission of data by dividing the stream into a number of individual bit streams, each of which has a considerably lower bit rate, and then employing those individual bit streams as the modulating source for more than one carrier. Multiple carrier modulation techniques, such as filter bank multicarrier (FBMC) and orthogonal frequency-division multiplexing (OFDM) systems, are utilized in this process. In comparison to the OFDM system, the spectral efficiency of the FBMC is significantly better. The results of the simulation were utilized to investigate the architecture, justification, and estimation of FBMC-offset quadrature amplitude modulation (OQAM), and a comparison with the OFDM system is observed. A doubly selective channel estimation is implemented, which is a basic time-selective and frequency-selective channel estimation. The channel estimation process indicates which is similar to have full channel availability, the expectation and maximization-based strategies are used twice. In comparison, it considers the current state of channel prediction processes. The signal-to-noise ratio (SNR) and measured bit error rate (BER) are calculated using estimated (recovered) data and input data sources for both systems.

Keywords Channel estimation · Doubly selective channel estimation · FBMC-OQAM · Multipath carrier · OFDM

1 Introduction

Advancement in science and technology plays a highly commendable role in the lives of humans. The technological innovations and progress will enhance the quality of life and plays an important aspect for the healthy living, which will help life of human beings to be social and communicate with others. The decade's progress in wireless technology had brought revolutionary changes in the lifestyle of people living in this

K. Pranathi (✉) · A. Naveena
G. Narayanamma Institute of Technology and Science (For Women), Hyderabad, India
e-mail: kpranathi26@gmail.com

earth. This progress is highly appreciable. The changes in signal characterization have brought a demand for the various innovations and developments in the wireless communication system [1]. The type and property of signal used will determine the wireless technology which will in turn lead to the improvement in the various performance parameters of the wireless system.

The environmental conditions and their properties will also affect the characteristic of the signal. The attenuation, interference, other specific parameters need to be considered while characterizing the signal as these properties will affect the quality of the signal. A frequency-selective channel into multiple frequency-flat subchannels can turn effectively at different subcarriers using the single input single output (SISO), single input multiple output (SIMO), and multiple input single output (MISO) techniques [2]. The simple one-tap channel equalization allows to limit the multipath effect.

The standardization of third-generation partnership project (3GPP) was selected as the main fifth generation (5G) waveform for millimeter-wave bands and sub-6 GHz [3]. To utilize the spatial multiplexing and diversity gain, FBMC techniques can integrate with the channel estimation methods that allow the systems for achieving the high throughput than the FS fading channels. In modern telecommunication systems, MIMO systems are ubiquitous, such as wireless local area network (WLAN) systems, and long-term evolution (LTE).

The LTE-based 4G and 5G communications [4] require the higher data rates to satisfy the requirements of user equipment (UE)'s. However, the conventional MIMO systems are failed to provide the maximum throughput, data rates to end users, which is challenging task [5]. Further, increment in the data rates, causing higher power consumption, resulted in reduction of SE and EE.

2 Literature Survey

1. Channel Estimation in OFDM Systems

The channel estimation in the OFDM is based on time-domain channel properties' use of multiamplitude signaling schemes in a wireless system which needs tracking of multipath fading radio channels by the usage of differential phase-shift keying (DPSK). A limited no. of bits per symbol results in a 3 dB SNR loss. The least square (LS) estimation is complex when compared to MMSE and prior knowledge of channel covariance and noise variance is assumed by MMSE. LS estimation is simple for higher and lower SNRs, and the modifications of LS and minimum mean-square error (MMSE) estimators are compromised between complexity and performance.

A 16-QAM system with the symbol error rate is presented to utilize the simulation results which are depended on the estimator complexity in SNR up to 4 dB which is gained over LS estimator and the SNR gain is less compared to the modified MMSE.

This paper mainly emphasized four aspects

- Estimation of the channel using the channel frequency response (CFR) according to the traditional method.
- Estimation of the channel based on the parametric model (PM).
- Iterative channel estimate.
- Estimation of the channels used in FBMCM systems.

Advantages of the system are

- Efficient usage of channel.
- Prior knowledge of channel covariance and noise variance.

Disadvantages are

- Error estimation is complicated.
- Loss of spectral efficiency due to CP.

2. Estimation of Doubly Selective Fading Channels

The accuracy of the existing basis expansion model (BEM) is improved with a simple windowing and de-windowing technique used at the receiver only on the branch of channel estimation and the WLS (Least squares windowed) technique, which is developed for a doubly selective fading channel. Additionally, by using these techniques, an minimum mean-square error (MMSE) estimator also increases its performance. A pilot pattern is designed to analyze the simulations which provide the improvements to verify the performance by using a simple estimator and compare it with the existing one.

The linear MMSE estimator can easily be used by the available channel statistics with the number of pilots which are observed to interpolate whether BEM is required or time Doppler is required to analyze and minimize the MSE of the channel estimation with the usage of optimum pilots, which results from the AWGN channel and the high-frequency components in the Doppler domain. The resultant mean-square error (MSE) weighted of the windowed truncation at the center is smaller than at the edges, and this motivates to retain results at the center to cover the total time domain by making use of the sliding window. This is the importance of all pilot sub-blocks, and the spaces must be identical. When the SNR is low, the MMSE estimators overcome the LS by making use of the channel statistics in order to reduce the channel estimate MSE. However, this MSE advantage does not seem to immediately transfer over to the BER comparisons, and it was noted that at low SNR the BER performance is similar with the WLS estimators, which, although having the MSE advantage, ends up having the lowest performance out of all of the MMSE estimators. This demonstrates that using a window strategy is beneficial. Because the schemes with windows only take the results between the two center pilots, they require p times as much computation as the ones with fastened windows with fixed ones and the transceiver design. This is due to the fact that the transceiver design and the ones with fastened windows with fixed ones are not sterilized. When the SNR is high, the performance of MMSE estimators with rectangular windows' area

units is worse than that of LS estimators with appropriately sized windows. Most importantly, the fixed rectangular window performs worse than all WLS estimators even when the signal-to-noise ratio is medium. This demonstrates that the straightforward windowed technique is beneficial for both LS estimators and MMSE estimators simultaneously.

Advantages

- Improves system performance.
- Improves the accuracy of DFT and BEM.

Disadvantages

- Data loss will be addressed due to the concept of windowing and de-windowing.
- Mismatching due to the frequency offset.
- Improper selection of windowing technique.

3. **Bit Error Probability for Pilot-Symbol-Aided Channel Estimation in FBMC-OQAM**

The closed-form equation for Bit Error Probability (BEP) comprises a channel estimator that is defined for both FBMC and OFDM systems, and the primary focus is to compare the two types of systems. We make the assumption of Rayleigh fading channel with a low delay spread and a low Christian Johann Doppler spread in order to make it possible to ignore the channel-induced interference in comparison to the noise. The pilot symbols are supported by channel estimation, which helps to cancel the unreal interference that is intrinsically created in FBMC at the pilot places. This interference may be canceled by auxiliary symbols or by writing code. The optimal power distribution between pilot symbols and the knowledge of data symbols is developed in order to reduce the BEP associate degree as much as possible. It is observed that the conventional OFDM system consumed 1 dB of power offset and proposed FBMC-OQAM resulted in 1 dB of power offset for various data sources. Additionally, for coding, such an offset leads to similar information of data symbol are estimated under the time-varying channel estimation conditions. The improper placement of pilot symbols in existing OFDM makes more complicated channels' estimation, so FBMC-OQAM system places the pilots in exact places between the data symbols. The FBMC uses a coding method to cancel the unreal interference at the pilot positions, and it may be found that BEP expressions were conjoint because they allow for the search of the optimal pilot-to-data power offset in closed-loop form. This discovery may lead to the conclusion that BEP expressions were conjoint. The performance is enhanced for constant transmission power, and a reduced BEP is attained, when such power offset is increased.

Advantages

- High power offset.
- Error rates are reduced by using the coding process.

Disadvantages

- Energy wasted.

3   OFDM System

OFDM is a method of MCM that divides a single data stream into a large number of narrowband channels operating at a variety of frequencies. This method is used to reduce signal distortion and crosstalk. Subcarrier orthogonality offers resilience against multipath effects and makes it possible to easily design an FFT block, as seen in Fig. 1.

It is able to use adaptive modulation methods as well as various bit-loading algorithms. On the other hand, the conventional OFDM method [6] has a variety of drawbacks but is used as a low-cost option for the purpose of meeting the complexities and needs of 5G wireless networks. In OFDM, the problem of time distortion is solved by inserting a cyclic prefix, which is a repeated period, between the two symbols. This makes it easier to equalize each subcarrier. Because of the existence of a cyclic prefix in the OFDM system [7], the available transmission time is used in an inefficient manner. This violates the optimum latency requirement, which in turn results in poor spectral fidelity and decreased throughput. These out-of-band emissions are the root of the problem known as inter-carrier interference (ICI). In networks with time-varying channel characteristics, the orthogonality of many of the subcarriers is distorted, which leads to ICI. The OFDM system is impacted by both Inter-Symbol Interference (ISI) and ICI. As a consequence, more strict synchronization processes are necessary, in addition to an increase in the amount of power that must be used [8]. The OFDM system suffers from a high Peak-to-Average Power Ratio (PAPR), a decrease in spectral efficiency as a result of the insertion of cyclic prefixes, and a loss of orthogonality as a result of the faulty synchronization. These deficiencies are addressed by the currently planned effort, which is based on the FBMC-OQAM system.
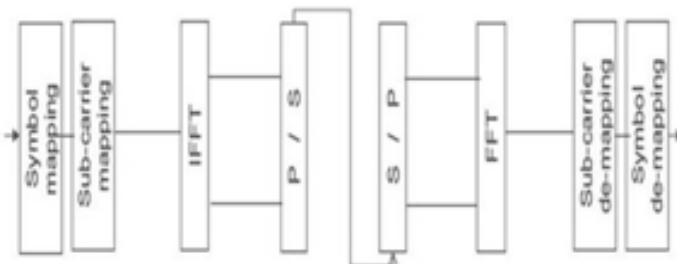
Fig. 1   OFDM block diagram

4 FBMC System

This framework, which is based on an extension of the OFDM scheme known as the OFDM with FBMC system, serves as the foundation for the vast majority of the wireless mobile networking infrastructure that exists today. A multicarrier method is one in which data symbols are sent concurrently across more than one frequency subcarrier. The FBCM and OFDM systems both are the MCM methods. However, the OFDM approach does not have additional preprocessing and posing filter banks, i.e., synthesis and analysis filter banks [9]. The absence of these filters resulted in improper correlation of cyclic prefix, which ultimately causes to increment of PAPR. Further, the FBCM can preprocess multiple users' data usin single filter bank at a time, which caused to reduce the BER. This filtering property is also not presented in conventional OFDM system. This system, which makes use of a filter bank in both the transmitter and receiver sections, as illustrated in Fig. 2, is the source of the most significant improvement. Due to the lack of a separate channel estimation tool in the FBMC scheme, an MMSE-based doubly selective channel estimation process with single-tap channel equalization is implemented on the obtained data with upgraded pilots at the receiver.

Symbol Mapping: The primary purpose of this building block is to convert binary data into symbols, after which it maps those symbols as frames to the subcarrier mapping that is being used as input.

Subcarrier Mapping: It is an essential step in the manufacturing of FBMC frames. Every frame of framed data has a preamble configuration that is used for fast tuning of carrier frequency, time synchronization, and the frequency of each subcarrier which is chosen from an orthogonal signal range, and those frequencies are known at the receiver to recover the signal [10]. This allows the receiver to quickly adjust the carrier frequency.
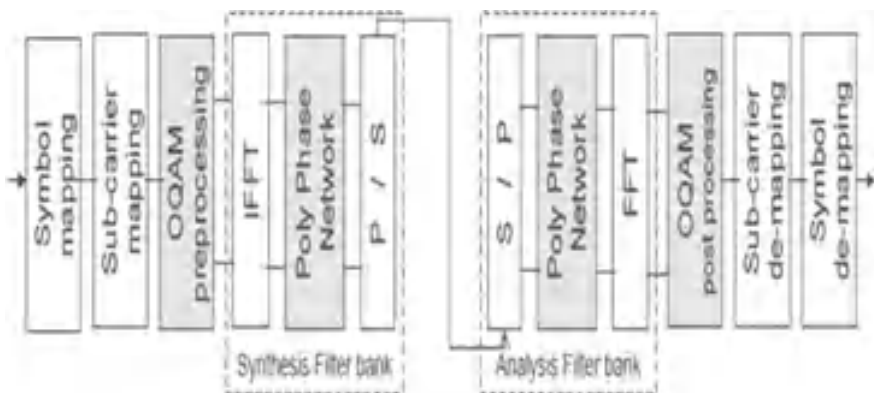
Fig. 2 Modulator reconstruction for FBMC utilizing several prototype filters and OQAM processing

IFFT: (Inverse Fast Fourier transform): It is a technique that converts an area or signals into the frequency domain. It is also known as the reverse Fourier transform.

FFT (fast Fourier transform): It is a method that converts a signal from its initial domain into the frequency domain.

OQAM: The prototype filter makes use of the OQAM approach. It is the first filter in the whole filter bank that is coupled to the zero frequency carriers that is able to give good functionality [11]. Preprocessing done by OQAM: It performs two distinct functions: The translation from abstract to concrete and multiplication by the use of sequences. Postprocessing is also done by OQAM. In addition to this, it features two operations: multiplication by the use of sequences and the transformation from real to complex.

TMUX (Transmultiplexer): It is the name given to a configuration for synthesis analysis, and a digital filter bank is a collection of filters that share a common input or output.

Synthesis filter bank: A MISO function is performed by the synthesis filter bank. It has M upsamplers and M synthesis filters, which are used to accommodate the synthesis filter bank that is depicted on the transmitter side in Fig. 2. The signals that are sent into $M/2$ are upsampled, and after that, they may be filtered using synthesis filters.

Analysis filter bank: Spectrum analysis is performed with the help of the analysis filter bank, which functions as Single Input and Multi Outputs (SIMO). As can be seen in Fig. 2, it has M analysis filters in addition to M downsamplers, which work together to compensate for the analytical filter bank that is employed on the receiver side. After being downsampled by a factor of $M/2$, each of the input signals will first be filtered, but solely with an analytical filter [12]. This will allow for the generation of the output signal.

4.1 Doubly Selective Channel Estimation

The estimation of the doubly selective channel is shown in Fig. 3, with the estimation and maximization operation with channel impulse response matrix, KF = kappa factor, h = time-based pilots, m = frequency-based pilots, ypn = error-free output, p = prediction value, e = estimated, $n-1$ = previous samples, $rn(k)$ = received pilot with samples k. The time and frequency-dependent pilots are generated by the doubly selective channel estimation using the EM operation, and the time-based pilot symbols are applied [6] using the expectation operation and the frequency-based pilot symbols are applied using the maximization operation. The updated channel matrix with high accurate pilots would be the product of this EM step. The data will be perfectly adjusted by using three steps to mitigate the error which is (1) Initialization Path, (2) Basic Error Correction, and (3) Final Error Correction.
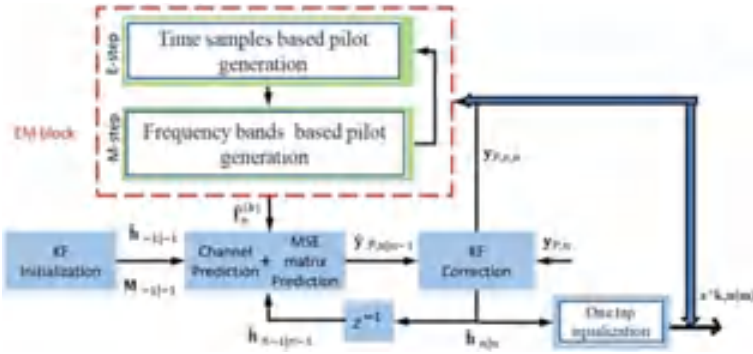
Fig. 3 Block diagram of doubly

1. Initialization Path: This process starts from KF Correction to Z^{-1} from Z inverse transformation to channel and MSE prediction method to KF Correction, this path is used to update the pilots at the receiver side every time.
2. Basic Error Correction: This error correction path starts from KF Correction to One Tap Equalizer and compares with EM-block; at last, it connects to KF Correction. In this path, the received pilots are compared with the transmitted pilots and result in error-free data output. If the errors are still present, further it is performed by the final error correction path.
3. Final Error Correction: Final path starts from KF Correction to One Tap Equalizer to EM-block and passes through channel and MSE prediction to KF Correction. In this path, if the multiple number of errors are present, then new channel prediction values are generated by the updated channel impulse matrix (KF). At last, all types of errors will be suppressed by using the updated values, and at each time and frequency samples, the data are adjusted perfectly.

4.2 FBMC-OQAM System with Doubly Selective Channel Estimation

In the FBMC-OQAM system with doubly selective channel estimation as shown in Fig. 4, the preamble insertion serves as a header and the pilots serve as protection for the subcarriers [8]. The frequency band pilots are created and used for perfect data estimation by expectation process, and they are concerned with how many data symbols are expected (recovered) across the transmitter and receiver sections, and in maximization, the frequency-based pilots are generated and used for perfect data estimation [2], across the subcarrier demapping, the receiver segment implements doubly selective channel estimation.
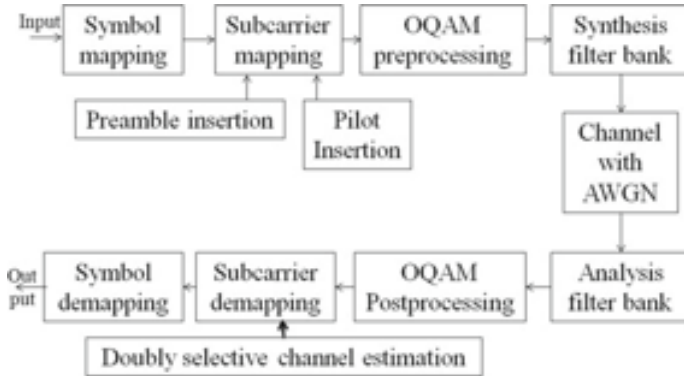
Fig. 4 Block diagram of FBMC-OQAM with doubly

4.3 Interference Cancelation Methodology in Doubly Selective Channel Estimation

- Channel equalization is highly relevant for a low-complexity interference cancelation scheme as the key problem of doubly selective channel estimation.
- Since the channel estimates at the pilot positions are distorted by interference, interference cancelation is also essential for the channel estimation process.
- The accuracy of channel estimation can be strengthened by canceling this interference. The working process of iterative channel estimation and interference cancelation scheme is:
- MMSE channel estimation of the transmission matrix $\hat{D}$ (0) as shown in Eq. (1), one element of the transmission matrix $\hat{D}$, at row position $\overline{l_1 k_1} = l_1 + L k_1$ and $\overline{l_2 k_2} = l_2 + L k_2$ column position can then be estimated by:

$$\left[\hat{D}\right]\overline{l_1 k_1, l_2 k_2} = \tilde{W}^H_{\overline{l_1 k_1, l_2 k_2}} \hat{h}^{Ls}_p. \tag{1}$$

- The weighting vector in Eq. (2) has a major influence on the channel estimation accuracy; the MMSE weighting vector can be calculated by:

$$\tilde{W}^H_{l_1 k_1, l_2 k_2} = R^{-1}_{\hat{h}^{LS}_p} r_{\hat{h}^{LS}_p, [\hat{D}]\overline{l_1 k_1, l_2 k_2}}. \tag{2}$$

- One-tap equalization and quantization are shown in Eq. (3) with

$$\hat{h}^{(0)} = \mathrm{diag}\left\{\hat{D}^{(0)}\right\},$$

$$\hat{X}^{(0)}_{l,k} = Q\left\{\frac{y^{(0)}_{l,k}}{\hat{h}^{(0)}_{l,k}}\right\}. \tag{3}$$

- Initialize with $i = 0$.
- Interference cancelation can be done by a signal as shown in Eq. (4),

$$y^{(i+1)} = y - \left(\hat{D}^{(i)} - \text{diag}\left\{ \text{diag}\left\{ \hat{D}^{(i)} \right\} \right\} \right) \hat{X}^{(i)}. \qquad (4)$$

- Improved estimation of the transmission matrix $\hat{D}$ $(i+1)$ enabled by a reduced interference at the pilot positions.
- Improved one-tap equalization and quantization are shown in Eq. (5),

$$\hat{X}_{l,k}^{(i+1)} = Q\left\{ \frac{y_{l,k}^{(i+1)}}{\hat{h}_{l,k}^{(i+1)}} \right\} \qquad (5)$$

- Repeat Step 4–Step 7. We consider $i = 0, 1, ..., 4$.

5   Simulation Results and Discussion

To evaluate the performance of the FBMC-OQAM and OFDM systems with doubly selective channel estimation which compares both systems with the data spreading method and auxiliary symbols method, the estimation of the system is carried out with the MATLAB software. The system parameters used for the simulation are shown in Table 1 (Figs. 5, 6, 7, 8; Tables 2, 3, 4, 5).

The BER performance of doubly selective channel estimation with CSI of interference cancelation scheme and one-tap equalizer with the three methods such as OFDM (data spread), FBMC-OQAM (data spread), and FBMC-OQAM (auxiliary method) is shown in the above graphs, and the BER performance concerning with the changes in SNR is shown in above tables. FBMC-OQAM system with the auxiliary symbol method gives better results in BER performance (BER reduces).

Table 1 Simulation parameters of the system

Parameter	Value
Modulation technique	256-QAM
Channel model	AWGN Channel
Used subcarrier	24
Number of transmit antennas	1
Number of receiving antennas	1
Number of iterations	4
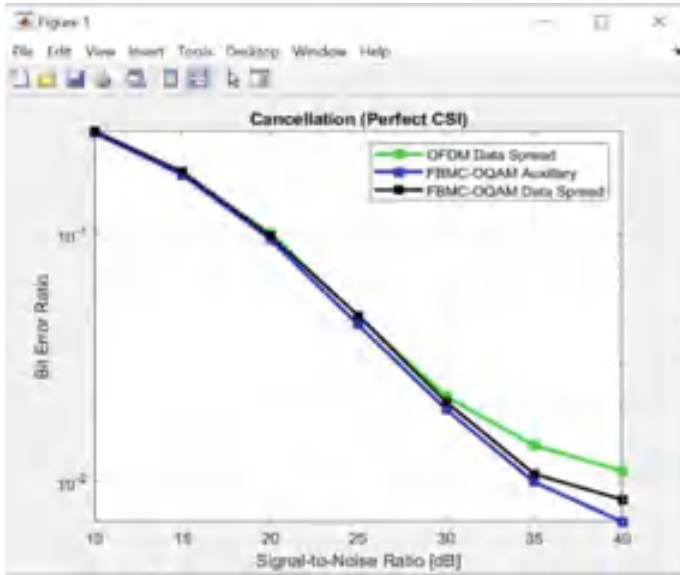
Fig. 5 Simulation outcome of BER versus SNR by the measurements of various techniques like OFDM and FBMC-OQAM with data spread and FBMC-OQAM auxiliary with perfect CSI
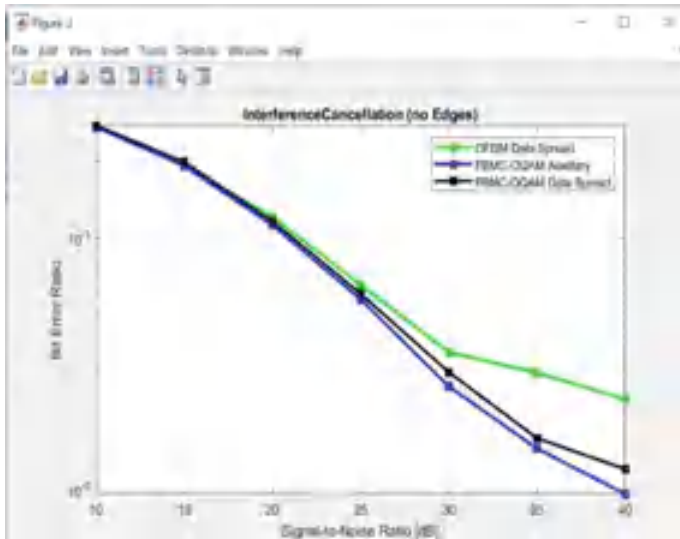
Fig. 6 Simulation outcome of BER versus SNR by the measurements of various techniques like OFDM and FBMC-OQAM with data spread and FBMC-OQAM auxiliary with no edges
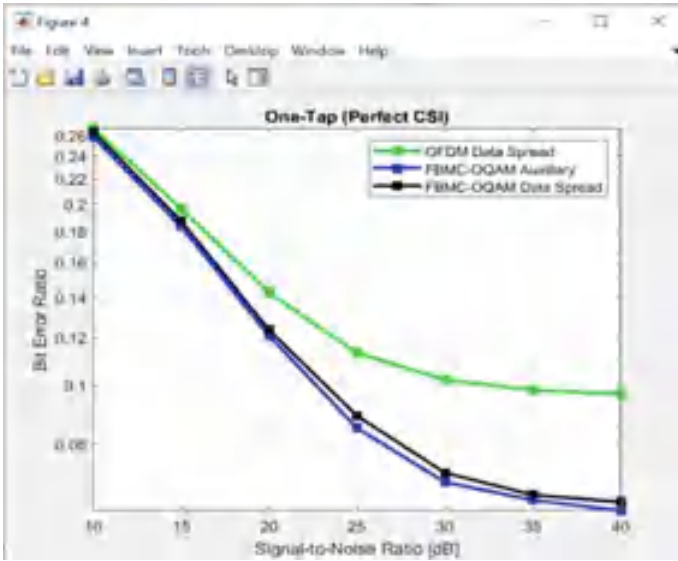
Fig. 7 Simulation outcome of BER versus SNR by the measurements of various techniques like OFDM and FBMC-OQAM with data spread and FBMC-OQAM with the auxiliary method of one-tap equalizer with perfect CSI
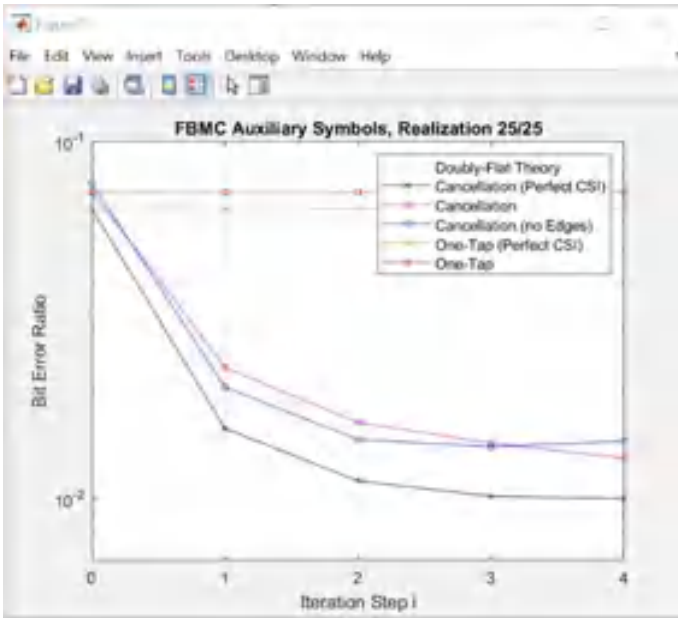
Fig. 8 Simulation outcome of BER versus iteration step '*i*' by the FBMC-OQAM system with the auxiliary method of one-tap equalizer, interference cancelation with perfect CSI and no edges

Table 2 BER values for OFDM, FBMC-OQAM (data spread), and FBMC-OQAM (auxiliary method) with perfect CSI

SNR (dB)	BER for OFDM data spread	BER for FBMC-OQAM data spread	BER for FBMC-OQAM Auxiliary
10	0.2603	0.2603	0.2603
20	0.09800	0.09846	0.09800
30	0.02206	0.02099	0.01961
40	0.01108	0.008488	0.006922

Table 3 BER values for OFDM, FBMC-OQAM (data spread), and FBMC-OQAM

SNR (dB)	BER for OFDM data spread	BER for FBMC-OQAM data spread	BER for FBMC-OQAM auxiliary
10	0.2787	0.2787	0.2787
20	0.1206	0.1159	0.1121
30	0.035	0.02915	0.02559
40	0.02276	0.012	0.009559

Table 4 BER values for OFDM, FBMC-OQAM (data spread), and FBMC-OQAM (auxiliary method) with perfect CSI of one-tap equalizer

SNR (dB)	BER for OFDM data spread	BER for FBMC-OQAM data spread	BER for FBMC-OQAM auxiliary
10	0.2637	0.2637	0.2586
20	0.1426	0.1236	0.121
30	0.1022	0.07156	0.06903
40	0.0968	0.06395	0.06191

Table 5 BER values for one-tap (imperfect and perfect CSI), interference cancelation (no edges), and (imperfect and perfect CSI)

Iteration step 'i' values	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
One-tap	0.07216	0.07216	0.07216	0.07216	0.07216
One-tap (perfect CSI)	0.06444	0.06444	0.06444	0.06444	0.06444
Interference cancelation (no edges)	0.07603	0.02051	0.01463	0.01397	0.01449
Interference cancelation (imperfect CSI)	0.07216	0.02328	0.01633	0.0143	0.01297
Interference cancelation (perfect CSI)	0.06444	0.01572	0.0112	0.01016	0.01

6 Conclusion

The present FBMC system employs a single filter which is a unique modulation method known as MCC. This system concludes that any linear modulation scheme such as OFDM and FBMC-OQAM systems will benefit from the proposed doubly selective channel estimation and interference cancelation scheme. The auxiliary symbol method of the FBMC-OQAM system outperforms the data distributing method, and this method also has the additional benefit of a higher data rate. The proposed solution employs a poly-phase filter bank, such as analysis and synthesis filter banks, for efficient filtering. In terms of BER, the FBMC-OQAM system is dependent on the auxiliary symbol approach outperforms than OFDM System. Filter banks are used for both spectrum sensing and data transmission.

7 Future Work

In future, new channel estimation techniques and filtering characteristics can be utilized in FBMC-OQAM with large MIMO systems to improve the BER and MMSE. Further, the complexity of the system can be reduced.

References

1. Nissel R, Schwarz S, Rupp M (2017) Filter bank multicarrier modulation schemes for future mobile communications. IEEE J Sel Areas Commun 35(8):1768–1782
2. Nissel R (2017) Filter bank multicarrier modulation for future wireless systems. Dissertation, TU Wien [online]. Available https://publik.tuwien.ac.at/files/publik265168.pdf
3. Nissel R, Rupp M (2016) Enabling low-complexity MIMO in FBMCOQAM. In: IEEE Globecom workshops (GC Wkshps), Dec 2016
4. Nissel R, Blumenstein J, Rupp M (2017) Block frequency spreading: a method for low-complexity MIMO in FBMC-OQAM. In: IEEE SPAWC, July 2017
5. Lee J, Kim Y, Kwak Y, Li Y (2016) LTE-advanced in 3GPP rel—13/14: an evolution towards 5G. IEEE Trans Commun Mag Commun Stand Suppl: 36–42
6. Nissel R, Rupp M (2017) OFDM and FBMC-OQAM in doubly-selective channels: calculating the bit error probability. IEEE Commun Lett 21(6):1297–1300
7. Naveena A, Pranathi K (2021) A survey on channel estimation in multicarrier modulation schemes. Sci Technol Dev J X(III). ISSN: 0950-0707
8. Hwang T, Yang C, Wu G, Li S (2009) OFDM and its wireless applications: a survey. IEEE Trans Vehic Technol
9. Nissel R, Zachmann E, Lerch M, Caban S, Rupp M (2017) Low-latency MISO FBMC-OQAM: it works for millimeter waves! In: IEEE international microwave symposium, June 2017

10. Tanda M, Fusco T, Petrella A (2010) Joint symbol timing and CFO estimation for OFDM/OQAM systems in multipath channels. Eurasip J Adv Signal Process 2010:897607
11. Caekenberghe SV, Bourdoux A, Louveaux J (2014) Preamble-based frequency-domain joint CFO and STO estimation for OQAM-based filter bank multicarrier. EURASIP J Adv Signal Process
12. Farhang-Boroujeny B (2011) OFDM versus filter bank multicarrier. IEEE Trans Signal Process Mag 28(3):92–112

Stress Detection in Classroom Environment Using Physiological Data

A. Sharada , **Bhageshwari Ratkal** , **M. Lalitha** , **K. N. V. Samhitha, and A. Rishika Reddy**

Abstract In current scenario, mental stress has detrimental impacts on one's health. To mitigate detrimental ramifications, high-level stress must be detected early on. Following the introduction of wearable devices that could become a part of our daily lives, according to study, researchers have started detecting extreme stress in people who used them during their daily routines. An automatic stress detection system is developed by incorporating physiological signals obtained from unobtrusive clever, smart portable wearable devices that can be carried during individuals' daily life routines. Using various approaches of machine learning, successfully differentiated contest stress, relatively higher cognitive load (lecture), and relaxed time activities using heart activity, skin conductance and accelerometer signals, and skin temperature. Main objective is to develop a wrist band which identifies whether the student is under stress or not based on various parameters like heartrate and temperature of human body. This information can be obtained by employing various sensors such as an electrocardiogram (ECG), galvanic skin response (GSR).

Keywords Electrocardiogram (ECG) · Galvanic skin response (GSR) · Radio-frequency identification (RFID) · Health care · Stress · Anxiety

A. Sharada (✉) · B. Ratkal · M. Lalitha · K. N. V. Samhitha · A. Rishika Reddy
G. Narayanamma Institute of Technology and Science, Shaikpet, Hyderabad 500104, India
e-mail: sharada@gnits.ac.in

B. Ratkal
e-mail: bhagya.ratkal@gnits.ac.in

M. Lalitha
e-mail: mlalitha@gnits.ac.in

1   Introduction

Common methods to measure stress include questionnaires, where the mental effort that applicants put into a task is evaluated. However, such methods can be subjective, i.e., depending on the opinion of the applicants, based on physiological or personal measures. Therefore, these methods are not accurate enough due to individuals' inconsistencies. Moreover, this process becomes challenging when the number of individuals to be evaluated increases in real time. Thus, automated stress detection algorithms that can correctly recognize and assess stress even with many subjects are important in identifying stress factors and facilitating stress management [1]. These approaches embrace the use of portable devices such as mobiles, remotes, or wearable sensing devices to gather physiological signals. Currently, there are several smart wrist watches in the market such as Empatica E4, Apple watch, Fit Bit, but there is no smart wristwatch specifically designed to detect stress in a classroom environment. The proposed technique for detecting stress in the classroom is to build a wrist band that can forecast stress using continuous real-time data obtained from physiological sensors [2].

2   Literature Survey

One of the biggest health issues in the world is stress. Depending on the type of stress, stress has a different impact on human health. For patients with heart issues, excessive arousal can result in heart attacks, arrhythmias and even abrupt death. According to public surveys, at least half of European workers experience work-related stress. Additionally, it is believed that stress from the workplace and psychosocial risks account for at least half of the lost working days in the commercial sector [1].

   There are existing technologies for stress recognition. There are various reports on physical effects and on the detection of physical changes occurring as a result of mental stress [3].

2.1   Cardiac Activity

The cardiac cycles can be studied in depth with a Photo Plethysmo Gram (PPG). ECG data often has less noise than PPG sensor data. This is caused to abnormalities that are driven on by movement, ambient light, or varying skin tones, which change how well the skin can reflect and absorb light. As a outcome, ECG sensors are frequently preferred over PPG sensors. PPG sensors can be affixed to a person's fingers, wrists, or ears. Due to compact form influence of the sensory system, PPG modality has several uses in smart watches and fitness trackers [4].

2.2 Electro Dermal Activity (EDA)

The finger, palm, or wrist are examples of body parts where sweat glands are concentrated and where EDA is usually tested. To measure EDA, the resistance among each electrodes is measured [5].

2.3 Skin Temperature (TEMP)

A "fight or flight" response restricts blood flow to the extremities, making some changes in the peripheral temperature an intriguing indicator. These changes in TEMP are able to be measured either using infrared thermopile or temperature-dependent resistor.

2.4 Inertial Sensors

Inertial sensors, integrating three-axis acceleration (ACC), which are gyroscope and magnetometer, are commonly used in the human activity recognition (HAR). In the AR field readings, the ACC signal could provide context evidence about the physical activity of the user. ACC data is used to classify six different activity types (lying, sitting, and standing) as depicted [3].

3 Stress Detection System

The proposed technique for the efficient identification of stress in classroom environment is to develop a wrist band which can predict stress based on continuous real-life data collected from physiological sensors. The objective is to develop a band which identifies whether the user is under stress or not based on various parameters like heart rate, skin conductance, skin temperature, and movement of the human body. This product is mostly used by students when they are inside the classroom that is it is activated based on the user's location.

The student's ID is mapped with RFID generated by the band so that it will be compulsory that every student must wear it. The proposed model for stress detector is designed to enhance the generalization ability in stress detection of students in universities. Physiological sensor data plays a vital role in stress detection. The focus of this product is to integrate multiple sensors in a wristband for recording mainly physiological parameters and detect stress levels among users. This product mainly identifies the stress levels at a very early stage so that its effects can be reduced to a great extent [5]. The rationale behind selecting numerous sensors is, research

shown that decision-making by recognition systems is influenced by the accuracy of multimodal data which is often close to 10% higher than that of unimodal data. GSR detects the different conductance level of the skin when a person is in stress [5].

When a person is stressed, the nervous system responds quickly by releasing sweat. ECG is used for measuring electrical activity of the heart by inferring features of heart rate (variability). A smart band device is created specifically for this purpose in order to detect varied skin conductance levels, heart rate variability, and determine whether or not the person is under stress. However, based just on skin temperature and heart rate variability, daily stress levels cannot be correctly predicted. Physical activity like jogging, not getting enough sleep, etc. can also trigger the physiological reactions brought on by stress. Classification should be done in order to estimate the stress level appropriately [5].

This band will be able to identify stress by analyzing many skin conductance-related factors, such as activity tracking and body temperature. The data is then sent to a machine learning model for stress prediction, and the results are examined by the professors. Multimodal physiological dataset named WESAD for the purpose of stress detection is utilized [6]. This dataset has been introduced and made publicly available by Schmidt et al. This dataset is a collection of motion data and physiological data from 15 participants. Participants were put into various study protocol conditions such as meditation, recovery, baseline, amusement, stress, and their physiological stimuli were documented.

The student first scans their RFID card with the RFID reader as shown in (see Fig. 1). The RFID reader maps the unique ID of the RFID card with the students' roll number present in the csv file. After the student is granted access to the system, the data is collected from various sensors. The data is collected in a csv file which is passed to the ML model using flask. The prediction is made and the amount of time that perhaps the user is stressed is displayed as a report on the end web page. The procedure is as depicted in (see Fig. 2). Data can be collected from multiple watches simultaneously. The data can be collected for longer durations and the various stress inducing factors can be analyzed. After analyzing the factors, the stress inducing factors can be identified and removed. Predictions made by the machine learning model can be analyzed using the biological factors that affect the human body during stress. If the student is detected with stress for the first time, he will be under level 1 [7]. If he/she is identified with stress continuously for 2 or 3 days, special care is to be taken for them. Stress experienced by students can further be classified as acute or chronic.
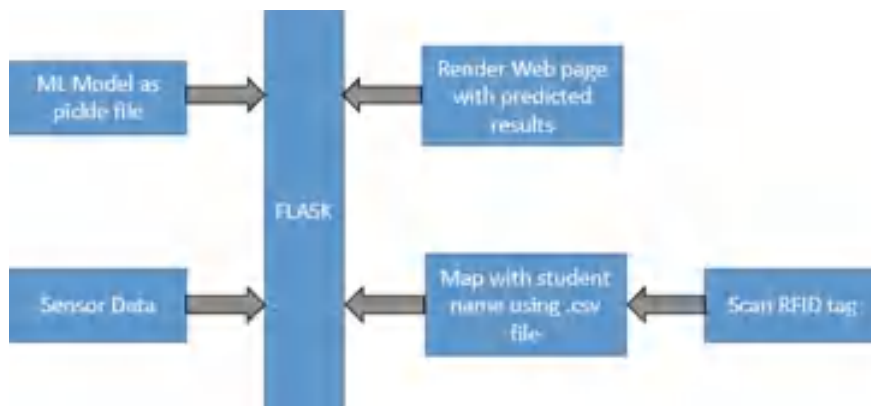
Fig. 1 Architecture of the band
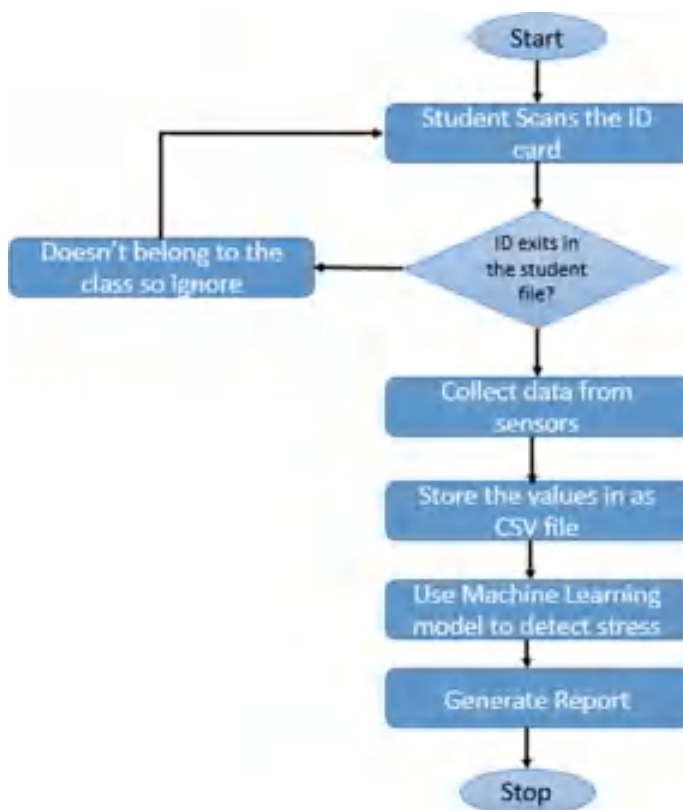
Fig. 2 Workflow of the band

4 Algorithms

In order to identify stress, a multimodal physiological dataset called WESAD is used. Schmidt et al. have introduced and made this dataset available to the general audience. This dataset consists of mobility and physiological information from 15 people. Participants underwent a variety of research protocol circumstances, including relaxation, recuperation, baseline, amusement, and stress, and their physiological responses were recorded. The following sensor modalities are included: blood volume pulse, ECG, EDA activity, electromyogram (EMG), respiration, body temperature, and three-axis acceleration. Two devices—a wrist-worn device (RespiBAN) and a chest-worn device (RespiBAN)—were used to record raw sensor data (Empatica E4). The RespiBAN gadget delivers the following sensor data: ECG, EDA, EMG, respiration, body temperature, and three-axis acceleration. The Empatica E4 gadget offers the ECG, (64 Hz), EDA (4 Hz), body temperature (4 Hz), and three-axis acceleration (32 Hz). In the data gathering, data from three different affective states (stress, amusement, and relaxed) were collected. The length of the stressed situation was 10 min, the amused situation was 6.5 min, and relaxed situation was 20 min.

Interquartile range has been applied to measure variability by dividing the dataset into quartiles. As per the rule of thumb observations can be qualified as outliers when they lie more than 1.5 Interquartile Range (IQR) below the first quartile or 1.5 IQR above the third quartile [8].

4.1 Decision Tree

The criterion used for decision tree classifiers is entropy. Entropy is a measure of the purity of the sub-split. Entropy always lies between 0 and 1. Entropy helps in the calculation of information gain, and information gain is the main key that is used by decision tree algorithms to construct a decision tree. Various pruning methods were applied on the decision tree, to avoid overfitting and improve the accuracy over the current data gathered from the students. The outcomes for the different combinations of the decision tree with entropy as criterion are displayed in the table.

Gini index is another criterion for the construction of decision trees. It is calculated by subtracting the sum of squared probabilities of each class from one. It favors larger partitions and easy to implement, whereas information gain favors smaller partitions with distinct values. As shown in Table 1, the combination of simple decision tree + normalization + cross-validation with criterion as Gini fetches the accuracy of 0.8661. Hence, it is seen that the combination of Simple decision tree + depth = 5 with criterion entropy produces the highest accuracy. The decision tree with depth 5 has the highest accuracy. The accuracy obtained is 93.82% for decision tree.

Table 1 Various combinations of decision tree with entropy as criterion

	Precision	Recall	F1-score	Support
Simple decision tree + depth = 5	0.96	0.92	0.94	245,784
Simple decision tree + depth = 5 + cross-validation + normalization	0.93	0.83	0.88	1,230,433
Simple decision tree + depth = 7 + normalization + cross-validation	0.91	0.84	0.88	1,230,433

4.2 Logistic Regression

The solver used in Logistic Regression algorithm is newton cg because it calculates Hessian explicitly which can be computationally expensive in high dimensions. Various performance measures were obtained: precision—0.8223, recall—0.8921, F1-score—0.8558, support—246,238. The accuracy obtained is 82.42% for Logistic Regression.

4.3 K-Nearest Neighbor

Various values for k are assumed and tested. The dataset is divided into k groups at random and then subjected to cross-validation. The rest of the groups are utilized as the training set, while one group is used as the test set. The training set is used to develop the model, while the test set is used to evaluate it. As soon as every distinct group has served as the test set, the procedure is repeated. The accuracy of the student data obtained is increased by the use of cross-validation. Table 2 shows the outcomes of implementing various KNN algorithm combinations.

The combination of KNN with k value as 3 along with cross-validation with number of holds as 10 and normalization gives the highest accuracy. The accuracy obtained is 85.57% for KNN model.

To choose the best algorithm, many evaluation metrics are required. Multiple machine learning methods are used to conduct the binary classification job, which discriminated between a stressed condition and a non-stressed state. Using the WESAD dataset, the performances of several classifiers were examined, and it was shown that machine learning techniques including the decision tree, logistic

Table 2 Various combinations of KNN algorithms

	Precision	Recall	F1-score	Support
KNN + cross-validation + K = 20	0.8295	0.8213	0.8720	1,230,433
KNN + cross-validation + K = 10 + normalization	0.8249	0.8200	0.8723	1,230,433
KNN + cross-validation + K = 3 + normalization	0.9245	0.8220	0.8802	1,230,433

Table 3 Comparison of algorithms

	Precision	Recall	F1-score	Support
K-nearest neighbors	0.8379	0.883	0.8564	2,100,456
Logistic Regression	0.8421	0.8028	0.8128	420,092
Decision tree	0.9341	0.9442	0.9382	420,092
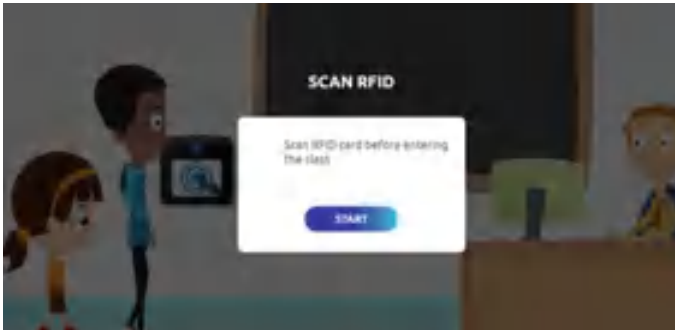
Fig. 3 Home page

regression, and K-nearest neighbor performed well. The greatest performance of the machine learning methods for binary classification is a decision tree (93.82%), which is better than the results of Logistic Regression (82.42%) and K-nearest neighbor (85.57%). The F1-score and accuracy are employed as evaluation criteria.

5 Results

Various performance measures were considered for the proposed work. The two metrics above precision and recall can be combined into a single metric called F-measure. It is a harmonic mean of precision and recall. Overall, decision tree has the highest accuracy as shown in Table 3.

The home pagein (see Fig. 3) is where the student is directed to scan the RFID card with the RFID reader. Data gathered from the sensors is collected in a csv file and the time duration is displayed for which the user is stressed (see Fig. 4).

Fig. 4 Report generated

6 Conclusion

A stress detection system is created for the classroom. The created system is simple for people to utilize in their daily lives because it uses unobtrusive wearable gadgets. It can monitor the amount of stress in real-time. Continuous stress detection using sensors of a smart watch was studied in this proposed work. Every RFID tag has a unique ID which is mapped with the corresponding student roll number. After gaining access, the data from various sensors in the wristwatch is collected and passed to a ML model built using the WESAD dataset. The ML model is integrated with the flask application which renders a webpage which displays the stress report. The stress report displays the amount of time the user was stressed [3].

References

1. Gjoreski H, Kozina S, Gams M, Luštrek M, Álvarez-García JA, Hong JH, Ramos J, Dey AK, Bocca M, Patwari N (2015) Competitive live evaluation of activity-recognition systems. IEEE Pervasive Comput 14(1):70–77
2. Sharada A (2020) Machine learning approach for stress detection using bio signals, robotics design and applications using wireless sensor networks, IoT & AI. IFERP
3. Gjoreski H, Kalužaa B, Gams M, Milić R, Luštrek M (2015) Context-based ensemble method for human energy expenditure estimation. Appl Soft Comput 37:960–970
4. Siirtola P, Koskimäki H, Röning J (2018) Personalising human activity recognition models using incremental learning. In: European symposium on artificial neural networks, computational intelligence and machine learning (ESANN), pp 627–632, Apr 2018
5. Incel O, Kose M, Ersoy C (2013) A review and taxonomy of activity recognition on mobile phones. Bio Nano Sci 3(2):145–171
6. Bobade P, Vani M (2020) Stress detection with machine learning and deep learning using multimodal physiological data. In: Second international conference on inventive research in computing applications

7. Handouzi W, Maaoui C, Pruski A, Moussaoui A (2014) Short-term anxiety recognition from blood volume pulse signal. In: IEEE 11th international multi conference on system signals devices, SSD 2014
8. Li R, Liu Z (2020) Stress detection using deep neural networks. BMC Med Inform Decis Making 20:285

Prediction of Next Words Using Sequence Generators and Deep Learning Techniques

P. Sunitha Devi, Chepuri Sai Tejaswini, Modem Keerthana, Manusree Cheruvu, and Minati Srinivas

Abstract Would not it be wonderful if your computer could anticipate what you may type as your next word? For a certain user's texting or typing, the next word prediction can be very helpful. People would be more productive as a result of the significant time savings. Predictive text is an input technology that facilitates typing on a mobile device by suggesting words that the end user may wish to insert in a text field. Predictions are based on the context of other words in the message and the first letters typed. We intend to bring this prediction to the next level. Recurrent neural networks are used in prediction of next word which is a neural application. Standard RNNs can be trained to tackle some issues, but it can be challenging to teach them to learn long-term temporal dependencies. LSTM networks are applied to solve this problem. By advancing the technology now in use, we hope to anticipate the next words that could fit the statement and create a user-friendly application for it.

Keywords Word prediction · LSTM · Bigram · Trigram · Sequence generator · Neural networks

1 Introduction

A likelihood distribution over word sequences may serve as an application of mathematics language model. When given a sequence of this type, let us say one of length m, the language model uses context to distinguish between terms and sentences that sound similar and assigns a likelihood $P(w1, w2, wm)$ to the entire series. For instance, in English, the words "recognise speech" and "destroy a nice beach" seem similar but mean two entirely different things. Language modeling is a problem that affects both natural language processing (NLP) and language comprehension (LM). In addition to encoding linguistic complexity like grammatical structure, models that can reliably distribute distributions over sentences also distill a significant amount

P. Sunitha Devi (✉) · C. S. Tejaswini · M. Keerthana · M. Cheruvu · M. Srinivas
G. Narayanamma Institute of Technology and Science, Hyderabad, India
e-mail: sunitha@gnits.ac.in

of information about the knowledge that some corpora may contain. The past few years have seen a significant advancement in language modeling analysis thanks to deep learning and recurrent neural networks (RNNs), which made it possible for academics to investigate a number of tasks that the robust conditional independence assumptions' area failed to fulfill. Even though N-grams are simpler models that only use a small sample of previous words to predict the future word, they are nevertheless an essential part of high-quality, low mental confusion language models. In reality, the most recent research on large-scale language models has shown that RNNs perform better when combined with N-grams because they have special features that enhance N-gram models, but when taken separately, they perform badly. Recurrent neural networks have networks with loops that enable the persistence of data. The current algorithm can only anticipate the potential next single word, not the potential next set of words. It is unable to assist the user in finishing simple words or even in creating an insightful paragraph. The main focus is to predict the next possible words that suit the sentence with the help of the suggested technologies, also by enhancing the current technology, and then create a user-friendly application.

2   Related Work

Predicting the words that will probably follow a certain text section is known as word prediction. Language modeling is the technique of predicting the next word. RNN comes into play because conventional neural networks lack the capacity to store large amounts of data. RNNs are networks with loops that aid in information storage. However, the issue arises when there are lengthy phrases and you have to go back and find the information. To solve the issue of enduring dependencies, the LSTM enters the picture [1]. A language model is created that predicts the likely next word given a group of existing words, mostly using the foundation for quick electronic communication. Word prediction algorithms must make informed judgments based on the previous word that is likely to continue with the few starting text fragments available. Their main objective was to assist the user in quick electronic communication by recommending appropriate terms. The aforementioned existing systems utilize machine learning algorithms that are limited in their ability to produce acceptable syntax and operate on a word prediction model that suggests the next word to be used immediately depending on the word that came before it. In addition to enforcing the multi-window convolution (MRN N) formula, they also developed a residual-connected lowest gated unit (MGU), a condensed version of the LSTM, in order to shorten training time and increase accuracy. However, using multiple layers of neural networks will result in latency when predicting n words. Models created using bidirectional LSTM algorithms are capable of handling more information quickly and making better predictions, making developing technologies more accurate than the currently used system technologies. The limitations of NLP include the necessity to apply various types of pattern discovery techniques targeted at removing noisy

data, big files or datasets that are difficult to process and yet require some optimization, and forming proper syntax [2]. Another piece of work shows how the system uses several processes including the closed-loop TensorFlow system to forecast and correct the system's next/target phrases. They made an attempt to create a model using Nietzsche's default text database that would predict the client's sentence after the client had written 40 letters. The model would understand 40 letters and predict the next seven to ten words using RNN neural organization, which would be carried out using TensorFlow. Their goal in developing this model is to anticipate 10 or more words as quickly as they can in the shortest amount of time. They chose a dataset from a Franz Kafka book that contains 25,107 words. To "encourage the creation and diffusion of eBooks," they looked into Project Gutenberg, a voluntary effort to digitize and archive cultural works, and were successful in finding databases for text data. From there, they gathered a multitude of accounts, documents, and textual data that was essential to their thesis. The major aim was to develop and test the best-fitting algorithm, and they mostly anticipated using an LSTM to achieve high accuracy. They developed a 3D vector layer for input and a 2D vector layer for output and fed the LSTM layer with 128 hidden layers in order to comprehend the issue statement at hand. The model's limitations include its inability to foretell how a paragraph would be composed of sentences. To train the model, which will review the weights to comprehend the fundamental characteristics of paragraphs and phrases and accurately forecast outcomes, more data are needed.

3 Existing Systems

Predictive text curates its predictions on texts that you use over time. It builds its glossary of words based on words and phrases that you type repeatedly. The system will then work to score these words by the probability of when you will use them again. The existing work possess various drawbacks and are addressed in the proposed system. An advanced artificial intelligence program called Grammarly is used to review English-language sentences. When we use *Grammarly* to write, the AI examines each sentence and notes all the areas for improvement, such as changing the verb tense, offering a more potent synonym, or providing a cleaner sentence structure. *WhatsApp* uses the input word entered using the keyboard to recommend the following terms that would be most suited. The existing systems only predict one word at a time; a collection of words which in turn can help the user to auto-complete the sentence is not done, which is a drawback.

4 Methodology

To build the main model that is to predict the next words, LSTM method is used whose architecture is shown in Fig. 1. A unique class of RNNs called long short-term memory networks is able to learn long-term dependencies. All recurrent neural networks have the shape of a series of neural network modules that repeat. This recurring module in typical RNNs will be made up of just one tanh layer, for example. Although the repeating module of LSTMs also has a chain-like topology, it is structured differently. There are four neural network layers instead of just one, and they interact in a very unique way. The dataset must be transformed into a series of numbers that can be represented by bi and trigrams in order to train the model. A trigram is three consecutive words in a sentence, while a bigram is two words in a row.

4.1 Dataset

The dataset is collected from different resources considering a particular domain. The domain that is considered is "Environment." The dataset contains data in the form of text sentences which is an aggregation of data from different related articles. The dataset contains 250 sentences.
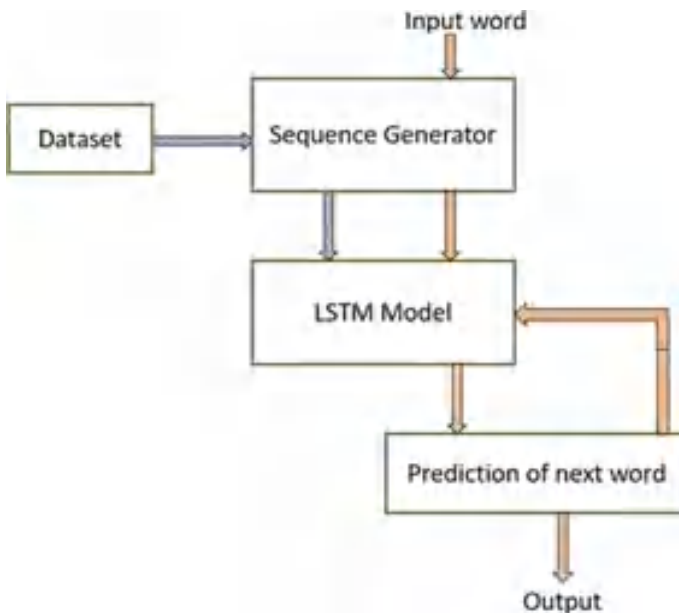
Fig. 1 Proposed workflow

4.2 Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. The created dataset contains set of statements. As the task is about the prediction of the next word, the preprocessing phase contains steps such as converting all the words into lowercases and converting the sentences in the dataset to sequence of tokens. The conversion of the dataset to tokens as the English vocabulary is not the one which a machine generally can understand, so tokens are converted to sequence of integers.

4.3 Building the Model

As LSTM is long short-time memory as shown in Fig. 2, it will understand the past text and predict the words which may be helpful for the user to frame sentences. Take the phrase "I grew raised in the U., I speak fluent English" as an example. It implies that the following word is probably the language's name, and if we want to know which language it is, we must return to the previous sentence since we need the context of the US. To address the issue of enduring dependence, the LSTM enters the picture.

Recurrent neural networks (RNNs) have a long-term dependency issue that LSTM networks were created expressly to address (due to the vanishing gradient problem). LSTMs differ from more conventional feedforward neural networks in that they feature feedback connections. With the help of this property, LSTMs may process whole data sequences (such as time series) without considering each data point separately. Instead, they can process new data points by using the information from earlier
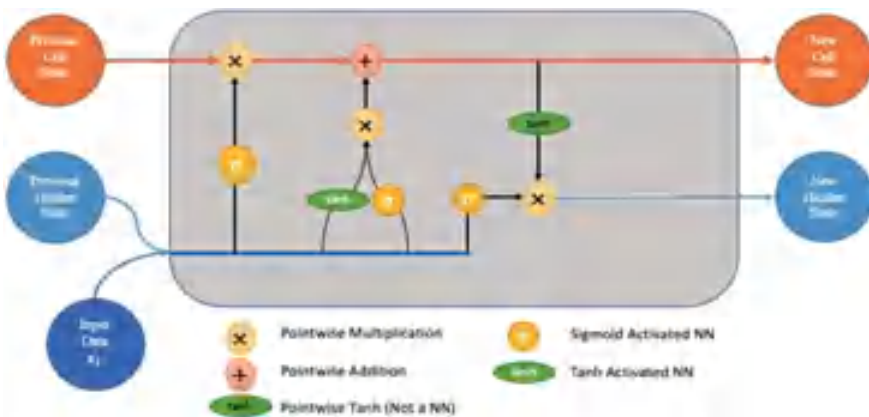
Fig. 2 LSTM architecture

data in the sequence. Because of this, LSTMs excel at processing data sequences like text, audio, and general time-series.

4.4 Predicting the Words

After the model has been trained, the words are predicted by factoring in the preceding two words. In order for the model to return the words with a higher degree of probability, the two words are taken into account and converted to sequences. The couple of words are then predicted by taking into account the current output word and the last previous word, and for bigram, the previous word is taken into account.

5   Implementation

5.1 Dataset

Data for this study were gathered from a variety of environmental articles. A dataset is created by combining the data from other articles. By predicting the next word in the sentences, this dataset of sentences has been utilized to comprehend how the model is used to predict the next word.

5.2 Preprocessing

The dataset is kept in a text file with the name as data.txt. The utf-8 encoding is used to retrieve the data from the.txt file. Tokenization is the subsequent preprocessing step. It alludes to dividing the larger text data, articles, or corpora into more manageable chunks. These more compact units may take the form of shorter documents or text data lines. They might also function as a word dictionary. By converting each text into a series of numbers as shown in Fig. 3, each of which represents the index of a token in a dictionary, and the Keras Tokenizer enables us to vectorize a text corpus. The texts will subsequently be turned into sequences. This method turns the text input into numbers so that we can conduct more accurate evaluations on them.

After tokenization is performed, sequences of length 3 are generated. Out of which, the first two words correspond to the input and the third word corresponds to the output as shown in Fig. 4.

Fig. 3 Tokenization

WORD	I	INTEGER
Industrialization		136
took		256
birth		257
around		137
the		1
18th		258
century.		259
This		11
is		7
when		71

Fig. 4 Sequence generation

Total Sequences:1956

$$\begin{array}{ccc} \text{INPUT} & : & \text{OUTPUT} \\ [136\ 256] & & 257 \\ [256\ 257] & & 137 \\ [257\ 137] & & 1 \\ [137\ \ \ 1] & & 258 \\ [\ \ 1\ 258] & & 259 \\ [258\ 259] & & 11 \end{array}$$

5.3 Building the Model

A sequential model is created as shown in Fig. 5. Next, the input and output dimensions are set for an embedding layer. One of the layers in Keras is the embedding layer. This is primarily utilized in NLP-related applications, such as language modeling, but it can also be applied to other neural network-based tasks. It enables to transform each word into a vector with a specified length and a predetermined size. The resultant vector is a dense one with real values instead of just 0's and 1's. The fixed length of word vectors helps us to represent words in a better way along with reduced dimensions. Given that the prediction will be made on just one word and that word will be the subject of the response, it is crucial to define the input length as 2. After that, design is extended with an LSTM layer. We will feed it through an 100 LSTM units for the subsequent layer.

1. Creating an embedding layer (input layer) with a specified input and output dimension.
2. Then, a LSTM layer is added with 100 neurons.
3. Dense layer is created with vocab_size as the dimension of output space and "softmax" activation function is used.

Finally, an output layer is created with a Softmax activation and the chosen vocabulary size. The dense layer is the typical layer of a neural network with many connections. The Softmax activation makes sure that a large number of probabilities are

```
Model: "sequential"
_____
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 2, 10)             7580
_____
lstm (LSTM)                  (None, 100)               44400
_____
dense (Dense)                (None, 758)               76558
=================================================================
Total params: 128,538
Trainable params: 128,538
Non-trainable params: 0
_____
None
```

Fig. 5 Sequential model

received for outputs equal vocabulary size. The last phase involves compiling and fitting the model. The model is trained and it is saved in .h5 format so that the model can be used as needed and training the model repeatedly can be avoided. Here, training data are used to train the model. You can, however, decide to train using both validation and test data. The categorical cross-entropy is a loss function that calculates the cross-entropy loss between labels and predictions. The model is built using Adam, an optimizer with a 0.001 learning rate. The loss and accuracy for each epoch are shown in Fig. 6.

```
Epoch 1/250
1956/1956 [==============================] - 1s 372us/sample - loss: 6.6023 - acc: 0.0506
Epoch 2/250
1956/1956 [==============================] - 0s 140us/sample - loss: 6.1315 - acc: 0.0547
Epoch 3/250
1956/1956 [==============================] - 0s 130us/sample - loss: 5.7829 - acc: 0.0547
Epoch 4/250
1956/1956 [==============================] - 0s 117us/sample - loss: 5.7167 - acc: 0.0547
Epoch 5/250
1956/1956 [==============================] - 0s 120us/sample - loss: 5.6607 - acc: 0.0547
Epoch 6/250
1956/1956 [==============================] - 0s 120us/sample - loss: 5.6129 - acc: 0.0547
Epoch 7/250
1956/1956 [==============================] - 0s 128us/sample - loss: 5.5713 - acc: 0.0547
```

Fig. 6 Loss and accuracy at each epoch

6 Predicting the Words

For the prediction, the h5-formatted tokenizer file is loaded that was previously stored. The next word model that has been saved will then be loaded. Each of the input sentences for which predictions must be made will be tokenized using the same tokenizer. Following this, the loaded model is used to generate predictions about the input phrase. Since the activation function that is used is Softmax which allows a probability-like output, it outputs the most probable next word for the given sequence of words. The model takes two words as input and predicts the next probable word as output. Since the model is expected to output the next possible three words, the present output and the last word of the previous input are considered as new input to predict the consecutive word. While running the predictions, try and except statements are used. These statements are used because we want the program to end the loop if the input word is out of domain. The model is then integrated with an application using Django to make it user-friendly, enabling the trained model to assist the user in finishing the task. The model makes the predictions on the interface as soon as the user enters the word and "SPACE" is encountered.

7 Result

7.1 Interface

A user-friendly application is built as shown in Fig. 7 which helps the user to write an article on Environment. A text area is provided where the user types a word and next predictions are made. Figure 8 depicts that as soon as the user types in a word and a space is pressed, the possible prediction of couple of words is listed at the bottom and it auto completes the sentence when a one of the possible options is selected.

Based on the accuracy and loss factors that are taken into account for each epoch, the performance of this model can be evaluated. One loop across the entire training dataset is referred as an epoch. A neural network often requires more than a few epochs to train. The model is trained with 250 epochs. The highest accuracy for a bigram as shown in Fig. 9 is at the 55th epoch, whereas for a trigram as shown in Fig. 10 is at the 45th epoch. As, in comparison we can conclude that in trigram we get to the maximum accuracy faster when the total number of epochs are 250.
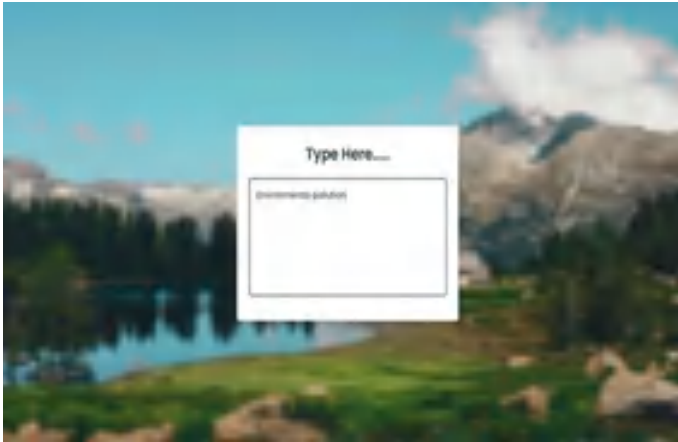
Fig. 7 Application interface
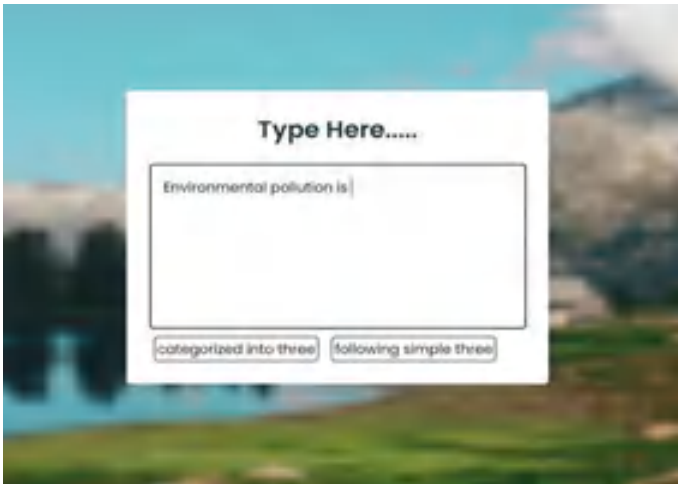
Fig. 8 Predictions at the bottom of text area

8   Conclusion

We were able to create a best next word prediction model for the environment dataset. In 250 epochs, considerable reduce in the loss is observed. On the available dataset, the next word prediction model that is created is fairly accurate. The results show that trigrams give better predictions when compared to the bigrams. The project would help the user to formulate the articles for the environment domain. This model is integrated with an application.
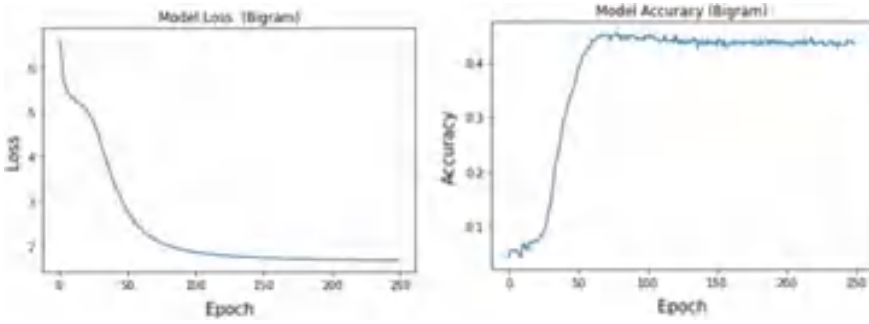
Fig. 9 Graphs related to bigram
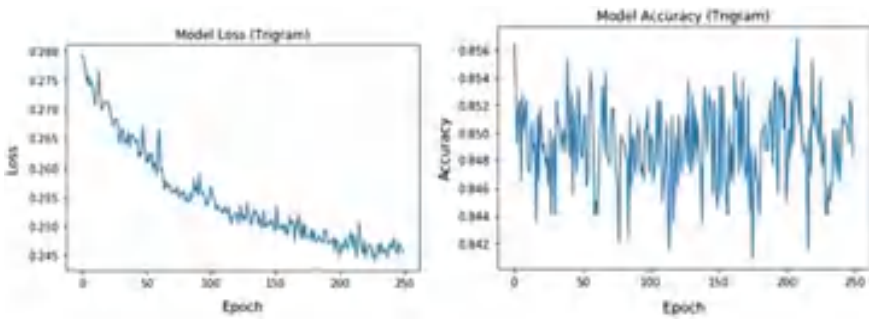
Fig. 10 Graphs related to trigram

9 Enhancements

A machine learning model to predict the next word in a given sequence of words was built using LSTM models. It was then extended to predict the next few words in the same sequence. The model developed can be used for predicting the next word in the different languages. This can effectively reduce the number of words that the user has to type, thus increasing the typing speed of the text inputs. It also helps in minimizing the spelling mistakes that are inputted by the user. The system can be extended for different language generation tasks like story auto-completion, poem auto-completion, etc. The model is limited to a specific dataset and more randomness can be incorporated into the model by enhancing the scope. The system can be adapted to new words that are not part of its vocabulary. This adaptation will be done when the model encounters a new word and adds the word to the vocabulary. This way the model becomes more generalized. By remembering the history of the user's model can become personalized. The system can be personalized to predict words based on the user's history. N-gram is the enhancement that can be made. By only looking at two to three words, the N-gram technique is a straightforward but effective way to represent language. Due to its simple and quick computation, it is highly useful in

various applications, such as word suggestion on mobile devices. However, N-gram is not reliable when it comes to creating lengthy sentences in which the current word is dependent upon the very first word. However, this has the drawback of taking 100 times longer for each epoch than bigram and trigram.

References

1. Keerthana N, Harikrishnan S, Konsaha Buji M, Jona JB (2019) Ego-vehicle speed prediction using a long short-term memory based recurrent neural network. Int J Autom Technol 20:713
2. Ambulgekar S, Malewadikar S, Garande R, Joshi B. Next words prediction using recurrent neural network
3. Asma Rashid HK, Saeed SA, Rashid TA (2021) Next word prediction based on the Ngram model for Kurdish Sorani and Kur Manji. Neural Compute Appl 33:4547–4566
4. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv:1409.0473
5. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5:157–166
6. Ganai F, Khursheed F (2019) Predicting next word using RNN and LSTM cells: statistical language modeling. In: 2019 fifth international conference on image information processing (ICIIP), pp 469–474
7. Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: continual prediction with LSTM
8. Graves A (2013) Generating sequences with recurrent neural networks. arXiv:1308.0850
9. Habib M, Al-Mamun AR, Siddiquee M, Shah Ahmed F (2018) An exploratory approach to find a novel metric based optimum language model for automatic Bangla word prediction
10. Hochreiter S, Schmidhuber J (1997) LSTM can solve hard long time lag problems. In: Advances in neural information processing systems, pp 473–479
11. Yang J, Wang H, Guo K. Natural language word prediction model based on multiwindow convolution and residual network. IEEE Access
12. Terada K, Watanobe Y (2019) Code completion for programming education based on recurrent neural network. In: 2019 IEEE 11th international workshop on computational intelligence and applications (IWCIA)
13. Mikolov T, Karafiát M, Burget L, Cernocký J, Khudanpur S (2010) Recurrent neural network based language model. In: Eleventh annual conference of the international speech communication association, 29
14. Wang M, Liu W, Zhong Y (1993) Simple recurrent network for Chinese word prediction. In: Proceedings of 1993 international conference on neural networks (IJCNN93-Nagoya, Japan), Nagoya, Japan, pp 263–226
15. Olah C (2015) Understanding lstm networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/
16. Barman PP, Boruah A. A RNN based Approach for next word prediction in Assamese phonetic transcription. Procedia Comput Sci
17. Sharma R, Goel N, Aggarwal N, Kaur P, Prakash C (2019) Next word prediction in Hindi using deep learning techniques. Int Conf Data Sci Eng (ICDSE) 2019:55–60
18. Gosavi R, Konduskar P, Chavan P, Kundale J (2014) Next word prediction using reccurent neural network
19. Sarwar SM, Abdullah-Al-Mamun (2016) Next word prediction for phonetic typing by grouping language models. In: 2016 2nd international conference on information management (ICIM), London
20. Siami-Namini S, Tavakoli N, Nami AS (2019) The performance of LSTM and BILSTM in forecasting time series. IEEE Int Conf Big Data (Big Data) 2019:3285–3292

Natural Scene Text Detection in Video with Hybrid Text Augmentation and Fusion-Transferred Learning

Mortha Manasa Devi, Maddala Seetha, and S. Vishwanadha Raju

Abstract We propose a robust approach of region proposal network through graph-based approach and hybrid text augmentation for detecting and recognizing the video text with different languages, fonts, complex background, and natural scene patterns. First, we use diverse regions using region proposal network (RPN) to identify the text that positioned in different location in the video with different sizes and scales. The locations are identified by segmenting the regions with similarity measure through graph-based approach that adds ability to find the correctness of the text locations. Then, along with the text augmentation, the ability of RPN is improved in locating and classifying the text by finding the bounding box coordinates. Second, a classification network though transfer learning from VGG19 is adopted to eradicate the false positives. Finally, we developed the fusion technique to obtain a clean scene text layer and verified the correctness of text by optical character recognition techniques.

1 Introduction

With the swift development of smart devices, automation, communication, e-learning, video has become the prominent medium. At the same time, with this medium, the increase of uploads in the social media recommends for efficient video

M. M. Devi (✉)
Ascent Fund Services, Bangalore, India
e-mail: manasadevinoolu@gmail.com

M. Seetha
Department of Computer Science and Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad 500008, India

S. Vishwanadha Raju
Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Jagityal 500085, India

indexing and retrieval, localization of preferred content from the colossal videos. Previous work on text detection from video [1–3] proposed various low-level techniques that perform well on edges, corners, textures. Though these methods proved to be effective, but it is highly recommended to have a more significant approach to handle the text in video with varied fonts, backgrounds, distortions, etc. [4], which applied SIFT and MSER to add popular on detecting the text for efficiency in performance and speed. Actually, there are two types of text in video [5]. Among these two, scene text usually depicts important information than caption text [6]. Mortha et al. [7] proposed convolution neural network (CNN) with unsupervised feature learning to detect and recognize the text in video. Transfer learning is outperforming the existing architectures like VGG16 [8], ResNet50 [9], and InceptionV3 [10] with its knowledge transfer approach with high performance. In order to control the nonlinear systems, fuzzy systems are recommended for powerful control methodology. Fuzzy c-means clustering algorithm (FCM) [11] performs well with text extraction from images. With the inspiration of these approaches and observations, we propose a novel approach for scene text detection and extraction of text features using graph-based segmentation and text augmentation. Later, the accuracy of detected text is computed by transfer learning classifier and finally separates the text candidate form background with fusion approach.

2 Related Work

There are diverse text detection techniques to locate and identify from the image like connected components based technique and sliding window technique. Based on these techniques, pixels with similar text features such as Maximally Stable External Regions (MSER) [12] and Stroke Width Transform (SWT) [13, 14] are widely used for localization of text components. Sliding window methods [15] define a sub-window to pass over images on all locations to extract the text features and utilize the classifiers like Support Vector Machines (SVMs). Later, with the evolution of neural networks such as convolution neural network (CNN) and recurrent neural network (RNN) have changed the working principle for scene text detection [16, 17]. Zhang et al. [18] exploited CNN to extract the text by eradicating false positives with the symmetrical property of text characteristics. The state of the robust techniques proposed by [19] applied the combination of SegLink and component level CNN classifier. It finds the correlation between the text and background through bipartite graph. Liang et al. [20] presented FOTS framework-oriented scene text detection. A novel approach of RoIRotate influenced the network to recognize the text efficiently. Zhang et al. [21] proposed multi-scale fusion to handle the lost features from pooling layer in order to predict the more predicted text box by mitigating the low ranking boxes.

3 Proposed Approach

Initially, frames are decoded from video and classified into five classes: grid reference, column, horizontal beam, vertical beam, and sloped beam. Then, a series of image preprocessing operations are applied to decrease the noise and improve the quality of frames for the detection step. Corner features are utilized which reflect the areas where text is existing. With the application of grayscale morphological processing and adaptive threshold binarization, complete text regions are obtained through corner features. But after observing the downsides of this approach toward the video text, then eventually tried with other techniques to justify proposed pipeline with robust process. Hence, first we verified our dataset with [22] and then using SegLink [19] and EAST [23]. Later, fusion transfer learning approach is used to verify the text regions in input frames with high-level features. The framework is evaluated with test dataset and compared to existing algorithms.

3.1 *Cited Techniques for Text Detection*

Even though edges and textures provision to find the text in video, they are not feasible solution because of false detection of other edges and increase the false positives. We have to find the technique which is rotation invariant, translation, and illumination invariant. To achieve this, corner features are important and interest points which ensure the same characteristics. With corners, there would be a significant gradient change in all directions. It means, shifting the window in any direction yields a large change in appearance and also more stable over the change of viewpoint. It is contradicted with edge and flat region. This is observed in Fig. 1a. Also, the overall workflow is shown in Fig. 1b.

Flat region
No changes in
all directions

Edge region
No change along
the Edge direction

Corner region
Significant changes in
all directions

Video Frames → Feature Extraction → Geometric Transformation → Text Detection
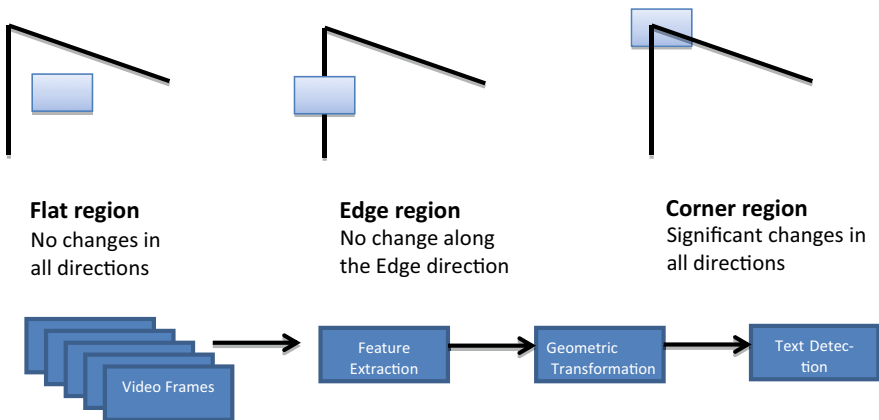
Fig. 1 **a** Direction of an object with flat, edge, and corner regions, **b** workflow

The main idea behind the corner detection technique is to identify the pixels with unique features. That means, small window is defined to compute around the pixel p in a frame and find all the pixels with unique property with respect to change in direction. Video text always has a regular distribution of corner points compared to background. The detailed mathematical intuition is given in [22]. So, initially the grayscale frames I that were obtained from preprocessing phase are taken over the window $W(x, y)$ and shifted by (u, v) to compute the change produced by the shift. This is done by using Eq. (1):

$$E(u, v) = \sum_W [I(x + u, y + v) - I(x, y)^2.$$

(1)

Here, $E(u, v)$ is the sum of all the sum squared differences (SSDs), where (u, v) are the x, y coordinates of every pixel in 3×3 window and I is the intensity value of the pixel. We check for the large SSD value that indicates changes in pixels' values in eight directions before and after the shift, i.e., $E(u, v)$ must be maximized for corner detection. That means, the second term has to be maximized. With the application of Taylor Expansion to Eq. (1) using few mathematical steps, we get the final equation as (2).

$$E(u, v) = [uv] \left(\sum \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right) \begin{bmatrix} u \\ v \end{bmatrix}.$$

(2)

Now, renaming the summed matrix as M,

$$M = \sum w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}.$$

(3)

When Eq. (2) is substituted in Eq. (1), then it becomes:

$$E(u, v) = [uv] M \begin{bmatrix} u \\ v \end{bmatrix}.$$

(4)

Here, M is called as Hessian matrix. If the two Eigen values of M are large and different positive values, a shift in any direction would cause a significant increase and a corner could be determined.

$$R = \det M - k(\text{trace } M)^2$$
$$\det M = \lambda_1 \lambda_2$$
$$\text{trace } M = \lambda_1 + \lambda_2$$

(5)

λ_1 and λ_2 are the Eigen values of M. These values decide whether a region is corner, edge, or flat. When |R| is small, the region is flat for small values of λ_1 and λ_2. When

$R < 0$, i.e., $\lambda_1 \gg \lambda_2$, the region is edge. Finally, when R is large, i.e., λ_1 and λ_2 are large and $\lambda_1 \sim \lambda_2$, the region is corner. In Eq. (5), k is tunable parameter.

From Fig. 2, there are corner points in text and non-text regions. From [22], Harris corner detector is utilized to extract the corner points, and with the series of morphological dilation operations, target text regions are formed. The perplexing point here is to choose the appropriate kernel value of dilation to detect the region of interest. If kernel value is too high, the interested points would collide with non-interest points; if kernel value is small, then we may lose the information to detect the desired one. To overcome this, we adopted combination of morphological operations and adaptive thresholding (Fig. 3) with corner features to obtain text targets accurately.

As scene text agrees with background, it is not simple to perform basic morphological operations to extract text region from the frames. This is because scene text comes with the background. Thus, it cannot be separated from background as done with the caption text. Hence, we employed character region awareness for text detection (CRAFT) technique to detect the text region by exploring each character region and affinity between the characters. Then, bounding box is obtained though minimum bounding rectangles on bitmap after the process of thresholding and affinity scores followed by a structure to learn the ground truth when data are low to identify.

As SegLink [19] works on two junctures, first is segmentation and second is links. Segmentation technique is used to segment the frames into chunks or regions. But this also generates false segments that represent non-text or text-alike characteristics. This approach improves the drawbacks of maximal regions. However, SegLink from
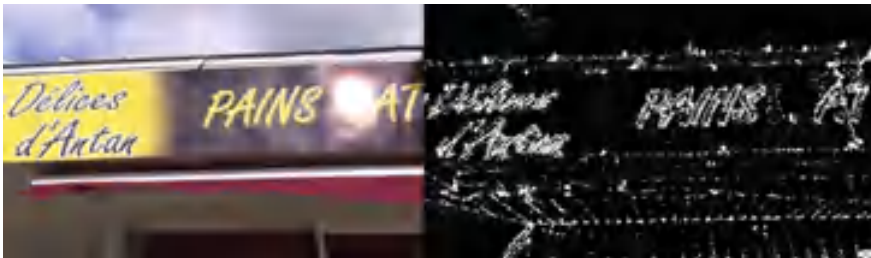
Fig. 2 Corner features of video frames

Fig. 3 Adaptive thresholding

Fig. 4 does not perform well on low illuminated and complex background because of lack of pixel data to connect the text points. Then, verified our dataset with EAST [23] which is robust detector that works on images and videos. It is observed that EAST cannot detect the text that is far, blurred. It is shown in Fig. 5. As from the figure, it can be seen that even the characters are visible, the model is unable to detect. And, it is clearly shown in Fig. 5 that few bounding boxes are generated even when no text is present. This proves that there are still false positives which are susceptible to less accuracy and demands for more robust algorithm to avoid such breakdowns.

To overcome these downsides of above-cited methodologies, proposed methodology utilizes the annotation data of the text candidates. Annotation helps to locate the text with the predefined bounding box information in $(x1, y1)$, $(x2, y2)$, $(x3, y3)$, $(x4, y4)$ format, where (x_i, y_i) denotes the four corners of the rectangle box around the text. But, this type of coordinates fails to retrieve the exact location of the text in video because of moving property. The position would be same for every frame irrespective of text location. To avoid this, we have converted these coordinates into $(x, y, width, height)$, where (x, y) is point and (width, height) gives the m (width) pixels wide and n (height) pixels high from the (x, y) position to embed the text within the desired bounding coordinates. Even though is it beneficial to obtain the desired text through annotation data, still it cannot ensure for the true positives because of DON'T CARES. Don't cares are those texts with invisible nature, poorly annotated, blurred. Hence, with additional preprocessing techniques, the proposed
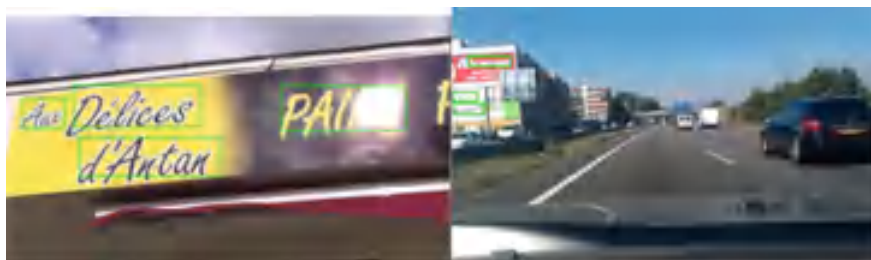
Fig. 4 SegLink technique

Fig. 5 EAST technique

approach tries to extract the DON'T CARES. Initially, we took all the (x_i, y_i), where, i varies from 1 to 4. These points are converted to $(x, y, \text{width}, \text{height})$. We considered top left coordinates as (x, y), i.e., $(x1, y1)$, and with the following Eqs. (6 and 7), corresponding width (w) and height (h) are computed using Euclidean distance.

$$w = \sqrt{(P_2[0] - P_1[0])^2 + (P_2[1] - P_1[1])^2}, \tag{6}$$

$$h = \sqrt{(P_3[0] - P_2[0])^2 + (P_3[1] - P_2[1])^2}. \tag{7}$$

Here, $P_1 = (x1, y1)$, $P_2 = (x2, y2)$, $P_3 = (x3, y3)$, $P_4 = (x4, y4)$.

After obtaining these four points to determine the bounding box for a text, we have verified by applying Faster RCNN [24] approach to validate the accuracy score for number of texts detected with proposed dataset. Figure 6 shows that this approach works well by finding the anchor boxes for possible positions though defined aspect ratios and anchor sizes. Later, normal maximum suppression (NMS) is applied to identify the target text area by calculating Intersection of Union (IoU). With huge anchors to find the desired box with text is like cumbersome and computationally challenging. To overcome this, by utilizing the region-based approach [24], given an input video, multiple regions are detected as regions of interest (ROIs) by a region proposal method with combination of graph-based segmentation approach. Initially, frames of a video are processed iteratively to find the regions with similar features like intensity, texture, and size. Then, each smaller chunk is combined to form a next layer region and repeated to obtain as many regions as possible till it cannot find further regions. Intensity similarity is robust to find the text with similar intensities because text has high intensity than any other object in the video. Hence, for each region, histogram is generated and similarity is computed by using Eq. (8).

$$T_{\text{intensity}}(r_i, r_j) = \sum_{n,k=1} \min\left(I_i^k, I_j^k\right), \tag{8}$$

$$S_{\text{size}}(r_i, r_j) = 1 - \left(\text{size}(r_i) + \text{size}(r_j)\right)/\text{size(frame)}. \tag{9}$$

Fig. 6 Region proposals

Once, we find the regions with similar intensity, then the corresponding neighboring text candidates are connected through intersecting similarity. This approach finds the text as a character or word with the level of intersection. It means if two corresponding characters of a word or two words of a sentence are merged by computing the intersection of corresponding characters or words.

$$T_{\text{overlap}}(r_i, r_j) = 1 - \big(\text{size}(BB_{ij}) - \text{size}(r_i) - \text{size}(r_j)\big)/\text{size(frame)}. \qquad (10)$$

The overall similarities are combined to find a final similarity to obtain the most prominent regions. It is shown in Eq. (11). Before this, Eqs. (9, 10) are used to find the size of the sub-region of whole frame and overlap score between the bounding regions and original ones.

$$\begin{aligned}
O_{(ri,rj)} = a1 * T_{\text{intensity}}(r_i, r_j) &+ a2 * S_{\text{size}}(r_i, r_j) \\
&+ a3 * T_{\text{overlap}}(r_i, r_j),
\end{aligned} \qquad (11)$$

where a_i is o for non-similarity and 1 for similarity.

The overall bounding box around the regions with the overlapping similarities acquired from Eq. (6) is considered along with the ground truth data obtained from the annotation details which is fed to the transfer learning classifier to isolate the false positives. The original Faster RCNN uses a feature map of convolution layer (conv5-3) layer in vgg16 model. The receptive field size in input image by the original region proposal network (RPN) is 228×228. Actually texts in video have various sizes and scales and single receptive filed would not be sufficient to detect the varied sizes of text. Because, if the text is too large and if the receptive filed small enough to detect the text, then it would represent noise. Therefore, the proposed procedure takes the advantage of VGG19 [21] with 47 layers and varied scales to improve the efficiency and robustness in detecting the text. We have defined adaptive RPNs for using different feature maps with different sizes [64^2, 148^2, 357^2, 512^2, 728^2]. The frame size of each video is considered to be uniform while extracting the key-frames from all the input videos. The uniform width and height that we have landed are 1280 and 720. Figure 6 shows the result of RPN scales and sizes that we have proposed, which determines that there are many texts which are not identified by the agreed parameters.

After obtaining the regions from RPN, the arrays generated from it are known as region of interest (RoI). These arrays are accumulated to into one array. To avoid the duplicates, we applied non-maximum suppression (NMS) and through finding the IoU with 0.7. These measures really affect the efficiency of our model when data augmentation pays a vital role as shown in Fig. 7. Text augmentation is a technique to fork original frame into number of copies with slightly modified pixels with respect to rotation, flips, mirror image, zooming, cropped, color modification, geometric transformations. This really helps to increase the ability of a model to learn the features to classify accurately.
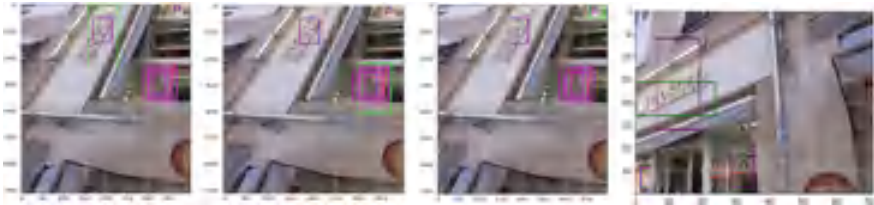
Fig. 7 Data augmentation with RPN

3.2 Text Augmentation

A survey on data augmentation [25] for a text in image is a common approach followed by most of the classifiers, therefore, for our dataset from video needed text augmentation to increase the frame data set for image classification task. While, after generating RoIs through convolution neural network (CNN), classified the regions obtained. But, we observed that the accuracy of a model was overfitting with the generated texts from RPN, and then, we employed data augmentation technique to acquire copies of frames with different vision angles as shown in Fig. 8. There are many techniques which are categorized into position augmentation, color augmentation, and Generative Adversarial Networks (GANs) [26]. The proposed method utilizes the simple geometric transformation techniques like rotation, flipping, scaling, and GANs.

3.2.1 Hybrid Text Augmentation

The proposed augmentation process considers the techniques of both geometric transformations and GANs to increase our dataset with mixed text patterns from frames. The novelty in our work takes the advantage of these to classify the frames with semi-supervised approach. GANs help in generating realistic frames in general. It includes Generator and discriminator. Generator is mainly used to generate the copies of original frames. Discriminator differentiates the real samples and generated samples
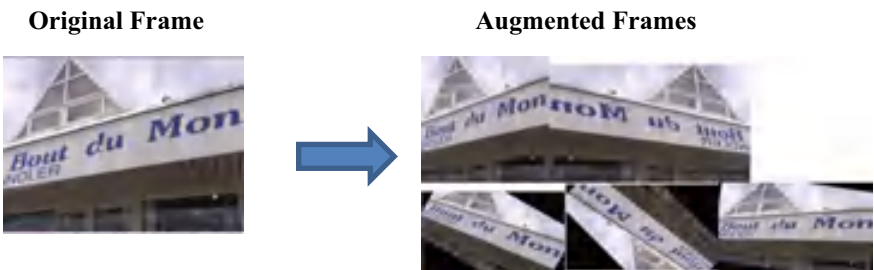
Fig. 8 Text augmentation

from generator. Discriminator assigns the scores based on their similarity index. During the training phase, generator uses the score of copies to improve the ability of creating more prominent frames. Discriminator also plays role in finding the classification loss. The main asset of GANs is that these two interact with each other to find the issues and improve the classifier well enough to generate relevant copies of frames/images. We have created the geometric transformed frames and distributed into train and test sets. The train set is further sent to GAN to generate more realistic transformed frames in Fig. 9.

Initially, the random samples of text train data are fed to the Generator for transmuting the input into instance and then forwarded to Discriminator to classify the generated frames. Then, the generator loss is calculated and also loss penalizes the generator for producing a frame that the discriminator can classify correctly. Later, discriminator tries to distinguish the original frames/images from the generated frames and computes the score to check how similar they are. Actually, the generated data with frames classified as fake are considered as negative samples during the training and the original data as positive samples. In order to reduce the loss, discriminator updates the weights through backpropagation. The equation of GANs for minimizing the reward is given as shown in Eq. (12).

$$\min_G \max_D V(D, G) = E_{x \sim \text{pdata}(x)}\big[\log D(x)\big]$$
$$+ E_{x \sim pz(z)}\big[\log(1 - D(G(z)))\big]. \qquad (12)$$

Thus, the name adversarial that denotes that generator and descriptor interacts each other in order to generate new and synthetic frame that can pass for original frames. Later, the probability distribution is imposed to calculate the distance loss between the distribution of generated and frame and the original frames. The minimax loss is formulated to determine the probability of corresponding frames data. Equation (13)
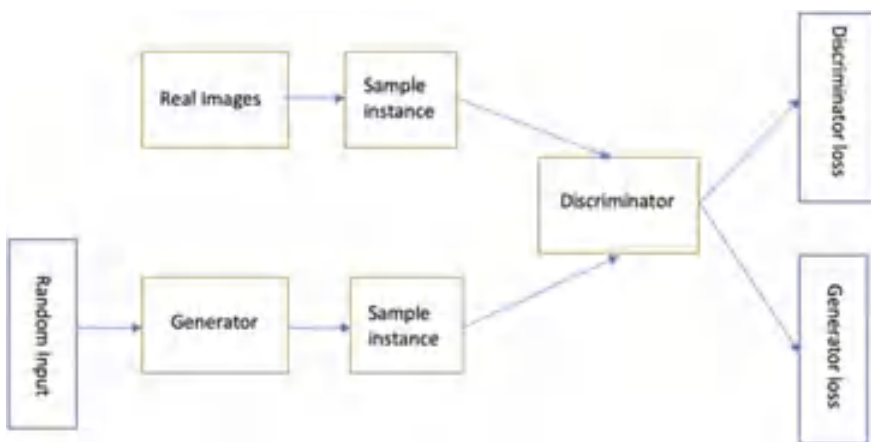
Fig. 9 GAN architecture

Fig. 10 Hybrid text augmentation using GANs

is minimized by the generator and minimized by the discriminator.

$$E_x\big[\log(D(x))\big] + E_z\big[\log(1 - D(G(z)))\big]. \tag{13}$$

Here, $D(x)$ is discriminator's estimate of the probability that real data instance x is real, E_x is the expected value of overall real data instances, $G(z)$ is the generator's output when noise z is given. This formula is derived from the cross-entropy between the real and generated distributions. E_z is the expected value overall random inputs to the generator. $D(G(z))$ is the discriminator's estimate of the probability that a fake instance is real (Fig. 10).

3.3 Fusion-Transferred Learning Architecture

With the help of feature pyramid [27], our model learned multiple scales to detect the text region with diverse scales and sizes. To acquire the desired text regions among the overall regions obtained by graph-based approach, we have fused the layers of features of different sizes.

The VGG19 [26] network model with all layers is similar to VGG16 model; however, we proposed VGG19 for options to detect the text with even bigger in size compared to VGG16. With this transfer learning approach by employing VGG19 based on multi-scale feature fusion [21] has increased the accuracy by extracting the prominent text features from convolution layers with different scales. So, out network defines varied RPNs with convolution layers of VGG19 with kernel size as 3×3, batch size as 256–512 as we increased the number of layers, stride with 2×2, and max pooling layers with 2×2 and 3×3 based on our text appearance in the video. The layers of VGG19 are given in [26] and the last layers are updated by above-proposed hyperparameters to detect the identify text robustly. Mainly, NMS with less than 0.6 is discarded to hold the less false detection regions. As demonstrated in Fig. 11, proposed pipeline combine feature maps after conv3 together by using 3×3 convolutions with stride 2. The generated two features maps are added up to predict region of interest (RoI) scores and bounding boxes. The later feature maps
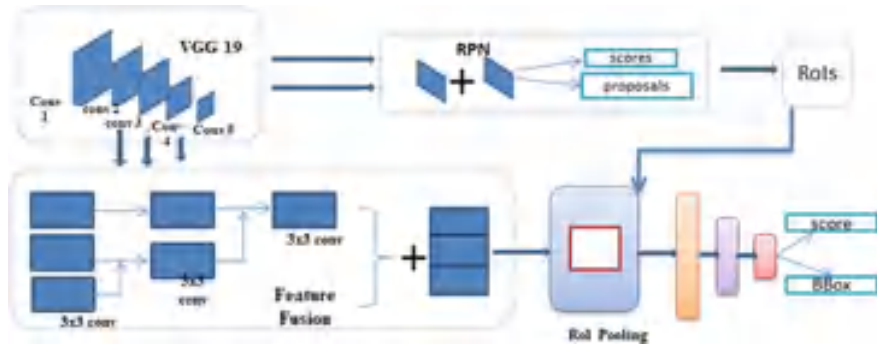
Fig. 11 Proposed architecture for text detection

are combined due to existence of semantic meaning of objects. Through fusion, the features of conv3, conv4, and conv5 are fused to obtain final high-resolution features containing string semantics.

4   Datasets and Evaluation Metrics

Datasets

We have taken ICDAR 2013 Text in Video [27, 31] Robust Reading Competition dataset. This includes train set with 25 videos and test set with 24 videos. Also, annotation file along with these sets gives bounding box information for all visible characters and words. There are also DON'T CARES with particular bounding box coordinates that represent text with low contrast, poor illumination, extremely blurred, long scaled. The text in this dataset is varied with respect to qualities like degradation, poor illumination, scaled in and scaled out, multi-linguistic words, shadowed words. Each video lasts from 10 s to 1 min with frame rate of 24–30 fps. The challenging in this dataset is motion configurations and nature noises, occlusion.

Experimental Results

We have mainly considered ICDAR 2013 dataset to process our algorithm and evaluate with other cited techniques. The hardware configuration includes NVidia GeForce with limited GPU memory and 16 GB RAM was good enough for small dataset or less number of epochs to run the classification model. However, we have utilized Google COLAB to run our model on GPU with 12–24 GB RAM. Our approach is compared with the Multi-RPN [21] and EAST [23]. We have tested on our dataset and compared to which type text like horizontal, vertical, rotated, blurred, these are performing. To overcome their downsides, we have proposed the novel approach and the performance is shown in Table 1.

Table 1 Comparison results of proposed and cited techniques

Method	Recall	Precision	F-measure
Proposed	**0.92**	**0.88**	**0.90**
Multi RPN [21]	0.81	0.90	0.85
EAST [23]	0.87	0.67	0.76
SegLink [19]	0.88	0.87	0.88
Corner features [22]	0.88	0.83	0.85
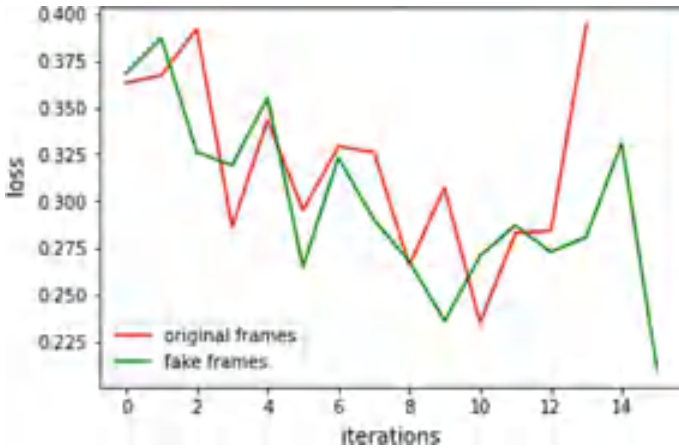
Fig. 12 Generator–-discriminator loss for one sample

With respect to hybrid approach of text augmentation, the accuracy of a model is determined by how efficiently the augmented frames are generated through the approach. This can be best understood by considering the tab 2. This shows the loss function between the discriminator and generator samples. In case, if the discriminator loss moves to zero and stays for long time, then it indicates the failed trained model. The loss function in Fig. 12 on our dataset with one sample states that model is able to generate the copies of fake frame samples from original frames as long as the discriminator is not remained on zero.

5 Conclusion

The proposed method of varied RPN with fusion transfer learning model though Faster RCNN for text detection is well employed to increase the accuracy through performance and recall. The approach was efficient than the traditional process of Faster RCNN. Also, with the hybrid text augmentation technique, the proposed pipeline took next level compared to state-of-the-art methods in classifying the

text and detects text reliably. With this, our method improved detection exactitude radically. With the influence of text augmentation mainly using GANs, text with horizontal, vertical, and multi-oriented in videos improves the detection algorithm commendably.

References

1. Chen X, Jin L, Zhu Y, Luo C, Wang T (2020) Text recognition in the wild: a survey. J ACM: 1–34
2. Cao D, Zhong Y, Wang L, He Y, Dang J (2020) Scene text detection in natural images: a review. Symmetry 12(12):1–27
3. Qin H, Zhang H, Wang H (2019) An algorithm for scene text detection using multibox and semantic segmentation. MDPI Appl Sci 9
4. Hu H, Zhang C, Luo Y, Wang Y, Han J et al (2017) WordSup: exploiting word annotations for character based text detection. In: IEEE international conference on computer vision (ICCV)
5. Devi MM, Seetha M, Raju SV, Rao DS (2019) Detection and tracking of text from video using MSER and SHIFT, vol 4. In: Advances in decision sciences; image processing, security and computer vision. Springer, pp 719–727
6. Jung K, Kim KI, Jain AK (2004) Text information extraction in images and video: a survey. Pattern Recogn 37(5):977–997
7. Mortha MD, Seetha M, Viswanadha Raju S (2019) Text spotting in video: recent progress and future trends. Int J Recent Technol Eng (IJRTE) 7:116–124
8. Wang T, Wu DJ, Coates A, Ng AY (2012) End-to-end text recognition with convolution neural networks. In: Proceedings of international conference on pattern recognition, pp 3304–2208
9. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations
10. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of IEEE conference on computing vision pattern recognition, pp 770–778
11. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of IEEE conference on computing vision pattern recognition, pp 2818–2826
12. Borisyuk F, Gordo A, Sivakumar V (2018) Rosetta: large scale system for text detection and recognition in images. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. ACM, pp 71–79
13. Epshtein B, Ofek E, Wexler Y (2010) Detecting text in natural scenes with stroke width transform. IEEE computer society conference on computer vision and pattern recognition, pp 2963–2970
14. Huang W, Lin Z, Yang J, Wang J (2013) Text localization in natural images using stroke feature transform and text covariance descriptors. In: IEEE international conference on computer vision, pp 1241–1248
15. Tian S, Pan Y, Huang C, Lu S, Yu K et al (2015) Text flow: a unified text detection system in natural scene images. In: IEEE international conference on computer vision (ICCV), pp 4651–4659
16. Shi B, Bai X, Belongie S (2017) Detecting oriented text in natural images by linking segments. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 3482–3490
17. Tian Z, Huang W, He T, He P, Qiao Y (2016) Detecting text in natural image with connectionist text proposal network. In: Computer vision—ECCV 2016, pp 56–72
18. Zhang Z, Shen W, Yao C, Bai X (2015) Symmetry-based text line detection in natural scenes. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), 2558–2567

19. Wang L, Wang Y, Shan S, Su F (2018) Scene text detection and tracking in video with background cues. In: International conference on multimedia retrieval (ICMR). ACM, pp 160–168
20. Liu X, Liang D, Yan S, Chen D, Qiao Y, Yan J (2018)FOTS: fast oriented text spotting with a unified network. In: IEEE/CVF conference on computer vision and pattern recognition, pp 5656–5685
21. Zhang MI, Pang K, Chen Y, Zhang B (2020) Character recognition based on multi-scale feature fusion transfer learning. In: Signal and information processing, networking and computers. Springer, pp 781–788
22. Lu W, Sun H, Chu J, Huang X, Yu J (2018) A novel approach for video text detection and recognition based on a corner response feature map and transferred deep convolutional neural network, vol 6. IEEE, pp 40198–40211
23. Zhou X, Yao C, Wen H, Wang Y, Zhou S (2017) EAST: an efficient and accurate scene text detector. In: IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, USA, pp 2642–2651
24. Zhong Z, Sun L, Huo Q (2019) An anchor-free region proposal network for faster RCNN-based text detection approaches. Int J Doc Anal Recogn (IJDAR) 22:315–327
25. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(1)
26. Agnieszka M, Michal G (2018) Data augmentation for improving deep learning in image classification problem. In: IEEE 2018 international interdisciplinary Ph.D. workshop (IIPHDW), pp 117–122
27. Zhu Z, Liao M, Shi B, Bai X (2018) Feature fusion for scene text detection. In: International workshop on document analysis systems (DAS), pp 193–198
28. Liao M, Shi B, Bai X (2018) TextBoxes++: a single-shot oriented scene text detector. IEEE Trans Image Process 27(8):3676–3690
29. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR)
30. Nguyen PX, Wang K, Belongie S (2014) Video text detection and recognition: dataset and benchmark. IEEE winter conference on applications of computer vision, pp 776–783
31. Tarakeswara Rao B, Ramakrishna Murty M (2020) A comparative study on effective approaches for unsupervised statistical machine translation. In: International conference and published the proceedings in AISC Springer conference, vol 1076, pp 895–905

Quality-Produced Agricultural Crop Price Prediction Using Machine Learning

Tumma Susmitha, Talla Prashanthi, and Rupesh Kumar Mishra

Abstract Agriculture is India's backbone. It is a key sector of the Indian economy, contributing roughly 17% of the country's overall GDP and employing over 60% of the population. We can use technology to improve product production in a variety of ways, but in the end, a farmer can only benefit if he makes money selling his crops. The Indian government has passed three legislations to promote agricultural produce trade throughout the country. Today, however, we can see farmers across the country battling for their rights against these rules. Farmers fear that they will be used as puppets by major retailers and that their products would be sold at a reduced price. After analysing the situation, we came up with the idea of developing an agricultural produce application that predicts the price of agricultural produce based on the quantity produced and previous years' sales rates, allows farmers to interact directly with retailers, and allows for product review and crop yielding rate prediction.

Keywords KNN · GPS navigation · Decision tree · Regression algorithm

1 Introduction

A robust market is required for agriculture produced in our country. Farmers have a difficult time for selling their products to clients. Farmers in India have little choice in markets to sell their products. Except for three states, all states require that farm produce can be sold and marketed through state-owned mandis, retail

T. Susmitha (✉) · T. Prashanthi · R. K. Mishra
Computer Science and Engineering, Chaitanya Bharathi Institute of Technology (A), Hyderabad, Telangana, India
e-mail: susmithat_cse@cbit.ac.in

T. Prashanthi
e-mail: prashanthit_cse@cbit.ac.in

R. K. Mishra
e-mail: rupeshmishra_cse@cbit.ac.in

markets where intermediaries pressure farmers to increase margins. Crop Cost Prediction, Language Translator, sorting based on geographical proximity for the farmer/customer, customising the app for that particular farmer's crop and profit, etc., using the Machine Learning, Deep Learning Algorithms—Decision Tree Regression Algorithm for Price Prediction and other techniques such as—GPS Navigation, KNN, Haversine, nearest neighbour search, load balancing, Market analysis, and few APIs for geographical proximity would make this application an excellent crop selling tool for farmers with good profits completely eliminating the mediators and the middlemen. An E-commerce application for the farmers with all the requirements by which a farmer can sell the products get to know about products' profits and cost and can interact with the customers directly by being in the safety of home. By this application, the farmers can interact with customers in their native language and also use the application without having much knowledge about technologies and mobiles, etc.

2   Related Work

In [1], the system is offering a platform at the government level, such as an Android app and a website app, where farmers may sell their crop products at various layers of the marketing chain (market, merchant, or end-user) with several alternatives. Farmers may use the platform to find out about nearby marketplaces, current stock levels, and demand for specific products in less time and with less effort. The web-based program will include information such as market details, merchant lists, farmer lists, end-user lists, and complaints lists, among other things. This will result in improved government management.

In [2], the focus of this research is on machine learning techniques for forecasting crop price using the support vector regression algorithm. Regression is a data mining methodology in which learning is used to calculate the crop price. Classification tasks and tasks with specific class labels will be considered for regression tasks. The crop price is computed by recognising our training dataset's patterns, which is provided as one of the algorithm's inputs. The user enters the parameter input values (Yield, Rainfall, Minimum Support Price, and Wholesale Price Index) into the algorithm. Probability, New Record Input, and the number of Dataset Parameters are the other parameters in the algorithm.

In [3], they propose that farmers can be protected from diminishing returns by increasing rice yield and through crop rotation. In order to maximise sales profit under the assumption of irrigation, this study used predictive data analysis to anticipate the selling prices of agricultural output suited for planting in addition to rice. The pricing data for six different alternative crop types grown in settings of limited resources—turnips, muskmelons, kailan, peanuts, cantaloupes, and water mimosas—were used in this study's monthly time series analysis of Hom Pathum rice. Least Square Method, Moving Average Method (3 months, 5 months, 7 months), Single Exponential Method, Double Exponential Method, and Winters' Method were used to analyse

sixty-month sale data of Hom Pathum Rice and six alternative crops. To evaluate the effectiveness of predictions, the Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Square Deviation were utilised (MSD). Then, for 12 case studies, the appropriate prediction approach was applied to predict the selling prices of rotation crops and rice for three different situations. They discovered Case Study 6 produced the highest level of profit (per cent increase when compared with a traditional method). In [4], understanding the structure of both languages is required to convert an English text into Telugu using a rule-based translation system. The structure and grammar of both languages influence the translation process. One of the most difficult aspects is dealing with prepositions. In Telugu, prepositions will be interpreted as postpositions. The purpose of this paper is to choose the proper postposition, create the prepositional phrase, and translate the supplied text from English to Telugu.

In [5], a natural foundation for modelling and analysing client interactions with E-Commerce search engines is customer interaction networks (CINs). The submission of a query based on an initial product intent kicks off customer interactions and then progresses through a series of activities for product engagement and inquiry reformulation. The amount of time a consumer spends with a product (e.g. clicks) reflects how relevant it is to their needs. Dissatisfaction with current outcomes or a change in the customer's product purpose signal reformulation to a new inquiry. We can uncover multiple query–query and query–product correlations by analysing such interactions inside and between sessions. The features of CINs developed using Walmart.com's product search records are the focus of this work. They show in their article that the features of CINs allow them to my purpose links between inquiries only depending on the structural data they contain. The paper demonstrates how these relationships can be used to (a) cluster inquiries based on purposes, (b) improve the quality of searches for underperforming queries, and (c) find the most influential (also known as "important") searches whose results have the biggest influence on the results of subsequent queries.

Predicting product prices is critical for selling items. The paper [6] uses two studies to forecast the product price: qualitative techniques and quantitative techniques. Qualitative cost assessment strategies are generally based on comparing a new product to prior goods in order to find similarities in the new one. The commonalities found aid in incorporating previous data into the new product, reducing the requirement to generate a cost estimate from start.

3   Methodology

3.1   Data Collection

The practise of acquiring and analysing data from a variety of sources is known as data collection. In order to leverage data to create workable artificial intelligence (AI) and machine learning solutions, the data must be gathered and maintained in a format that makes sense for the business challenge at hand. The real-time data are collected from the Government Websites and Offices of Food and Agriculture development.

3.2   Data Augmentation

Data augmentation is a group of methods for generating extra data points from already-existing data to fictitiously increase the volume of data. Examples of this include using deep learning models to generate more data points or making little modifications to the data augmentation can enhance the performance and outcomes of machine learning models by adding new and diverse examples to training datasets. When the dataset is large and sufficient, a machine learning model performs better and more accurately. For machine learning models, data gathering and labelling can be time-consuming and expensive. By changing datasets using data augmentation techniques, businesses can reduce these operational costs.

3.3   Training

The proposed work has the data of different features. The data are trained using different algorithms to understand the working of these algorithms. Every Feature as different training sets and algorithm.

3.4   Visualisation

We can observe how the data appear and what sort of correlation the properties of data hold with the help of data visualisation. It is the quickest approach to check if the features match the output. Confusion matrix is used to visualise the accuracy. This work interprets the model by generating heat map to visualise the data.

Fig. 1 Block diagram of proposed system
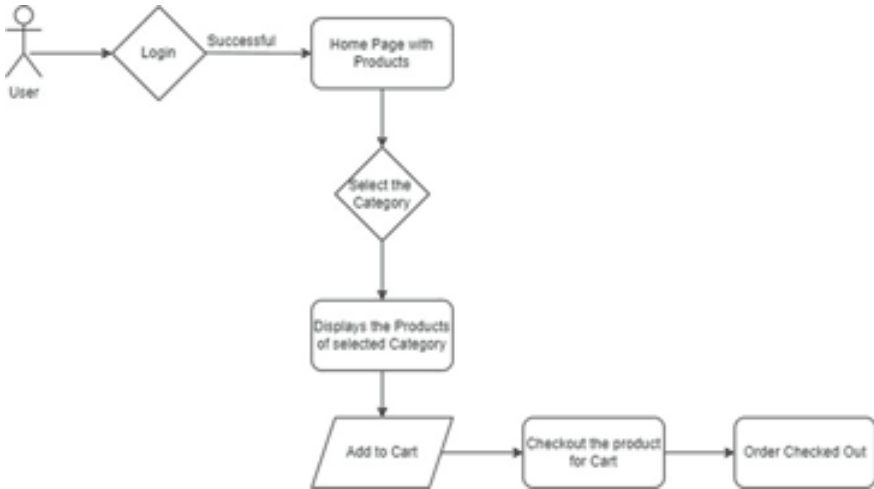
Fig. 2 Diagram of the proposed system

3.5 Testing and Accuracy

The models are tested and their respective accuracies are calculated. The models are also tested with the real-time data collected and thus validated the performance of the models (Figs. 1 and 2).

Crop Prediction

Crop prediction is very useful for farmers when they are deciding the crop for the season. So, crop is predicted using the values that are suitable for the conditions in the area. For these, Random Forest Classifier is used since it gives the best accuracy.

Random Forest Classifier

A well-known machine learning algorithm called Random Forest makes use of supervised learning strategies. It can be used in machine learning for both classification and regression problems. It is based on ensemble learning, a technique for combining several classifiers to tackle a challenging problem and improve the performance of the model. "Random Forest is a classifier that contains a number of decision trees on various subsets of a given dataset and takes the average to enhance the predicted accuracy of that dataset," according to the name. The Random Forest gathers the predictions from each decision tree and predicts the ultimate result based on the
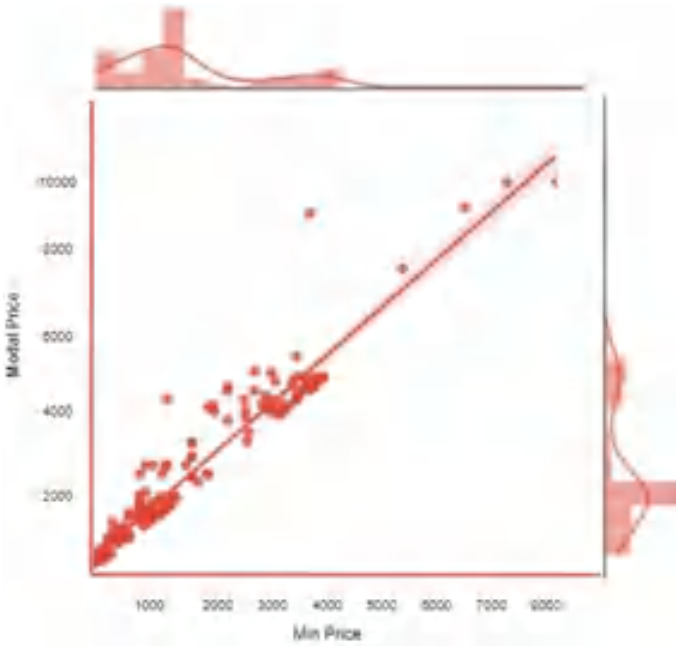
Fig. 3 Min versus modal comparison

majority votes of predictions, as opposed to depending just on one decision tree (Figs. 3, 4 and 5).

Crop Price Prediction

Crop Price Prediction is useful to get the prices of the products in the application dynamically. It can be done by training a model to get the accurate price for the crop according to the market prices.

Prices can be calculated using min–Max Models. Here, the algorithms are considered and compared to the best suitable algorithm for the price prediction.

Linear Regression

A supervised learning machine learning algorithm is linear regression. It performs a regression operation. Regression creates a value for the goal prediction based on independent factors. It is generally used for forecasting and figuring out how variables are related to one another. The kind of relationship that is assessed between the dependent and independent variables, as well as the number of independent variables used, varies between different regression models.
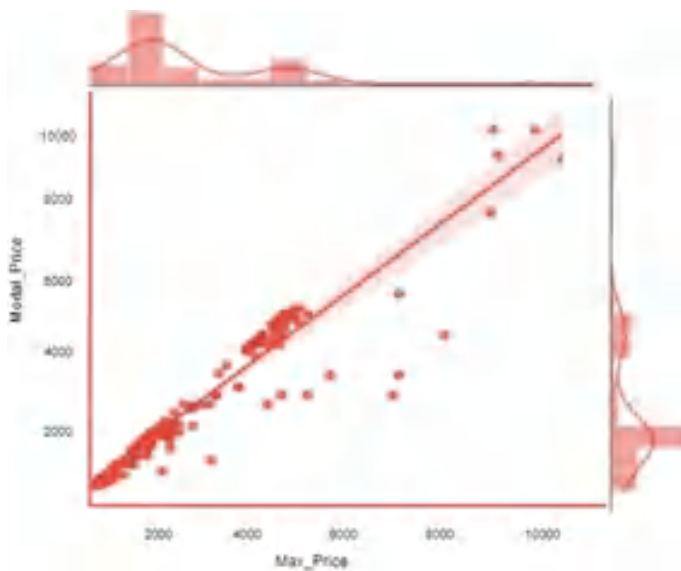
Fig. 4 Max versus modal comparison
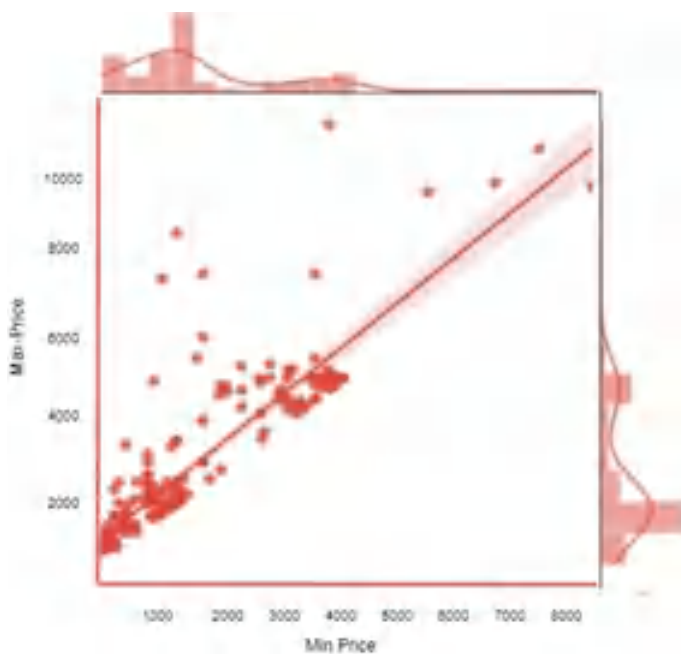
Fig. 5 Min. versus max. comparison

Decision Tree

The decision tree is the most effective and popular categorization and prediction tool. Each leaf node (terminal node) carries a class label, whereas each internal node symbolises a test on an attribute, each branch a test result.

4   Dataset Description

The dataset is collected from the Government Website of Telangana with valid fields, and it is trained well also using data obtained from online.

5   Conclusion and Future Work

Hereby, we conclude that our application helps farmers to grow economically and provides them with personalised data with statistics of their products. The application reduces the manual work, makes it easy for the farmer to sell his products, and gives him the insights of the market trends. The application uses the machine learning algorithms to predict crop and also price which helps farmers in many ways.

We hope to convert the application web to mobile in future. It can also modify in easy way for all the users. Banking and other processes availability for the transactions. Transportation facilities for the Product from the Application End.

References

1. Bhende M, Avatade MS, Patil S, Mishra P, Prasad P, Shewalkar S (2018) Digital market: E-commerce application for farmers. In: 2018 fourth international conference on computing communication control and automation (ICCUBEA)
2. Rohith R, Vishnu R, Kishore A, Chakkarawarthi D (2020) Crop price prediction and forecasting system using supervised machine learning algorithms. Int J Adv Res Comput Commun Eng (IJARCCE)
3. Ruekkasaem L, Sasananan M (2018) Forecasting agricultural products prices using time series methods for crop planning. Int J Mech Eng Technol (IJMET)
4. Lingam K, Rama Lakshmi E, Ravi Theja L. Rule based machine translation from English to Telugu with emphasis on prepositions. In: First international conference on networks & soft computing. IEEE
5. Adhikari B, Sondhi P, Zhang W, Sharma M, Aditya Prakash B. Mining E-commerce query relations using customer interaction networks. Department of Computer Science, Virginia Tech, WalmartLabs
6. Dai JS, Balabani S, Seneviratne L. Product cost estimation: technique classification and methodology review. J Manuf Sci Eng

An Optimized Control on Delay and Transmission Rate Over Wireless Video Streaming Channels

K. Maheswari and N. Padmaja

Abstract The quality of video data transmitted in real-time over internet is often affected with an end-to-end delay. It further burdens the video data transmission in addition with data rate distortion. Hence, it is necessary to optimize the delay and rate behavior using a diagnostic optimistic mechanism. In this paper, we develop a machine learning mechanism that considers the delay and data as an important factor in delivering the video to the destined user without delay or buffering. The machine learning using artificial neural network (ANN)-based Markov Chain maintains the trade-off between the delay and data rate and considers both the factors as a severe constraint in affecting the performance of data delivery over wireless channels. The machine learning offers scheduling of data packets that tends to reduce the data minimization, increases the transmission rate, and reduces the delay. The scheduling sets the priority of each data and avoids retransmission of data. The experimental verification shows the efficacy of the model over existing state-of-the-art mechanism in terms of transmission errors and delay.

Keywords Video data · Streaming · Transmission rate · Delay

1 Introduction

Heterogeneous environments have played an important part in the digital streaming systems of LTE networks in recent decades. The LTE multimedia service network provides customers with a streamlined solution in terms of unified coverage and better resource sharing with multiple users [1].

K. Maheswari (✉)
Research Scholar, Department of ECE, Jawaharlal Nehru Technological University Anantapur, Anantapur, Andhra Pradesh 515002, India
e-mail: kmaheswari2009@gmail.com

N. Padmaja
Professor, Department of ECE, School of Engineering, Mohan Babu University (erstwhile Sree Vidyanikethan Engineering College), Tirupati, Andhra Pradesh 517102, India

The LTE networks enable broad-range transmission, but long-range communications remain a bottleneck that provides consumers with big data services. As the demand for more data services in LTE networks grows, Quality of Service (QoS) suffers significantly, resulting in customers receiving poor Quality of Service from network operators. Such high-demand communication services impact the LTE network's voice services because of their reduced bandwidth [2, 3].

Some networks, such as vehicle highways, require a low level of jitter and maximum reliability. In addition, due to the bandwidth usage on the network, high performance is required [4]. Similarly, in the case of the higher data rate on LTE networks, a high-quality requirement is needed [5]. However, with its high-quality users, video streaming services need a variety of data, particularly on networks such as LTE, with many users as the key concern, and QoS needs to be addressed. Owing to the efficient distribution of network resources to its customers, latency is often a key concern in the network. Again, this is the second big restriction [5].

The following are several reports that help existing research to mitigate the above-mentioned impacts by increasing video streaming on LTE networks [6].

To increase the time-to-watch ratio, ANN-MC management has been improved, and video stream quality on LTE networks has been improved [7]. The Quality of Experience (QoE) is computed for each user and re-buffered for each user, in addition to the QoS parameter, for the overall buffering time. This is called a joint problem of optimization by reducing the problem needed to a single problem of optimization [8]. In conclusion, the multi-user dynamic video stream consists of sub-optimal solutions that have an ideal matching algorithm [9]. Many methods are used to mitigate the interference effects of LTE video streaming networks, including cache-induced joint encoding, which improve QoS and reduce efficiency measurement loss in LTE networks.

In addition, thanks to video streaming, the overall energy demand of these networks is considerably reduced [10]. The improved QoS further reduces latency by reducing bandwidth allocation to background voice users [11]. The user who takes semi-Markov judgment, which achieves a smooth video replay and helps to keep the background user without any loss in output, is involved in other related techniques. It appears that analysis emphasis on QoS metrics is boosted by an expanded resource assignment methodology in LTE network video streaming. This research therefore further aims to increase the overall QoS levels to improve network stability and production [12–15]. The research is focused on technical formulations that use a knowledgeable network adaptive streaming protocol to improve the efficiency of the video provided to the user in wireless LTE. The present analysis also includes the presence, delay metrics, and propagation metrics in order to transfer the video to the viewer with the correct bandwidth and content evaluation. This research further examines the use of a data quality evaluation for an end-to-end distortion model that takes note of the encoding speed required to increase LTE video streaming network video bit rate.

The present study covers the formulation needed for an improved bit rate encoding and for high video quality for target users. The application of video streaming servers has been significant in current research, but the inefficacy of the dependent streaming

service in relation to the content-centric network is demonstrated. The suggested approach is also using the prioritization method to boost video delivery capability, which enhances video playback in streaming servers and further decreases the overall network time with improved throughput.

2 Background

QoS is a significant parameter in the multi-user environment for video streaming services through the LTE network. The video streaming network services are also a subject to latency and communication errors that occur during buffering during the video streaming phase. The buffering takes place because the bandwidth is restricted and the video quality is selected according to consumer requirements [16, 17]. Thus, this chapter takes account of latency and takes into account the steps to reduce this latency. The QoS is increased in this way, even though it appears inappropriate to allocate the bandwidth over LTE networks during video streaming services [18].

The latency issue is avoided with the encoder, which sends the packet to the streaming LTE network with a shortened delay [19]. This contributes to reducing the delay in LTE bandwidth allocation in video streaming networks, which rely entirely on the multiple users' service requirement. In addition, this analysis aims to increase LTE network QoS by reducing delays by improving network distribution of resources and realistic aggregate performance [20–23].

The research considers the use of a uniform solution, which uses the ANN controller to overcome problems due to delay and ANN-MC. The video output is well adjusted with an ANN-MC. The improvements to the ANN controller are carried out to optimize the QoS for the network, which utilizes the previously assigned bandwidth for each user and the potential bandwidth to be assigned. Through this feedback, judgment feedback is changed and the bandwidth is automatically allocated with an optimum bandwidth range.

Finally, with the Markov Chain algorithm, which is paired with the ANN controller to pick the appropriate bit rate, the judgment is further enhanced. This allows the multi-cast setting to be assigned to the resource with improved bit rate selection and encoder selection further increases bit rate selection. The selection of the bit rate is achieved successfully on the basis of the users' service requirements and includes the infrastructure in order to minimize network congestion with enhanced QoS for massive data transmission across the LTE video streaming network.

3 Proposed Method

The video streaming framework is a video streaming protocol for IP network distribution and, unlike UDP streaming, represents a client-driven approach. The request for the retrieval of video information from the client end is submitted, and the device

output is balanced by the normal HTTP streaming protocol. The video output is separated by the server into video fragments, which are separately decoded on the user or customer interface. Client and HTTP server connectivity is developed using the media presentation definition as shown in Fig. 1. The MPD consists of individual or several time intervals in which many examples of one video material are shown per time. The example is largely consistent with a different coding process, which has a varied bit rate. An initial section with metadata contents has an entity representation or an example of the video sector. For the consumer, this is really helpful for knowing a certain video or section metadata. Customers are able to adapt their average transmitting capability to the varying streaming rate of the video segments during video streaming.

The proposed study uses the LTE downlink method, with n videos sent over eNode B, while ANN-MC shares the N bandwidth between devices. The bandwidth here consists of 12 subcarriers, located next to each other. The 12 subcontractors belong to the n bandwidth m subcarrier. Finally, the feasible data rate for the one-user is seen in the following terms:

$$R(l, m(n)) = B\log_2(1 + p(l, m(n)h(l, m(n))), \tag{1}$$

where

$p(l, m(n))$ required allocated power to user from subcarrier.
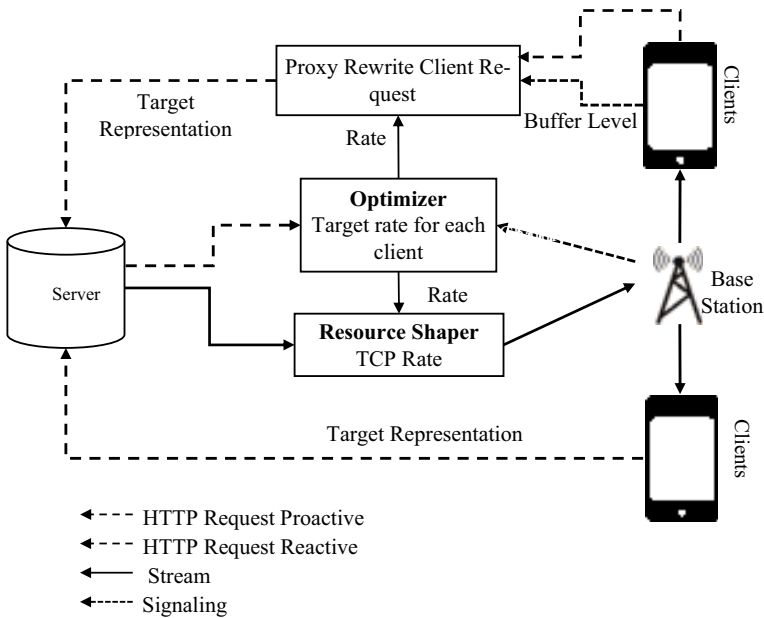
Fig. 1 Proposed approach on video streaming

$h(l, m(n))$ channel condition.

I interference.

i instant channel gain.

N_0 noise power spectrum density is given as below:

$$h(l, m(n)) = i/(N_0 B + I). \tag{2}$$

Therefore, the data rate of individual user l is represented as:

$$R(k) = \sum nc(k, n) \cdot \sum m R(k, m(n)), \tag{3}$$

$$R(k) = \sum nc(k, n) \cdot \sum m \, B\log_2(1 + p(k, m(n)H(k, m(n))), \tag{4}$$

where c is the bandwidth allocated nth for the lth user in the network.

The video streaming QoS formulation is considered as:

$$Q = x \cdot \ln R(k) + y. \tag{5}$$

Thus, any user achieves a maximum level of QoS, which is taken into consideration with the corresponding data rate $R(k)$. The LTE network uses a programmed service class that specially indicates network users' achievement of video streaming output, which is used to achieve the nominal fraction value. This is the multi-objective function that is expressed in the form:

$$\max \sum_{a=1}^{n} \sum_{r \in RB_a} r_{a,r}, \tag{6}$$

$$\forall a \in N: \lim_{t \to \infty} \inf r_a(t) \geq \gamma_a \bar{r}, \tag{7}$$

$$\forall a \in N: \min(l_a) \text{ and } \min(d_a) \tag{8}$$

subjected to

$$\forall a, b \in N, a \neq b: RB_a \cap RB_b = \theta, \tag{9}$$

$$\forall a \in N, k \in K: d_{a_k} < D_k, \tag{10}$$

$$\forall a \in N, k \in K: l_{a_k} < L_k, \tag{11}$$

where

a   assigned user
N   total users where $N = 1, 2, \ldots, n$.

The resource block allocated based on each user classes is defined as:

$$RBa \subset RB = \{1, 2, \ldots, r\}, \tag{12}$$

where

RBa   allocated resource block to dedicated users, *a*,
ra, r   achieved throughput.
γa overall throughput.
Dk   maximum packet delay.
Lk   maximum packet loss.
θ fairness throughput.
dak   packet delay.
Iak   packet loss factor.

4   Results and Discussions

On the basis of quality measurement that includes throughput and latency, the proposed study is considered efficient. By better selecting its associated parameters, the controller with a correct interval is considered simple and efficient.

The weighted average is used to select these three tuning parameters, proportional, integral, and derivative parameters. The controller's output is closer to the value of one and this helps to reflect the use of three tuning parameters to achieve the user service data rate. In comparison with three conventional methods, the proposed method for improvement of the video service is the Markov chain and exhaustive research.

The result is estimated based on the above-mentioned parameters, thus providing the path to improved use of network resources even if the bandwidth is limited. The current performance, past ANN-MC, and current ANN-MC are the key parameters for each single user to achieve their main goal.

Figure 2 shows that the ANN-MC controller does the correct execution of the bit rate needed in multimedia video streams to improve compared to other methods. Here, the proposed MC algorithm controller, as discussed above, is compared to other methods.

The evaluation criteria are associated with parameters for the entire LTE network user and the reachable aggregate performance. The results have shown that the proposed method has greater throughput and a more comprehensive search than the Markov chain. The ANN-MC mechanism results are initially generated without using an ANN-MC controller and test results are obtained. The ANN-MC controller will also be tested with the results of the proposed method. The results show that
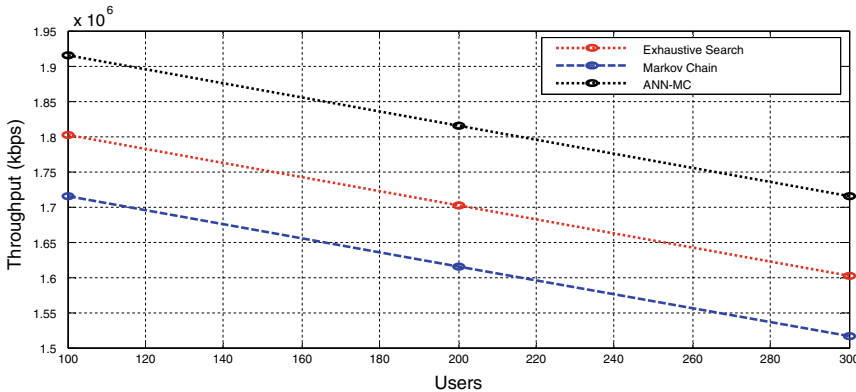
(a) mobile users
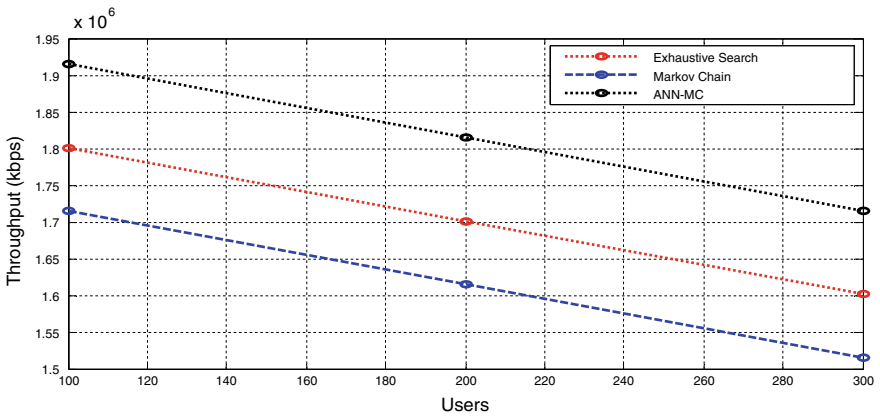
(b) static users

Fig. 2 Aggregate throughput

ANN-maximum MC's aggressive rate is high compared to the other mechanisms. The higher level of performance is because the users in the LTE network allocation are higher than the other methods. The proposed mechanism therefore achieves the criteria for meeting users with maximum QoS.

Similarly, Fig. 3 results show that the ANN-MC mechanism, based on the PID controller, has a greater interval than the other methods. The latency is defined as the reciprocal of the LTE system delay parameter. This shows that with an increased interval ratio, the delay value was significantly reduced. The technique shows a better interval ratio without the use of the ANN-MC algorithm than other methods. The proposed method with the use of the controller could further be determined by means of the feedback findings derived from past and present values to offer a better interval ratio.

(a) mobile users

(b) static users
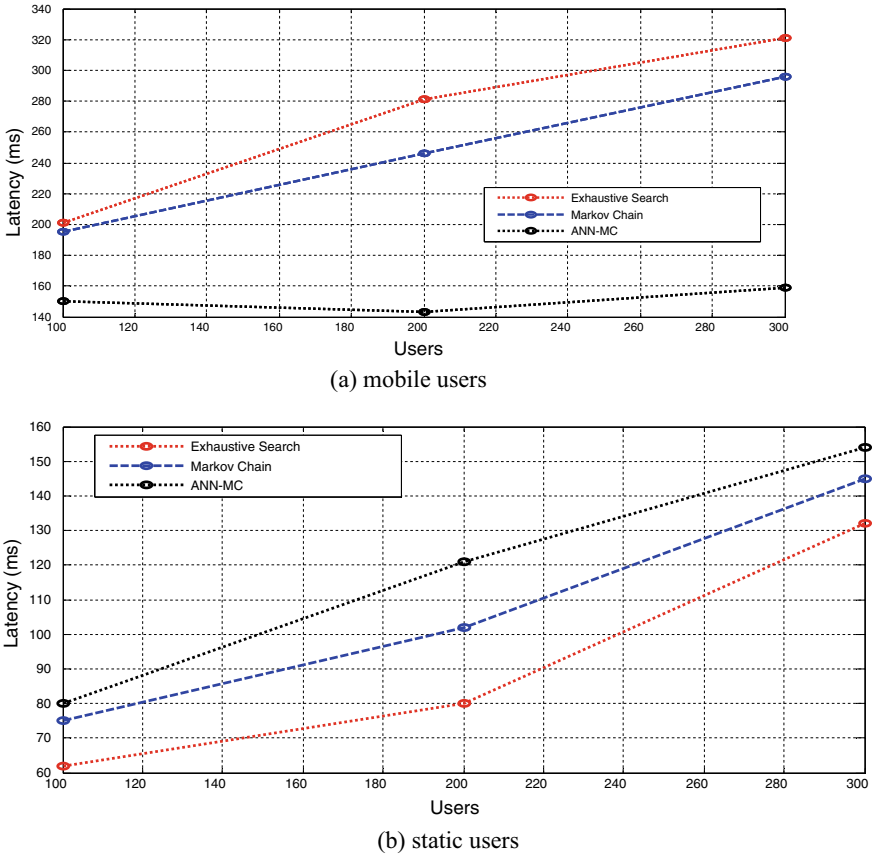
Fig. 3 Latency

Figure 4 shows the PSNR of video streaming system for varying bit rates between 20 and 200 kbps. The rate of average PSNR shows that the coding efficiency is high in proposed ANN-MC model than Markov chain and exhaustive search mechanism. It is further seen that the average PSNR is 1.15% higher than Markov chain and 1.25% higher than exhaustive search model. The ANN-MC algorithm proposed method has better interval than the ANN-MC algorithm non-ANN method. The ANN-MC algorithm aids in the improvement of interval dynamics and improves with a ratio of 0.8. The interval ratio is more important than other conventional methods, which demonstrate the efficiency of the ANN-MC algorithm linked to server-based LTE networks.

Finally, the reduced latentness with an improved overall performance could be seen to increase the QoS user. In addition, the LTE network user experience is improved with the LTE mechanism, and a high interval ratio allows the video to be offered to the user without buffering.
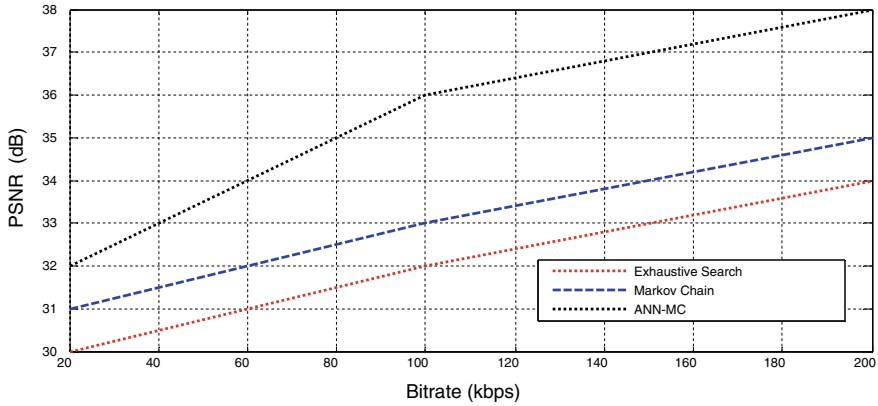
Fig. 4 PSNR of video streaming system

5 Conclusions

In this paper, we developed ANN-based Markov chain to maintain the trade-off between the delay and data rate in wireless video streaming channels. The ANN-based Markov chain offers scheduling of data packets that tends to reduce the data minimization, increases the transmission rate, and reduces the delay. The scheduling sets the priority of each data and avoids retransmission of data. The experimental verification shows the efficacy of the model over existing state-of-the-art mechanism in terms of transmission errors and delay. Hence, providing QOS guarantees for application services over next-generation wireless network a common QOS frame work will be proposed for future research.

References

1. Ozmen M, Gursoy MC (2017) Secure transmission of delay-sensitive data over wireless fading channels. IEEE Trans Inf Forensics Secur 12(9):2036–2051
2. Wang CC, Lin ZN, Yang SR, Lin P (2017, June) Mobile edge computing-enabled channel-aware video streaming for 4G LTE. In: 2017 13th international wireless communications and mobile computing conference (IWCMC). IEEE, pp 564–569
3. Huang T, Zhang RX, Zhou C, Sun L (2018, Oct) Qarc: video quality aware rate control for real-time video streaming based on deep reinforcement learning. In: Proceedings of the 26th ACM international conference on multimedia, pp 1208–1216
4. Ahmed AA, Ali W (2018) A lightweight reliability mechanism proposed for datagram congestion control protocol over wireless multimedia sensor networks. Trans Emerg Telecommun Technol 29(3):e3296
5. Chen J, Wei Z, Li S, Cao B (2020) Artificial intelligence aided joint bit rate selection and radio resource allocation for adaptive video streaming over F-RANs. IEEE Wirel Commun 27(2):36–43

6. Du J, Yu FR, Lu G, Wang J, Jiang J, Chu X (2020) MEC-assisted immersive VR video streaming over terahertz wireless networks: a deep reinforcement learning approach. IEEE Internet Things J 7(10):9517–9529
7. Han S, Noh Y, Lee U, Gerla M (2019) Optical-acoustic hybrid network toward real-time video streaming for mobile underwater sensors. Ad Hoc Netw 83:1–7
8. Hamdoun H, Nazir S, Alzubi JA, Laskot P, Alzubi OA (2021) Performance benefits of network coding for HEVC video communications in satellite networks. Iran J Electr Electron Eng 17(3):1956–1956
9. Feng H, Llorca J, Tulino AM, Molisch AF (2018) Optimal control of wireless computing networks. IEEE Trans Wireless Commun 17(12):8283–8298
10. Sun L, Duanmu F, Liu Y, Wang Y, Ye Y, Shi H, Dai D (2018, June) Multi-path multi-tier 360-degree video streaming in 5G networks. In: Proceedings of the 9th ACM multimedia systems conference, pp 162–173
11. Li C, Xiong H, Zou J, Wu DO (2017) Joint dynamic rate control and transmission scheduling for scalable video multirate multicast over wireless networks. IEEE Trans Multimedia 20(2):361–378
12. Guo Y, Yang Q, Yu FR, Leung VC (2018) Cache-enabled adaptive video streaming over vehicular networks: a dynamic approach. IEEE Trans Veh Technol 67(6):5445–5459
13. Vo NS, Duong TQ, Tuan HD, Kortun A (2017) Optimal video streaming in dense 5G networks with D2D communications. IEEE Access 6:209–223
14. Al-Halafi A, Oubei HM, Ooi BS, Shihada B (2017) Real-time video transmission over different underwater wireless optical channels using a directly modulated 520 nm laser diode. IEEE/OSA J Opt Commun Netw 9(10):826–832
15. Huang XL, Tang XW, Hu F (2019) Dynamic spectrum access for multimedia transmission over multi-user, multi-channel cognitive radio networks. IEEE Trans Multimedia 22(1):201–214
16. Yuvaraj N, Srihari K, Dhiman G, Somasundaram K, Sharma A, Rajeskannan SM, Soni M, Gaba GS, AlZain MA, Masud M (2021) Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking. Math Probl Eng
17. Raja RA, Kousik NV, Johri P, Diván MJ (2020) Analysis on the prediction of central line-associated bloodstream infections (CLABSI) using deep neural network classification. In: Computational intelligence and its applications in healthcare. Academic Press, pp 229–244
18. Veerappan Kousik NG, Natarajan Y, Suresh K, Patan R, Gandomi AH (2020) Improving power and resource management in heterogeneous downlink OFDMA networks. Information 11(4):203
19. Karthikeyan T, Praghash K (2021) An improved task allocation scheme in serverless computing using gray wolf optimization (GWO) based reinforcement learning (RIL) approach. Wireless Pers Commun 117(3):2403–2421
20. Wu J, Tan R, Wang M (2018) Streaming high-definition real-time video to mobile devices with partially reliable transfer. IEEE Trans Mob Comput 18(2):458–472
21. Khorov E, Ivanov A, Lyakhov A, Akyildiz IF (2018, Oct) Cloud control to optimize real-time video transmission in dense IEEE 802.11 aa/ax networks. In: 2018 IEEE 15th international conference on mobile ad hoc and sensor systems (MASS). IEEE, pp 193–201
22. Choi M, Kim J, Moon J (2018) Wireless video caching and dynamic streaming under differentiated quality requirements. IEEE J Sel Areas Commun 36(6):1245–1257
23. Nimmagadda P, Swamy KA, Prathima S, Chintha S, Alex ZC (2022) Short-term uncleaned signal to noise threshold ratio based end-to-end time domain speech enhancement in digital hearing aids. Indonesian J Electr Eng Comput Sci 27(1):131

Classifier-Free Guidance for Generative Adversarial Networks (GANs)

Ananya Chakravarthi and H. S. Gururaja

Abstract Classifier-free guidance is a recently discovered technique to enhance the images generated by diffusion models. This paper aims to use the same technique in the context of Generative Adversarial Networks (GANs) and contrast and examine the images produced. Its main focus is determining if classifier-free guidance is unique to diffusion models or if can it be applied to GANs too, and if the answer was found to be affirmative, in what manner. Previous research has shown that classifier-free guidance in diffusion models produces better images with a minimal impact on conditional modeling performance. With a guidance scale above the value of unity, the images shown are better in quality than those produced conditionally or unconditionally. Classifier-free guidance has yielded considerably better results for diffusion models. The images generated after performing classifier-free guidance show promise for generating high-quality images after sufficient training. Some images are already more photorealistic than those produced just conditionally. This shows that classifier-free guidance can be made conducive for other image generation models as well. It opens doors for further research on enhancing the image generation quality of models other than diffusion models as well and helps in making models like GANs successfully tackle the Generative Learning Trilemma.

Keywords Classifier-free guidance · Diffusion models · Generative adversarial network · Conditional GANs · Generative Learning Trilemma · Convolutional network · Dynamic thresholding

A. Chakravarthi (✉)
Department of EEE, B.M.S. College of Engineering, Bengaluru, India
e-mail: ananya.ee18@bmsce.ac.in

H. S. Gururaja
Department of ISE, B.M.S. College of Engineering, Bengaluru, India

1 Introduction

Any machine learning model can be categorized under two main types: discriminative and generative. The former deals with predicting results, while the latter generates the results. They are generally used for classification or regression. The goal of generative models is to create new samples by learning from the distribution of a dataset's distribution probability. Mathematically, generative models depend on the joint distribution probability $P(X, Y)$, whereas discriminative models depend on the conditional probability $P(X|Y)$ [1].

Image generation models have witnessed an incredible pace of advancement in recent months. Although GANs were until recently considered to be the best model for the purpose, it was subsequently shown that diffusion models outperform GANs in image synthesis [2, 10]. Denoising Diffusion Probabilistic Models (DDPMs) use the Markov chain process to incrementally denoise samples in order to generate images [3]. One of the main reasons for the superior performance of diffusion models is their ability to capture diversity better and provide better mode coverage. Their sampling procedure is also flexible and is amenable to modifications that result in faster sampling rates. Additionally, GANs suffer from mode collapse and vanishing gradients on occasion [4].

Although a variety of image generation models have been invented and the progress recorded has only accelerated, these models are yet to satisfy three key cornerstones of a robust image generation framework: high-quality samples, mode coverage, and fast sampling [5]. Novel techniques like classifier-free guidance and dynamic thresholding are at the heart of the rapid progress recorded [6]. However, the scope of the application currently limits its utility to a certain class of models. Making these techniques and innovations possible for image generation models like GANs can explore possibilities of achievement of higher quality images and effectively solve the Generative Learning Trilemma, as shown in Fig. 1 [5].

Classifier-free guidance was mentioned for the first time by Ho and Saliman [7]. It was later implemented practically by OpenAI in GLIDE [8] and DALL.E 2 [9] and most recently by Google's Imagen [6] to yield very high-quality results. Classifier-free guidance is considered to be critical in producing high-quality images but seems to limit diversity in the images generated. While it greatly improves the overall quality of the image and its correlation with the conditioned signal, it inevitably results in a compromise on diversity [7]. There are many other ways to enhance guidance for diffusion models too, as is evident from dynamic thresholding [6].

High guidance weights can often cause the pixel ranges of the training and testing dataset to mismatch, where the pixel values for the outputs exceed the bounds at a timestep. This can be addressed by static or dynamic thresholding. While static thresholding clips the distribution to the range of $[-1, 1]$, it creates oversaturation in the images. Dynamic thresholding overcomes this problem by operating on the pixel values to bring them within an absolute range gravitated toward zero.

With tricks like dynamic thresholding further solidifying the effectiveness of classifier-free guidance, the question then arises if classifier-free guidance can be
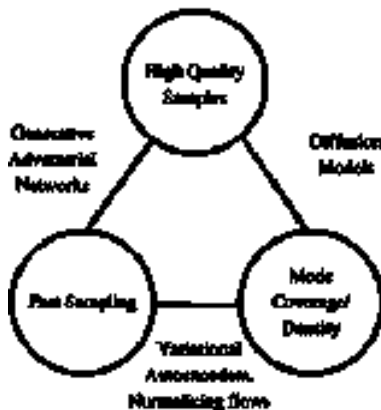
Fig. 1 Generative Learning Trilemma for three of the most commonly used models. GANs perform well when sampling rate and visual quality is concerned. Diffusion models show good results for visual quality and mode coverage. Variational autoencoders (VAEs) yield optimal images with fast sampling and good mode coverage. None of these models are yet to satisfy all three key requirements

implemented in models other than diffusion models after suitable modifications. Although classifier-free guidance has been studied in detail, its use for models other than diffusion models has received inadequate attention. The previous models dealing with this technique have been within the realm of diffusion models where it has proved to be very powerful in enhancing the results [7]. By exploring the application of classifier-free guidance for GANs, the scope of this guidance technique is widened. Since both GANs and diffusion models work on joint probability distributions, it is easier to implement guidance with GANs than on others like autoregressive models. This will serve as an impetus for further research in adapting classifier-free guidance for other techniques and attempting to overcome the drawbacks of each model. The implications of this deserve more attention and research.

The objective of this paper is to establish the suitability and potential of classifier-free guidance in the context of GANs. We investigated the images generated as a result and identified the limitations posed by adopting this approach. This paper first implements classifier-free guidance for GANs and then goes on to examine the results obtained and possible limitations. It also scrutinizes how the application of the technique in GANs differs from its application in the context of diffusion models.

2   Related Work

Before diffusion models achieved popularity, GANs were considered to be the best method available for the purpose of image generation. The idea of GAN architecture was first proposed by Ian Goodfellow et al. in "Generative Adversarial Networks"

[10]. GANs have been scrutinized, modified, and investigated over the years extensively [11–13]. Deep Convolutional Generative Adversarial Networks (DCGANs) have proven to be a powerful class of Convolutional Neural Networks (CNNs) [14]. StyleGAN is another GAN that is known to produce realistic images of the human face [15]. The original denoising diffusion method was introduced by Sohl-Dickstein et al. [16]. It was soon discovered that a classifier can transform an unconditional diffusion model to a conditional one in a post-hoc manner. This was called classifier guidance and was extensively explored in Dhariwal and Nichol [7]. It was shown that the quality of the samples generated increased considerably after classifier guidance. Apart from diffusion models and GANs, Google's Parti uses an autoregressive model which uses sequence-to-sequence modeling to provide a text-to-image translation [17]. OpenAI's DALL.E used variational autoencoders called Vector Quantized Variational AutoEncoders (VQ-VAE) [18].

Research comparing diffusion models to GANs has shown that the former has superior performance records as juxtaposed with the latter [2]. While the Markov Chain process and the parameters used for diffusion models prove to be the main contributing factor, the techniques used to amplify its effectiveness play a crucial supporting role. Diffusion models are also easily scalable. Classifier guidance has been implemented along with the main framework of diffusion models from the beginning to boost its results. It is also demonstrated to be computationally efficient [7].

Classifier-free guidance was first enunciated by Ho and Saliman in Classifier-free Diffusion Guidance in 2021 [7]. It was subsequently used by OpenAI in GLIDE where its impact was visible in the very high-quality and photorealistic images generated using text-guided diffusion [8]. Classifier-free guidance builds upon the enhancing efforts by introducing conditioning dropout where for a certain percentage, the conditioning information 'y' is replaced with null. By taking into account the scores of an unconditional model, this method utilizes differences in the latent space to create realistic, directionally aligned images. For semantic image synthesis, classifier-free guidance helps in better matching with semantic label maps and disentangles the noise from the conditional component with the unconditional component. The major trade-off is maintained by the guidance scale hyperparameter [19].

It was also used to good effect in DALL.E 2 by OpenAI in April 2022 [20]. It was observed that with a guidance strength > 1, there was a very noticeable impact on the quality of images produced. DALL.E 2 used CLIP which works on a multimodal latent space with both text and image points, resulting in better text-to-image translations [21]. It is also a good example of zero-shot learning where predictions are based upon samples not encountered during training [9]. Most recently, it was used by Imagen by Google where it was a critical contributor to the results the model yielded [6]. Notably, Imagen also initiated the use of dynamic thresholding to enhance the guidance process.

The question of whether classifier-free guidance is unique to diffusion models has not been studied or examined yet in adequate detail. It is also unknown if, with required modifications, classifier-free guidance can be made tractable for models

like GANs. This paper attempts to initiate research in this direction and investigate the questions posed.

3 Proposed Method

To implement classifier-free guidance, we need both conditional and unconditional results. The results obtained from both functions are then combined, and using a guidance scale, the guidance is implemented.

3.1 Preliminaries

The generator and the discriminator are the two major components used to design GANs [10]. The generator produces a sampled output after receiving random noise as input. The random noise vector is sourced from the latent space. Original data are fed into the discriminator which then acts as a binary classifier to determine if the output from the generator is fake or real. The discriminator seeks to maximize the loss/value function, whereas the generator seeks to minimize it. Hence, both components work adversarially with each other to achieve individual objectives. However, this results in a feedback flow that finally converges Pg (probability distribution of the generator) with Pdata (probability distribution of original data), as illustrated in Fig. 2.

Fig. 2 Basic structure of a GAN model. The generator creates a fake image using input of random noise. The discriminator compares a real and a fake image and, using binary classification, determines whether the image from the generator is real or fake. The discriminator's goal is to increase the combined loss function, whereas the generator wants to reduce it
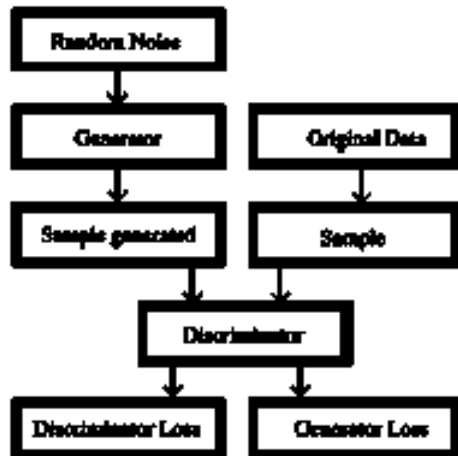
3.2 Classifier-Free Guidance for Diffusion Models

Latent variable models include diffusion models. Gaussian noise is added at each timestep to gradually destroy the original data and corrupt it. The forward process can be expressed mathematically as:

$$q(x1:T|x0) = \Pi Tt = 1q(xt|xt-1)$$
$$= \Pi Tt = 1\mathcal{N}\left(xt; \sqrt{1-\beta txt-1}, \beta tI\right), \tag{1}$$

where β is the variance schedule. In the backward process, the model gradually denoises a noise-corrupted image to learn the joint probability distribution as shown in Fig. 3:

$$p\theta(x0:T) = p(xt)\Pi Tt = 1 p\theta(xt-1|xt)$$
$$= p(xt)\Pi Tt = 1\mathcal{N}\left(xt-1; \mu\theta(xt,t), \sum \theta(xt,t)\right). \tag{2}$$

Diffusion models predict the score function, $\nabla x\log p(x)$ instead of $p(x)$, where $p(x)$ is the high-dimensional data distribution. When classifier guidance is applied to diffusion models, we get:

$$\nabla x \log p\nabla(x|y) = \nabla x \log p(x) + \gamma\nabla x \log p(y|x), \tag{3}$$

where γ is the called the guidance scale and $\nabla x\log p(y|x)$ is the conditioning term.

Further, from Bayes rule, we have:

$$p(y|x) = p(x|y).p(y)/p(x). \tag{4}$$

This gives us:

Fig. 3 Forward and backward diffusion processes. An original image X0 is gradually corrupted with Gaussian noise at each timestep. This corrupted image is then denoised by the diffusion model. In the process, the diffusion model learns how to generate an image. The entire flow is a parameterized Markov Chain
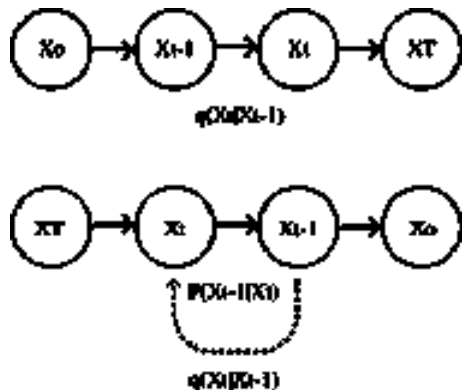
$$\log p(y|x) = \log(x|y) + \log p(y) - \log p(x). \tag{5}$$

Applying the score function, we get:

$$\nabla x \log p\nabla(y|x) = \nabla x \log p\nabla(x|y) - \nabla x \log p(x). \tag{6}$$

To perform classifier-free guidance, we then obtain the following equation after substituting the above values in the formula for classifier guidance:

$$\nabla x \log p\nabla(x|y) = (1 - \gamma)\nabla x \log p(x)$$
$$+ \gamma \nabla x \log p\nabla(x|y). \tag{7}$$

3.3   Unconditional GANs

In our study, the Fashion MNIST dataset was utilized. It consists of 60,000 tiny 28 × 28 pixel square grayscale photos of objects divided into ten categories of apparel. Additionally, it has 10,000 images in the test set [22].

GANs use both discriminator and generator models. A point in the latent space is provided to the generator as the input. In response, the generator creates a single 28 × 28 grayscale image as an output. The discriminator model then predicts whether the image created by the generator is fake or real. It takes one 28 × 28 grayscale image as input and gives a binary prediction on the authenticity of the image.

Taking the loss functions of the generator and the discriminator together, we get:

$$\min \max V(D, G) = \min \max(Ex \sim \text{Pdata}(x)\big[\log D(x)\big]$$
$$+ Ez \sim P(z)\big[\log(1 - D(G(z)))\big]. \tag{8}$$

This is called the combined loss function or the value function of the entire model. The model begins generating acceptable images after training for 20 epochs. Both the parts of the GAN are trained simultaneously until the generator starts producing plausible items and the discriminator classifies the images generated by the generator as real about 50% of the time.

3.4   Conditional GANs

The unconditional GAN model is updated by making the input image to the discriminator conditioned on the class label [23, 24], demonstrated in Fig. 4. The generator model also takes the class label to make the latent space now conditional. The combined loss function is modified as:
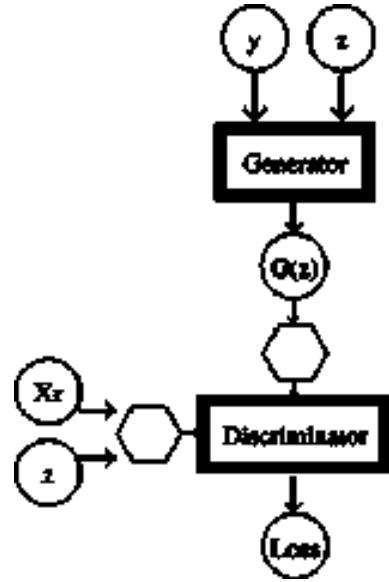
$$\min \max V(D, G) = \min \max(Ex \sim \text{Pdata}(x)$$
$$[\log D(x|y)] + Ez \sim P(z)[\log(1 - D(G(z|y)))]. \tag{9}$$

The architecture of GAN is modified. An additional input channel is created to accept the class label as the conditional inputs to the generator and the discriminator.

An embedding layer with a fully connected layer is implemented. Linear activation is then performed that scales and then concatenates the size of the image to the model, which then acts as an additional channel.

3.5 Classifier-Free Guidance for GANs

After plotting images for conditional and unconditional GAN, we use the results generated to carry out classifier-free guidance, as explained in Fig. 5 [7]. We define a guidance scale (γ) and use it to scale the difference between the conditionally and the unconditionally generated images. This can be shown as:

$$\grave{\varepsilon}\Theta(xt|c) = \varepsilon\Theta(xt|\Phi) + s.(\varepsilon\Theta(xt|c) - \varepsilon\Theta(xt|\Phi)), \tag{10}$$

where c is the conditioning label, ε_Θ is the prediction, and $\grave{\varepsilon}_\Theta$ is the modified prediction.

Fig. 5 Difference between the conditionally and unconditionally generated data points is multiplied by the hyperparameter '*s*' (indicating the guidance scale), and it is vectorially added to the unconditional data point
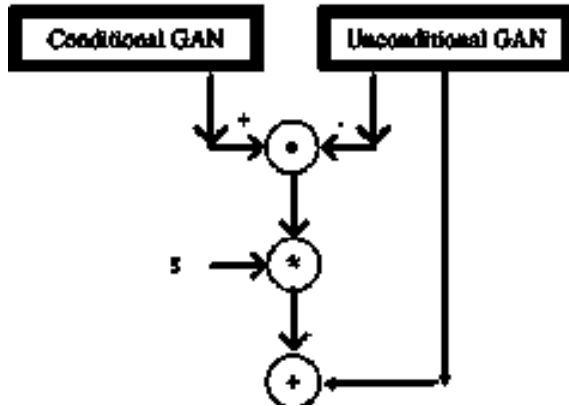
Algorithm: Stochastic Gradient Descent for Training of Proposed GAN with Classifier-Free Guidance

Input: For input vector: Z, samples $X0, X1, \ldots. Xn$, condition labels $C0, C1, \ldots., Cn$
'k' and 's' are hyperparameters

for no. of training iterations do
for steps of k do
Update the discriminator and ascend the stochastic gradient descent:
$\nabla \Theta d (1/m) \Sigma i = 1 \big[\log D + \log(1 - D(G(z))) \big]$
end for
Generator Update is Performed and descend the stochastic gradient descent:
$\nabla \Theta d (1/m) \Sigma i = 1 \big[\log(1 - D(G(z))) \big]$
end for
Apply classifier free guidance using the formula:
$\nabla \Theta (xt|c) = \varepsilon \Theta (xt|\Phi) + s.(\varepsilon \Theta (xt|c) - \varepsilon \Theta (xt|\Phi))$

4 Result Analysis

4.1 Dataset

The Fashion MNIST dataset [22] is used in the experiment. It has 785 columns and two sets: a training set of 60,000 photos and a testing set of 10,000 images. Each image has a height of 28 pixels and a width of 28 pixels for a total of 784 pixels. An integer from 0 to 255 is represented by each pixel's value. They are in grayscale and each is a fashion or a clothing item. They include *t*-shirts, sneakers, dresses, sweaters, etc. There are ten classes based on which the images can be labeled and associated.

There are 7000 images per category. Fashion MNIST is a dataset that can be used in place of the MNIST dataset for benchmarking purposes.

4.2 Implementation Details

Convolutional neural networks are used to implement the discriminator model. It utilizes the LeakyReLU activation function. A 2×2 stride and a slope value of 0.2 are used for downsampling. We utilize the existing Adam version for stochastic gradient descent and set the learning rate to 0.0003 and the momentum to 0.6. An input to the generator model is a random noise vector that represents a point in the latent space. Additionally, it uses a hyperbolic tangent activation function and LeakyReLU activation. With the help of this, it decodes a point in the latent space and outputs a low-resolution image. This is upsampled two times using transpose convolutional layers. The latent space dimension is given as a parameterized value to the function argument. The model is then trained for 50 epochs.

It begins producing legible images after 20 training epochs. After performing the entire methodology on conditional and unconditional GAN models, we combine the results using the formula for classifier-free guidance with the guidance scale to obtain results. We use values of $\gamma = 3, 5, 6$ on the dataset for the hyperparameter of the guidance scale. The whole experiment is implemented using Keras, TensorFlow, and Google Colab.

4.3 Comparison with Previous Results

Figures 6 and 7 correspond to the images generated via conditional and unconditional GAN. Figures 8, 9, and 10 correspond to results from the implementation of classifier-free guidance with the hyperparameter of guidance scale set to 3.0, 5.0, and 7.0, respectively.

4.4 Model Evaluation

The primary goal of the task is to assess the photos' visual quality. To measure the quality, the well-known Fréchet Inception Distance (FID) metric is used [25]. The FID score computes how similar two images are and produces a score using computer vision statistics. The FID score and the similarity of the image with the original data distribution have an inverse relationship; as the score decreases, the similarity increases.

While it is observed from Table 1 that the FID score of images generated by classifier-free guidance performed on GANs is still lower than the others with a

Fig. 6 Unconditional GAN performed on fashion MNIST dataset

Fig. 7 Conditional GAN performed on fashion MNIST dataset

Fig. 8 Results from classifier-free guidance with $\gamma = 3.0$

Fig. 9 Results from classifier-free guidance with $\gamma = 5.0$

Fig. 10 Results from classifier-free guidance with $\gamma = 7.0$

Table 1 FID scores for other models using the fashion MNIST dataset

Models	FID scores for the fashion MNIST dataset
Proposed classifier-free guidance for GANs	32
PAE [26]	28
PeerGAN [27]	21.73
Sliced iterative generator [28]	13.7

best score of 32 on the experimental setups performed so far, the technique shows considerable promise for further improvement.

As can be seen from Fig. 11, the FID score of classifier-free guiding rises as expected as the model is trained with a larger number of epochs. Better the convergence with the original (actual) sample distribution, lower the FID score [25].

5 Conclusion and Future Work

While the effectiveness of classifier-free guidance for diffusion models has been established, its utility for other image generation models has remained largely unexplored. By testing the applicability of this technique for Generative Adversarial Networks, we find that with sufficient training, classifier-free guidance can enhance
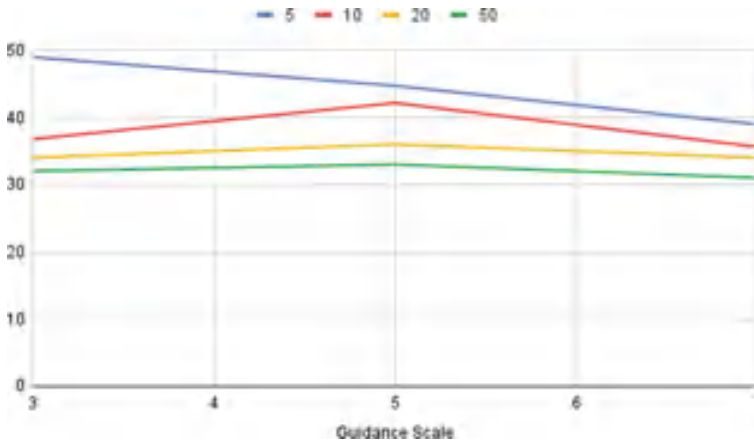
Fig. 11 Plotting the FID score against the number of training epochs across guidance scales 3.0–7.0

the results of the model. For some images, the quality produced is noticeably better than when generated by conditional GANs. As more and more epochs are trained, the standard of images generated by the classifier-free guidance technique also shows considerable improvement, thereby displaying promise for improvement in GANs.

Future research can be conducted for other image generation models like variational autoencoders and their compatibility with classifier-free guidance [29]. While some experiments with classifier-free guidance and autoregressive models have been done, there are some significant differences in their application to generative models and discriminative models like autoregressive ones. Autoregressive models work with sequential logits and conditional distributions. This requires another approach for guidance as applying the technique to joint and conditional probability distributions differ in mathematical and computational implementation.

Further improvement in using the technique for GANs is also possible. It must be determined whether static and dynamic thresholding can be used with GANs [6]. Also, different conditioning dropout percentages can be compared to observe their impact on the results. Further innovations are also possible in latent spaces from where generators source the random noise inputs. Multimodal latent spaces can be explored to establish relationships between image and text data points [30]. Different datasets can be used and the convergence over epochs trained can be tracked.

Although the model used in this paper was a class-conditional GAN, GANs can be conditionally aligned with images and entities like text too. Future research can involve classifier-free guidance image-to-image translations and other conditional inputs. It can also explore guidance-enhancing tricks and devise new ones apart from dynamic thresholding. These tricks can also help in compensating for the loss of diversity that results from the use of classifier-free guidance.

References

1. Nian YW, Gao R, Han T, Zhu S-C (2018) A tale of three probabilistic families: discriminative, descriptive and generative models. arXiv:1810.04261
2. Dhariwal P, Nichol A (2021) Diffusion models beat GANs on image synthesis. arXiv:2105.05233
3. Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. arXiv:2006.11239
4. Thanh-Tung H, Tran T (2020) Catastrophic forgetting and mode collapse in GANs. arXiv:1807.04015
5. Xiao Z, Kreis K, Vahdat A (2022) Tackling the generative learning trilemma with denoising diffusion GANs. arXiv:2112.07804
6. Saharia S, Chan W, Saxena S, Li L, Whang J, Denton E, Ghasemipour SKS, Ayan BK, Mahdavi SS, Lopes RG, Salimans T, Ho J, Fleet DJ, Norouzi M (2022) Photorealistic text-to-image diffusion models with deep language understanding. arXiv:2205.11487
7. Ho J, Salimans T (2022) Classifier-free diffusion guidance. arXiv:2207.12598
8. Nichol A, Dhariwal P, Ramesh A, Shyam P, Mishkin P, McGrew B, Sutskever I, Chen M (2021) GLIDE: towards photorealistic image generation and editing with text-guided diffusion models. arXiv:2112.10741
9. Ramesh A, Pavlov M, Goh G, Gray S, Voss C, Radford A, Chen M, Sutskever I (2021) Zero-shot text-to-image generation. arXiv:2102.12092
10. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks. arXiv:1406.2661
11. Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of GANs for improved quality, stability, and variation. arXiv:1710.10196
12. Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X, He X (2018) Attngan: fine-grained text to image generation with attentional generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1316–1324
13. Tang H, Qi X, Xu D, Torr P, Sebe N (2020) Edge-guided GANs with semantic preserving for semantic image synthesis. arXiv:2003.13898
14. Radford A, Metz L, Chintala S (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434
15. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. arXiv:1812.04948
16. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B (2021) Score-based generative modeling through stochastic differential equations. In: International conference on learning representations
17. Yu J, Xu Y, Koh JY, Luong T, Baid G, Wang Z, Wu Y (2022) Scaling autoregressive models for a content-rich text-to-image generation. arXiv:2206.10789
18. Gu S, Chen D, Bao J, Wen F, Zhang B, Chen D, Guo B (2022) Vector quantized diffusion model for text-to-image synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10696–10706
19. Wang W, Bao J, Zhou W, Chen D, Yuan L, Li H (2022) Semantic image synthesis via diffusion models. arXiv:2207.00050
20. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M (2022) Hierarchical text-conditional image generation with CLIP latents. arXiv:2204.06125v1
21. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning transferable visual models from natural language supervision. arXiv:2103.00020
22. Xiao H, Rasul K, Vollgraf R (2017) Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. arXiv:1708.07747
23. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv:1411.1784
24. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B (2018) High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8798–8807

25. Heusel M, Ramsauer H, Unterthiner T, Nessler B (2018) GANs trained by a two time-scale update rule converge to a local nash equilibrium. arXiv:1706.08500
26. Bohm V, Seljak U (2022) Probabilistic autoencoder. arXiv:2006.05479
27. Wei J, Liu M, Luo J, Li Q, Davis J, Liu Y (2021) PeerGAN: generative adversarial networks with a competing peer discriminator
28. Dai B, Seljak U (2021) Sliced iterative normalizing flows. arXiv:2007.00674
29. Kingma DP, Welling M (2019) An introduction to variational autoencoders. arXiv:1906.02691
30. Nair NG, Bandara WGC, Patel VM (2022) Image generation with multi-modal priors using denoising diffusion probabilistic models. arXiv:2206.05039

Segmentation of Lung Regions for the Detection of Juxta-Pleura Nodules in CT Scan

B. Sasidhar [ORCID]

Abstract Extraction of lung fields helps in reducing the time taken for the recognition of juxta-pleura lung nodules. The existing approaches cannot extract pleura with juxta-pleura nodules. In proposed approach, active contour, region growing, morphological operators, convex hull, and XOR operation are used to extract pleura with juxta-pleura nodules. The proposed approach is tested on 100 images and an average overlap measure (AΩ) of 99.0% is calculated.

Keywords Pleura's · Morphological operators · Thresholding · Active contour · Region growing · Convex hull

1 Introduction

Good health is crucial for the well-being of mankind. However, modern lifestyle and toxic environment have increased the rate of life-threatening diseases like cancer. Cancer has become one of the primary reasons for death around the world. National Cancer Registry Programme reveal a dreadful fact that the main types of cancers among men are oral cavities, pulmonary, trachea, and abdomen and among women are cervical, breast, and oral cavities. Among these, the oral and lungs in men and cervix and cancer of breast in women, account for over 50% of all cancer deaths in India.

The second leading cause of mortality rate in both male and female is lung cancer [1]. About one out of four cancer mortalities is from lung cancer. Lung carcinoma can be cured at the early stage, and Computer-Aided Diagnosis (CAD) plays a significant role in aiding radiologist for better identification for the carcinoma of lungs using Computed Tomography (CT) Scan images. Carcinoma of lungs is curable at the early stage, but detecting it at this early stage is an onerous task. In recent years, computer

B. Sasidhar (✉)
Department of Computer Science and Engineering, GNITS, Hyderabad, India
e-mail: bolasasi37@gmail.com

233

vision and deep learning approaches are used in diagnosis of lung cancer, which is known as CAD from various imaging modalities.

CAD system comprises segmentation of lungs, detection of nodules, and classification of tumour as benign or malignant. Pleura extraction helps in speed up the process of identification of nodules, and it reduces radiologist time in diagnosis of carcinoma of lungs. The difficult task is to isolate lungs with nodules at the pleura walls. In Fig. 1, nodule at the border of the pleura is shown with the rectangular box in the red colour.

The previous approaches [2, 3] could not segment the lungs with the nodules at the lung walls. In proposed approach, active contour, region growing, morphological operations, convex hull, XOR operations are used for the lungs. To segment the lungs, intensity values for different parts of the CT scan image are needed. In Table 1, the attenuation values of the different tissues of Computed Tomography (CT) scan images are given.

Dataset Used

The datasets are downloaded from the Lung Image Database Consortium image collection (LIDC-IDRI) which contains diagnostic and carcinoma of lung CT scans with marked annotated lesions.

Datasets used computer vision techniques that can guide in tumour identification on CT images. The use of such techniques may considerably increase the sensitivity

Fig. 1 Juxta-pleura nodule in the bounding box (red label)
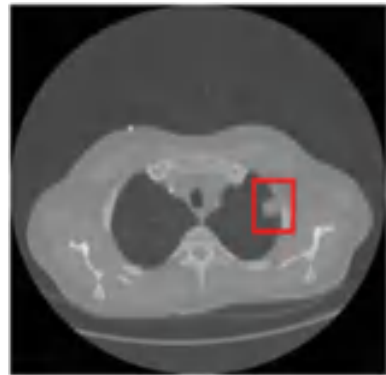
Table 1 Attenuation values of the different tissues [4]

Substance	HU range
Air	− 1000
Lung	− 400 to − 600
Nodule	− 150
Fat	− 60 to − 100
Water	0
Soft tissue	+ 40 to + 80
Bone	+ 400 to + 1000

and specificity of CT pleura images and also reduce radiologist time needed for diagnosis. The proposed approach is tested and experimented on a dataset comprising various kinds of CT images.

The remaining paper is divided into Sect. 2 which is related work, Sect. 3 for proposed methodology, Sect. 4 is for results and discussion, and Sect. 5 is for conclusion.

2　Related Work

Tseng and Huang [1] have used threshold technique for the lungs, but it does not extract the nodules at the lung walls. Itai et al. [5] have used the active contour technique for the extraction of the lungs, but the drawback is major disadvantage is that semi-automated [6, 7]. Active Contour Model (SNAKE) goes towards the boundary of the desired object and to extract borders of the object. Mesanovic et al. [2] have proposed the region growing technique for the extraction of lungs. Al-Fahoum et al. [8] have segmented both lungs using thresholding and labelling.

Arfan Jaffar et al. [9] used Genetic Algorithm and morphological operators for the extraction of lungs and the proposed approach is automated. Lai et al. [10] have used active contour for the lungs with juxta-pleura nodules. El-Regaily et al. [11] have applied morphological operators for the extraction of lungs. Vas and Dessai [12] have applied closing of morphological operators for the extraction of pleura's.

Paing et al. [13] had applied threshold and morphological operators for lungs. Nadkarni and Borkar [3] applied morphological operation on input images for the Pleura's. Vas and Dessai [14] have used binary of input image, morphological opening and closing operations for pleura's. Halder et al. [15] used threshold algorithm in an iterative way with morphological closing operations for pleura extraction. Chithra and Renjen Roy [16] used "Otsu's thresholding" for the pleura extraction in CT images. Gite et al. [17] have done lung segmentation using U-NET++. [18] have applied threshold on the input image, then morphological processing, elimination of blob's for the pleura extraction.

Gite et al. [19] have used U-NET model for the pleura of the CT images and [20] have applied watershed with marker-based approach for the extraction of pleura. From the above existing approaches, automated approaches are not applied and other approaches average overlap (AΩ) needs to increase.

3　Proposed Methodology

The proposed approach is based on Morphological Operators, Convex Hull, and XOR operations for the segmentation of pleura's with the tumours at the lung walls. The process of the proposed approach contains: thresholding [4] is used to get different components of image and then apply attenuation value of the blood vessels in lungs.

Compute the locations of left and right lungs [i.e. $(x1, y1)$, $(x2, y2)$]. Active contour is applied with blood vessels as starting point in the lungs to grow in the lungs. Use region growing [21] with seed points for the lungs without nodules at the lung walls. To extract pleura with nodules at the lung walls, morphological operators, Convex Hull, XOR operation are used for generation of mask. Keep the mask on the input image for the segmentation of pleura. The proposed algorithm is shown in algorithm 1.

Algorithm 1: Pseudo Code for the Extraction of Lungs

Input: CT Scan Image
Output: Pleura Nodules with lung regions
Start

Step 1.   Thresholding is applied for the extraction of different lung parts.
Step 2.   Find the attenuation value of the blood vessels in the lungs (HU)
Step 3.   Find the location of blood vessels for the left lung $(x1, y1)$ and right lung $(x2, y2)$.
Step 4.   Region growing [21] with seed point as $(x1, y1)$ for the segmentation of left lung.
Step 5.   Region growing [21] with seed point as $(x1, y1)$ for the segmentation of left Lung.
Step 6.   Combine step 4 and step 5 for the generation of mask.
Step 7.   Use active contour with the blood vessels as the starting points on left and right lungs for the generation of the mask.
Step 8.   For the extraction of lungs with the nodules at the lung walls, convex hull, XOR and morphological operations are used on step 6 and 7.
Step 9.   Combine lung mask of step 6 and step 7.
Step 10.   The result of step 9 is overlapped with the DICOM input image for the segmentation of lungs.

Stop

4   Results and Discussion

The proposed approach is tested on 15 patients' datasets with 100 images, and the performance metric for the proposed approach is overlap measure (Ω) [22].

The average overlap (A Ω) measure in % can be calculated by using (1)

$$A\Omega = \frac{(\Omega 1 + \Omega 2 + \Omega 3 + \Omega 4 + \cdots + \Omega n)}{n}, \qquad (1)$$

where Ω_i (overlap measure), $i = 1$ to n, and n is number of images.

In Fig. 2, input image, lungs with the help of [2, 3], and proposed system Ω are shown.
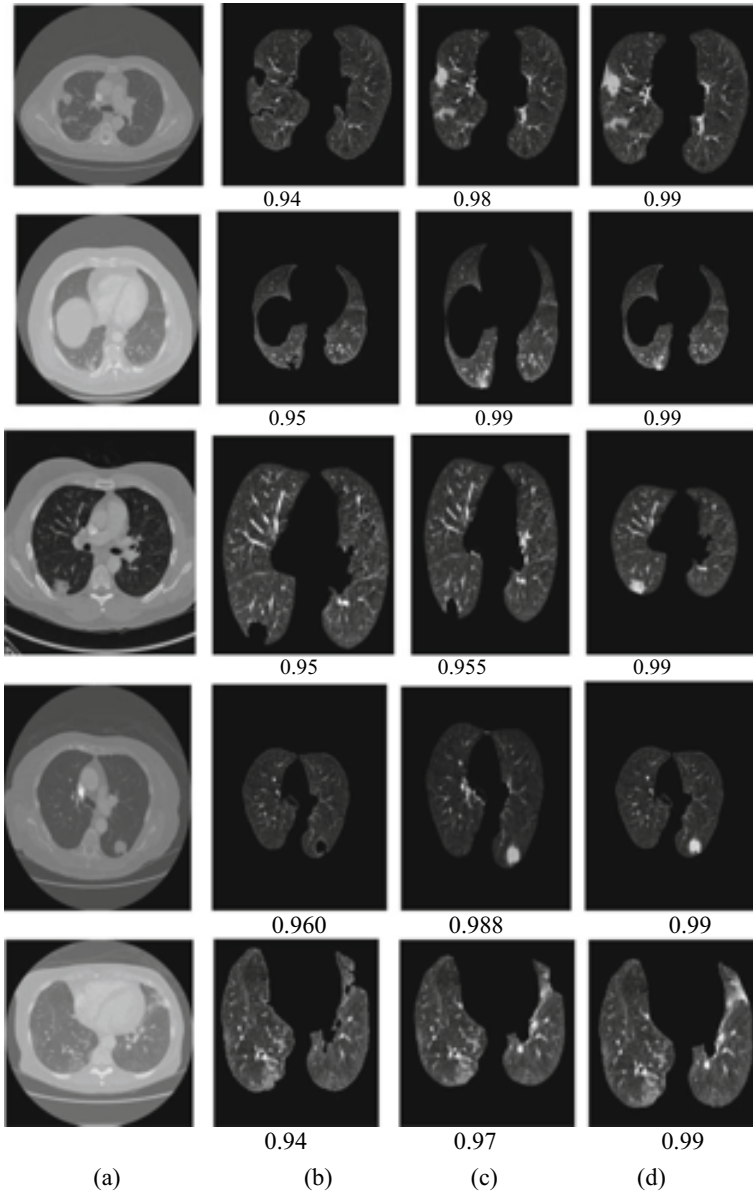
Fig. 2 **a** Input image, **b** lungs using [2] with Ω, **c** lungs using [3] with Ω, **d** proposed approach for the segmentation of lungs with Ω

Table 2 Average overlap measure (AΩ)

AΩ region growing [2]	AΩ morphological operators [3]	AΩ proposed system
94.8	97.66	99.0

In Table 2, AΩ of existing approaches, proposed approaches [2, 3] is calculated.

5   Conclusion

AΩ of [2, 3] was not able to get the lungs with the nodules at the lung walls. The proposed approach segments lungs with the nodules at the lung walls.

The proposed approach is better than existing approaches in terms of accuracy which is 99.9%.

References

1. Tseng L-Y, Huang L-C (2009) An adaptive thresholding method for automatic lung segmentation in CT images. IEEE AFRICON
2. Mesanovic N, Grgic M, Huseinagic H, Males M, Skejic E, Smajlovic M (2011) Automatic segmentation of the lungs with region growing algorithm. IWSSIP, June 2011
3. Nadkarni NS, Borkar S (2019) Detection of lung cancer in CT images using image processing. In: Proceedings of the third international conference on trends in electronics and informatics (ICOEI 2019). IEEE
4. Habib MSL (2009) A computer aided diagnosis system (CAD) for the detection of pulmonary nodules on CT scans. Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Giza, Egypt
5. Itai Y, Kim H, Ishikawa S, Katsuragawa S, Ishida T, Nakamura K, Yamamoto A (2005) Automatic segmentation of lung areas based on SNAKES and extraction of abnormal area. In: Proceedings of the 17th IEEE international conference on tools with artificial intelligence (ICTAI'05)
6. Talakoub O, Alirezaie J, Babyn P (2007) Lung segmentation in pulmonary CT images using wavelet transform. ICASSP, IEEE
7. Amutha A, Wahidabanu RSD (2013) Lung tumor detection and diagnosis in CT scan images. Int Conf Commun Signal Process (ICCSP) 2013:1108–1112
8. Al-Fahoum AS, Jaber EB, Al-Jarrah MA (2014) Automated detection of lung cancer using statistical and morphological image processing techniques. J Biomed Graph Comput 4(2)
9. Arfan Jaffar M et al (2008) GA and morphology based automated segmentation of lungs from CT scan images. IEEE
10. Lai J et al (2009) Active contour based lung field segmentation. In: International conference on intelligent human-machine systems and cybernetics. IEEE
11. Amin El-Regaily S, Salem MAM, Aziz MHA, Roushdy MI (2017) Lung nodule segmentation and detection in computed tomography. In: The 8th IEEE international conference on intelligent computing and information systems (ICICIS 2017)
12. Vas M, Dessai A (2017) Lung cancer detection system using lung CT image processing. In: International conference on computing, communication, control and automation (ICCUBEA), IEEE

13. Paing MP, Hamamoto K, Tungjitkusolmun S, Pintavirooj C (2019) Automatic detection and staging of lung tumors using locational features and double-staged classifications. Appl Sci 9:2329
14. Vas M, Dessai A (2017) Lung cancer detection system using lung CT image processing. In: International conference on computing communication control and automation (ICCUBEA). IEEE Explore
15. Halder A, Chatterjee S, Dey D (2020) Superpixel and density based region segmentation algorithm for lung nodule detection. In: Proceedings of 2020 IEEE Calcutta conference (CALCON)
16. Chithra AS, Renjen Roy RU (2018) Otsu's adaptive thresholding based segmentation for detection of lung nodules in CT image. In: Proceedings of the 2nd international conferences on trends in electronics and informatics (ICOEI 2018). IEEE Xplore
17. Gite S, Mishra A, Kotecha K (2021) Enhanced lung image segmentation using deep learning. In: Neural computing and applications. Springer
18. Abdullah MF, Sulaiman SN, Osman MK, Karim NKA, Setumin S, Isa IS (2022) A new procedure for lung region segmentation from computed tomography images. Int J Electr Comput Eng (IJECE)
19. Khan A, Garner R, Rocca ML, Salehi S, Duncan D (2022) A novel threshold-based segmentation method for quantification of COVID-19 lung abnormalities. Signal Image Video Process
20. Bougourzi F, Contino R, Distante C, Taleb-Ahmed A (2021) Recognition of COVID-19 from CT scans using two-stage deep-learning-based approach: CNR-IEMN. Sensors
21. Adams R, Bischof L (1994) Seeded region growing. IEEE Trans Pattern Anal Mach Intell 16(6)
22. Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Xue Z, Karargyris A, Antanu S, Thoma G, McDonald CJ (2013) Lung segmentation in chest radiographs using anatomical atlases with non-rigid registration. IEEE Trans Med Imag 33(2)

Analysis of Serial-In Parallel-Out Finite Field Multiplier Using Various Domino Logic Styles

Bhavana Majji and K. Ragini

Abstract Digital privacy and authenticity are the main factors that have developed as a result of migration from paper to electronic media. Effective cryptographic solutions are necessary to preserve digital privacy. Reordered normal basis (RNB) is used for cryptographic solutions. In this paper, a 11-bit serial-in parallel-out (SIPO) RNB multiplier is implemented using various domino logic styles. The XOR-AND-XOR (XAX) module is implemented using various logic styles like domino, domino keeper, and NP domino logic in the Tanner EDA tool utilizing the 45 nm technology. The design parameters like area, power dissipation, multiplication delay, CPD, area-delay product, and power-delay product of the multiplier implemented using these logic styles are evaluated and compared. When compared to other domino logics, the delay and area of the multiplier employing NP domino logic are lower, whereas the power dissipation is similar to other domino logics. Hence, NP domino logic is best suited for high-speed applications.

Keywords Reordered normal basis · Finite field multiplier · Domino keeper logic · NP domino logic · 45 nm technology

1 Introduction

In the realm of communications, Galois fields have gained a lot of attention where error-correcting codes [1] and cryptography [2] are just a few of the uses. For switching theory and digital signal processing, the finite field GF (2^m) is very appealing [3]. As a result, good multiplication and inversion algorithms are required in finite fields [4]. In hardware implementations, the normal basis is more efficient because a simple cyclic shift can be used to square a field element over GF (2^m). For

B. Majji (✉) · K. Ragini
Department of Electronics and Communication Engineering, G. Narayanamma
Institute of Technology and Science, Hyderabad, India
e-mail: bhavana.maroju@gmail.com

cryptosystems that use frequent squaring, this makes the normal basis more desirable [5]. The type-II optimal basis multiplier, also known as the RNB multiplier, allows multiplication operation in the closed form rather than matrix operation [6]. The multiplier [7] presented in this paper produces the outputs in parallel fashion following the RNB multiplication operation.

In the literature, several potential architectures for sequential RNB/type-II ONB multipliers have been presented. Massey-Omura [8] developed an efficient normal basis bit-serial multiplier with 2 XOR and 1 AND plane. It was later extended to bit-parallel architecture. Gao-Sobelman [9] has the same complexity as the Massey-Omura multiplier except the planes are rearranged as XOR-AND-XOR planes. Namin [7] proposed a SIPO RNB multiplier and used domino keeper logic to implement the idea in order to build the word level RNB of the XAX module. This design was modified [8] to decrease the contention current, which had a significant improvement in functionality at high operating speeds. Later, Abhinav. V. Deshpandey [10] implemented the multiplier presented by Namin [8] using the pseudo NMOS logic.

In this paper, 11-bit SIPO RNB multiplier is implemented using the various logic styles like domino logic, domino keeper logic, and NP domino logic. The design parameters like area, power dissipation, multiplication delay, CPD, area-delay product (ADP), and power-delay product (PDP) of the multiplier implemented using these logic styles are evaluated and compared. The optimal logic style is hence determined by analysing these design parameters. The organization of this paper is as follows: The information in Sect. 2 pertains to the RNB multiplication process. The functionality of the SIPO RNB multiplier is detailed in Sect. 3. The NP domino logic along with domino logic and domino keeper logic is explained in Sect. 4. In Sect. 5, the results of 11-bit SIPO RNB multiplier and analysis of its design parameters are presented. Section 6 covers conclusion of this paper.

2 RNB Multiplication Process

Let the polynomials A and B that are a part of GF (2^m) be represented in RNB basis [7] as

$$A = \sum_{i=1}^{m} a_i \gamma_i,$$
$$B = \sum_{i=1}^{m} b_i \gamma_i.$$

The element C which is the multiplication of two elements is stored in the same basis as

$$C = \sum_{i=1}^{m} c_i \gamma_i, \quad \text{where} \quad a_i, b_i, c_i, \in F_2.$$

The product coordinates, c_i are compared as given below:

$$c_i = \sum_{j=0}^{m} a_i \big[b_{s(i+j)} + b_{s(j-i)} \big], \quad i = 1, 2, \ldots, m,$$

where $s(i)$ maps the set of integers to the set $\{0, 1, \ldots, 2m + 1\}$ which is defined as

$$s(i) \triangleq \left\{ \begin{array}{ll} i \bmod 2m + 1, & 0 \leq i \bmod 2m + 1 \leq m \\ 2m + 1 - i \bmod 2m + 1, & \text{otherwise} \end{array} \right\}.$$

3 Operation of Serial-In Parallel-Out Multiplication Operation

The multiplier designed by Namin [8] is a very regular structure. When compared to other structures identical to [7], the architecture utilized in [8] will have the shortest CPD. Figure 1 depicts a 11-bit SIPO RNB multiplier presented in this paper.

The inputs A and B are multiplied, resulting in an output C. The multiplier's critical path incorporates a series of XOR-AND-XOR gates. The function $((b_0 \oplus b_1).a_{11}) \oplus c_{11}$ is implemented by using the XOR-AND-XOR path, where b_0, b_1, a_{11}, c_{11} are the inputs to the XAX module. The process of multiplication is clearly represented in form of a flow chart which is shown in Fig. 2.
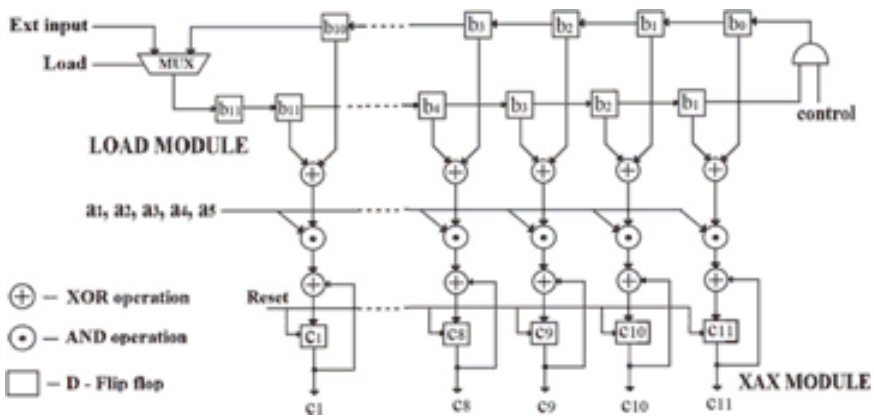
Fig. 1 Block diagram of 11-bit SIPO RNB multiplier
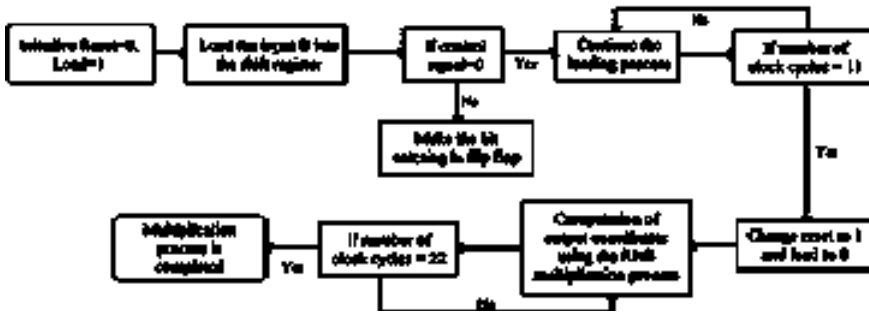
Fig. 2 Flow chart of multiplication process of SIPO RNB multiplier

The multiplier circuit runs on the clock's negative edge, giving the domino logic styles enough time to evaluate the result. The input is loaded into the shift register when the multiplexer's load signal is high and cuts off after loading the input, resulting in a circular shift register. As soon as the load signal goes low, the multiplication process starts. Initially, b_0 is forced to zero.

The control signal is made zero in the first clock cycle and then changed to 1. The input coordinates of B will be shifted circularly in each clock cycle. The input-A will be serially fed into the XAX module. The clock signal is used to synchronize the flip-flops. The reset signal is given to the output flip-flop in order to reset the previous value of the flip-flop. Throughout loading, the reset value is 0, and during multiplication, it remains 1. At the end of each clock cycle, partial products are created. After five clock cycles, the output coordinates are generated.

4 Different Domino Logic Styles

To minimize the power dissipation caused by short circuit current in static CMOS logic, dynamic logic was introduced. Dynamic logic works in two phases: precharge and evaluation. The dynamic logic circuit is driven by a clock signal to ensure that it works properly [11]. To solve the monotonicity problem in dynamic circuits, an inverter is used in domino circuits at the output node [12]. The logic 0 of the first stage's output is transformed to a rising signal for the following stage, which helps cascaded circuits [13]. The schematic of the XAX module implemented using the domino logic is depicted in Fig. 3.

A weak keeper transistor is employed at the dynamic node to compensate for leakage current, which is nothing but domino keeper logic. When the pull-down network (PDN) is turned off, it holds the high voltage level at the dynamic node during the evaluation phase [14]. The schematic of the XAX module implemented using the domino logic is depicted in Fig. 4.
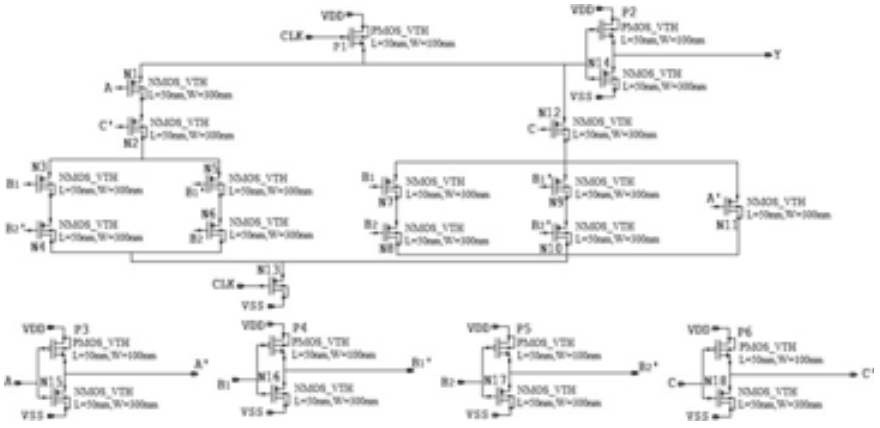
Fig. 3 Schematic of XAX module using domino logic
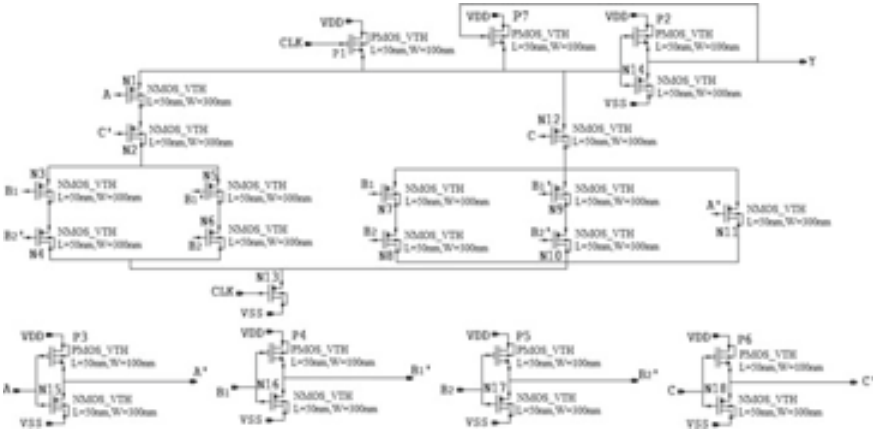
Fig. 4 Schematic of XAX module using domino keeper logic

NP domino logic is a race-free dynamic CMOS logic for implementing pipelined logic functions. An NP dynamic CMOS circuit is formed by cascading alternating *n*-type and *p*-type dynamic logic gates [15].

To avoid the clock-race condition, two clock signals, CLK given to PDN and CLKB given to pull-up network (PUN), are used, where a CLK stage is preceded and followed by a CLKB stage [12]. The basic implementation of XAX module using NP domino logic is depicted in Fig. 5.

PDN block's output is high during precharge, but PUN is low [16]. The logic block's output is assessed as a function of the inputs [17]. During the evaluation phase, inputs can make only one transition. The logic block's output is assessed as a function of the inputs [14]. In a pipelined system, this output will be transferred
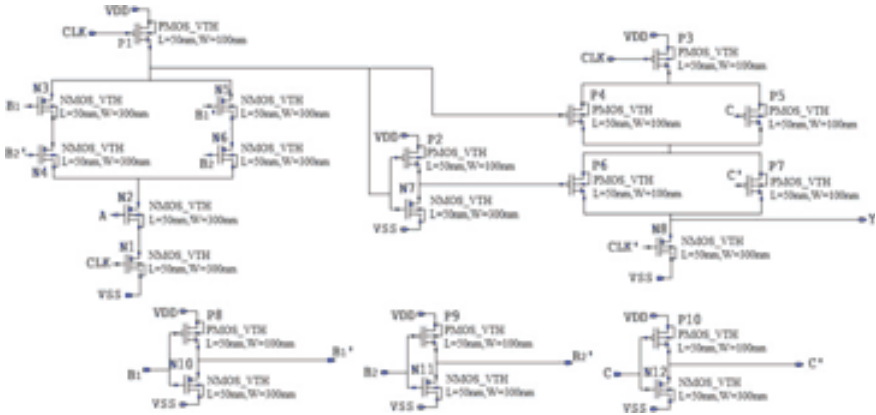
Fig. 5 Schematic of XAX module using NP domino logic

to the next pipelined section. During the evaluation phase, inputs can make only one transition.

In the NP domino circuit, both inverting and non-inverting outputs are possible. As there is a need to use an inverter to eliminate the monotonicity problem in domino circuits, the NP domino logic becomes handy as it handles this issue efficiently. The charge sharing issue [17] causes erroneous outputs which can be solved by supplying independent clocks to the precharge transistor and the evaluation transistor [18]. As a result, the NP domino logic family can be rejuvenated as a reasonable solution to conventional domino circuits [15].

5 Results and Discussions

The Mentor Graphics tool is used to simulate the multiplier design. The full custom design is used to implement the design using the 45 nm technology. The PMOS and NMOS transistors used in the schematics are derived from 'NCSU devices FreePDK45' library. Increasing the transistor sizes will help in reducing the delay. The length of both the PMOS and NMOS transistors is 50 nm. The width of PMOS transistor is 100 nm and that of NMOS is 300 nm for all the logic styles. The design parameters of 11-bit SIPO RNB multipliers are calculated and compared for different logic styles in Table 1.

The output waveforms of 11-bit SIPO RNB multiplier are shown in Fig. 5. The *X*-axis label in the output waveform is time (sec) and the *Y*-axis label is voltage (V). The CLK signal is used to synchronize the flip-flops in the circuit. The input is fed to the multiplier when the LOAD signal is high. The output flip-flop was given a RESET signal until the LOAD signal went low. Once the LOAD signal goes low,

Table 1 Comparison of design parameters of 11-bit SIPO RNB multiplier using different logic styles

Design parameters	Domino logic	Domino keeper logic	NP domino logic
Area (μm^2)	10,985	11,395	10,625
Power dissipation (nW)	84.6	84.2	84.9
Critical path delay (ps)	221.40	222.57	214.77
Multiplication delay (ps)	4870.8	4896.54	4724.94
Area-delay product (μm^2/GHz)	53,505	55,796	50,202
Power-delay product (nW/GHz)	412.06	412.28	401.14

no input combination is provided. As the RESET signal goes high, the outputs are generated.

The area of the multiplier is calculated by multiplying the length and width of each transistor with the number of transistors in the circuit. PDP is a figure of merit associated with the energy efficiency of a logic gate, whereas ADP is termed as the area of the circuit times the clock delay. The XOR-AND-XOR path is responsible for the critical path delay of the circuit. The time taken by the circuit to generate the output of multiplication is referred to as the multiplication delay.

From Table 1, it is clear that the area of domino logic is 3.6% less than the domino keeper as a keeper transistor is introduced.

But the NP domino logic is 6.75% less than domino keeper logic because the NP domino logic involves pipelining of n-type and p-type dynamic logic gates. As a result, its area will be less. As a keeper transistor is used in domino keeper logic, it dissipates less power than domino logic.

Since NP domino logic and domino logic, with the exception of pipelining, have similar features in the precharge and evaluate modes, their power dissipation is almost equivalent. From the table, it is evident that the domino keeper logic consumes 0.83% less power when compared to NP domino logic. The similar is the case with domino logic. Hence, domino keeper logic has the least power dissipation.

The XAX path is responsible for the critical path delay of the circuit. The size of the transistors in the circuit can be changed to control the delay of the dynamic domino circuits. The length of both the PMOS and NMOS transistors is 50 nm. The width of PMOS transistor is 100 nm and that of NMOS is 300 nm for all the logic styles. The CPD of NP domino logic is reduced by 2.9 and 3.5% when compared to domino and domino keeper logic styles. The critical path delay of the NP domino logic is the least of all the logic styles. The multiplication delay follows a similar decreasing pattern. The ADP of NP domino logic is the least and highest for domino keeper logic as the area in domino keeper logic is dominating the other logic styles. But as the power dissipation is almost same, the PDP is same for both domino and
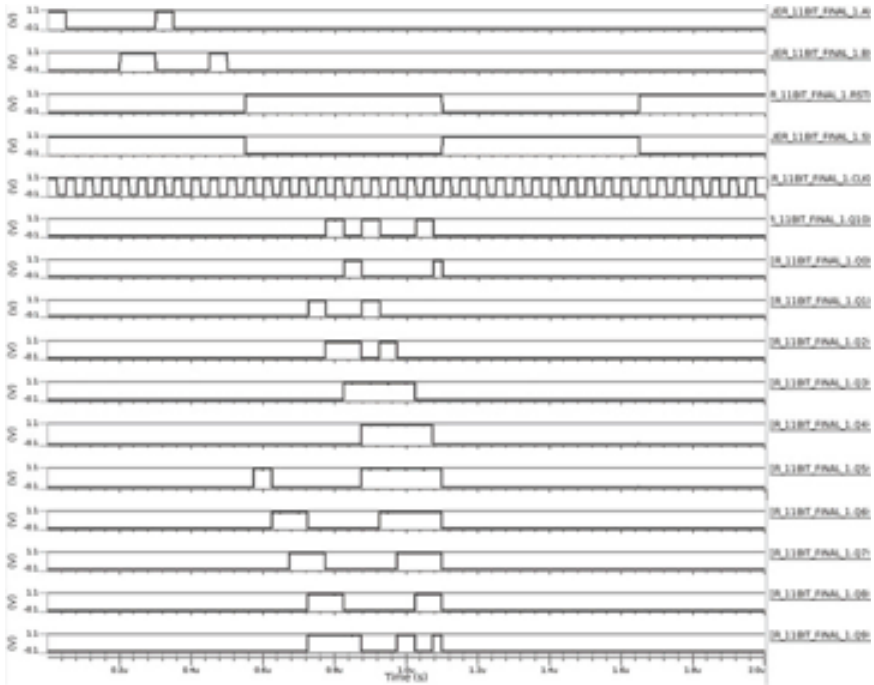
Fig. 6 Output waveforms of 11-bit SIPO RNB multiplier

domino keeper logic styles. As usual PDP is least for NP domino logic. These results prove that the NP domino logic is way better when compared to other logic styles.

As the RNB multipliers find its use in cryptographic applications, the field size of 233 is large enough to fall within the acceptable range for elliptic curve cryptography (ECC). Furthermore, it is one of the few fields that the National Institute of Standards and Technology for ECC applications. The 233-bit SIPO RNB multiplier can use NP domino logic making it suitable for cryptographic application.

6 Conclusions

The SIPO RNB multiplier has been implemented in this paper using the various domino logic styles in the Tanner EDA tool utilizing the 45 nm technology. The design parameters are examined, and the most appropriate logic styles for each design parameter are determined. In terms of power dissipation, domino keeper logic circuits have been proven to be efficient. The NP domino logic has the least critical path delay, area, ADP, and PDP. Hence, the NP domino logic is area-efficient and has high-speed compared to domino logic styles. This multiplier designed using NP domino logic can be further extended to GF (2^{233}) which was recommended

by the National Institute of Standards and Technology for its use in cryptographic applications.

References

1. Peterson WW, Weldon EJ (1972) Error-correcting codes. MIT Press, Cambridge, MA
2. Denning DER (1983) Cryptography and data security. Addison-Wesley, Reading, MA
3. Reed IS, Truong TK (1975) The use of finite fields to compute convolutions. IEEE Trans Inform Theor IT-21:208–213
4. Feng G-L (1989) A VLSI architecture for fast inversion in GF (2^m). IEEE Trans Comput 38(10):1383–1386
5. Azarderakhsh R, Reyhani-Masoleh A (2013) Low-complexity multiplier architectures for single and hybrid-double multiplications in Gaussian normal bases. IEEE Trans Comput 62(4):744–757
6. Mullin RC, Onyszchuk IM, Vanstone SA, Wilson RM (1989) Optimal normal bases in GF (pn). Opt Normal Bases GF(pn) 22(2):149–161
7. Namin AH, Leboeuf K, Muscedere R, Wu H, Ahmadi M (2010) High-speed hardware implementation of a serial-in parallel-out finite field multiplier using reordered normal basis. IET Circ Dev Syst 4(2):168–179
8. Namin PH, Roma C, Muscedere R, Ahmadi M (2018) Efficient VLSI implementation of a sequential finite field multiplier using reordered normal basis in domino logic. IEEE Trans Very Large Scale Integr (VLSI) Syst
9. Gao L, Sobelman GE (2000) Improved VLSI designs for multiplication and inversion in GF (2^m) over normal bases. In: Proceedings of 13th annual IEEE international ASIC/SOC conference, Sept 2000
10. Deshpande AV (2021) High speed VLSI implementation of a serial-in parallel-out finite field multiplier. Asian Res J Curr Sci 3(1):131–136
11. Ragini K, Satyam M, Jinaga BC (2010) Variable threshold MOSFET approach (through dynamic threshold MOSFET) for universal logic gates. Int J VLSI Des Commun Syst (VLSICS) AIRCC 1(1):34–44
12. Weste NHE, Harris DM (2000) CMOS VLSI design a circuits and system perspective, Asia. Pearson Education (Asia) Pvt. Ltd.
13. Anugraha Rose V, Durga Devi S, Avudaiammam R (2017) Design and performance analysis of 2:1 multiplexer using multiple logic families at 180 nm technology. In: 2nd IEEE international conference on recent trends in electronics information & communication technology (RTEICT), India, May 2017
14. Shinde J, Salankar S (2014) Multi-objective optimization for VLSI circuits. In: IEEE international conference on computational intelligence & communication networks, India, Nov 2014
15. Rabaey JM, Chandrakasan A, Nicolic B (2003) Digital integrated circuits: a design perspective, 2nd edn. Prentice-Hall, Upper Saddle River, NJ
16. Rai UK, Mehra R, Rasaily D (2016) CMOS design of low power high speed NP domino logic. IOSR J VLSI Signal Process (IOSR-JVSP) 6(1)
17. Ling N, Bayovmi MA (1988) A testable NORA CMOS serial-parallel multiplier. In: International symposium on circuits and systems (ISCAS), USA
18. Hang G, Zhou X, Yang Y, Zhang D (2014) NORA circuit design using neuron-MOS transistors. In: 10th international conference on natural computation, China

A Deep Learning Paradigm
for Classifying Personality

K. V. Raghavender, S. Ravi Kumar Raju, S. Alankruthi, M. Ashritha, G. Poojitha, and B. Avyaktha

Abstract Personality refers to a person's distinctive ways of thinking, feeling, and behaving in a variety of contexts. These traits are what consistently characterize a person's behavior. It influences how we come to decisions, work through issues, resolves disputes, and handle stress. The Myers-Briggs type indicator is currently the most popular psychological type assessment in use worldwide. This dataset was obtained via Kaggle. It enables us to positively deal with people's diversity by allowing us to anticipate certain personality traits in specific individuals. Many IT businesses are employing candidates with strong interpersonal communication abilities in today's society. To measure these, we developed a method to determine their personality, which can help in decision-making. Using different machine learning approaches like SVM, DT, and LR, several past research has attempted to categorize people into different personality types. LSTM, a deep learning model, is now being used to help classify people's personalities. In contrast to more conventional feedforward neural networks, LSTMs feature feedback connections. This property allows LSTMs to process entire data sequences without taking into account each point in the series individually, instead maintaining useful information about earlier data in the sequence to help in the processing of incoming data points to effectively classify the user texts. Finally, this supports the management, selection, and advancement of policies within organizations.

Keywords Long short-term memory (LSTM) · Myers-Briggs type indicator (MBTI)

K. V. Raghavender (✉) · S. Ravi Kumar Raju · S. Alankruthi · M. Ashritha · G. Poojitha
Department of Computer Science, G. Narayanamma Institute of Technology
and Science (For Women), Hyderabad, Telangana, India
e-mail: drkvraghavender@gmail.com

B. Avyaktha
Masters in Business Analytics, University of Texas at Dallas, Richardson, TX, USA

1   Introduction

Learning, remembering, thinking, perceiving, and emotions are just a few of the cognitive processes and mental states that are the subject of research in the broad discipline of cognitive science. Among the cognitive types mentioned above, personality, plays a critical role in dictating how people act in social contexts. Our personality has a significant impact on our lives; it has an impact on our lifestyle decisions, overall health, and a variety of other preferences. Numerous crucial real-world uses can be made of automatic personality trait recognition. For instance, when using sentiment analysis, a user should only be given goods and services that have received favorable reviews from other users who share their personality type.

Since the same notion might have distinct meanings to different sorts of people, personality detection can also be used to clarify word polarity in sentiment lexicons. Certain diagnoses in mental health diagnosis are correlated with particular personality qualities. Knowing personality traits in forensics aids in narrowing the suspect list. Personality qualities have an impact on a person's fitness for a certain position in human resources management. Long-term research has been ongoing in the field of computer-based personality detection and classification.

It is possible to analyze personalities using a variety of media, including text, images, audio, and video [1]. The topic of taxonomy of emotional perception from text on social networks has emerged in response to a variety of challenges facing academics in cognitive computation. Numerous studies have already been done in this area, and many more could be. There are numerous practical applications for cognitive-based text sentiment analysis (SA), making it more than just a theoretical field.

1.1   Problem Statement

Cognitive-based SA applications have gained popularity in recent years among online communities as a way to learn about people's attitudes and personality traits about certain topics, laws, and other things. However, because social media information is so varied, it is time-consuming to analyze text using current approaches to identify personality traits in such content. Therefore, the automatic classification of personality traits used in social media content extraction and analysis has become crucial. The areas of text-based SA [2], lexicon creation [3] cognition [4], visual-based SA [5], and aspect-based SA [6] have all seen a lot of research. But additional research is needed in the field of cognitively oriented social media, focusing on exploitation and classification of personal qualities.

1.2 Problem Definition

Models that only have a few personality traits are the foundation of current research on categorizing personality traits. Moreover, for effective classification of user's behavior on social media content, such algorithms require a sophisticated feature set. The issue of categorizing personality traits is formulated as follows in this project: Consider the following text reviews beginning with the letter T: With labels T_i, $T = t1, t2, t3,\dots$ pn for $I = 1$ to T). Ith text review is a psychological feature associated with eight personality types. We see the classification issue of personality trait identification, with the main goal of categorizing input text into the personality traits E (Extroversion)-I (Introversion), S (Sensing)-N (Intuition), F (Feeling)-T (Thinking), and P (Perception)-J (Judging). The study's goal is to build a robust and trustworthy personality trait identification model that can recognize and categorize text as one of the 16 types of personality trait combinations.

1.3 Contribution

The following are the main contributions to the research work: In order to extract word semantics at a deeper level and retrieve past contextual data from texts, we created the deep learning model—LSTM. For lengthy texts, the LSTM model effectively learns the long-range dependencies between word sequences.

The remaining part of the paper follows the same format. By describing current methodologies and systems, Sect. 2 illustrates the relevant research on classifying personality traits. Section 3 presents the suggested approach for categorizing personality traits. Section 4 of the proposal finishes with a presentation of the results together with any limitations and any future recommendations.

2 Related Work

The approaches suggested and used in other studies that we have previously looked at are detailed below.

One of the methods that is most frequently used for academics to study spoken and written text is supervised machine learning, commonly known as the corpus-based technique. There are a lot of benefits, like the ability to look at word usage, collocation, frequency, and concordance [7]. It tackled the issue of subject-independent emotion recognition by using a network-based fuzzy inference system. The suggested system can generalize well and pick up new information from training data. One of the main limitations of such strategies is that an annotated corpus is necessary for both the training and testing phases of the supervised or corpus-based technique (CBAclassifier) [8].

The chosen studies that concentrate on machine learning techniques are summarized here. In [9], a system for detecting personalities using unsupervised techniques and five personality traits (The Big Five) for extracting and analyzing user personality attributes from various social media networks such as friend feeds is presented. They developed a personality model by utilizing many language traits associated with personality. For a given text, the system computes personality scores and produces acceptable results. However, the inter-user interaction elements were not addressed by their system. When analyzing domain-dependent data, the unsupervised technique is simple to use.

The user's personality was inferred from written data using the Naive Bayes method [10]. The user creates a self-descriptive sentence that will be used to determine their personality before being matched with a dating partner on the Internet. It is spoken in Indonesian. The Four Temperaments are the preferred personality model. Four Sanguine, choleric, melancholic, and phlegmatic components are present.

In [11] emotion was recognized from speech. Emotion recognition is to choose key elements that convey a lot of emotional information about the voice signal. Gaussian mixture model makes use of parameters like MFCC, spectral centroid, spectral skewness, and spectral pitch chroma that are retrieved from the form of the speech signal to recognize emotion.

Text can also be used to predict emotion [12]. K-nearest neighbors were used to analyze papers for emotion. Indonesian language was employed in this experiment. The text document is a news piece from the Internet. Joy, rage, fear, sadness, contempt, and surprise are examples of basic emotions.

Previous studies on the Big Five personality traits from Facebook-based social media are found in [13] and [14]. The My Personality dataset was utilized as the corpus [15]. WEKA is used in these trials to analyze and forecast the outcome. Several built-in algorithms were used, with accuracies ranging from 52 to 61%.

3 Methodology

We proposed LSTM model for classifying personality characteristics from texts in social networking sites. By preserving information in one direction, the use of the LSTM model makes it easier to interpret the context. The suggested technique consists of three independent sections: (a) data acquisition; (b) pre-processing of data; and (c) deep learning model implementation. The subsequent sections provide a comprehensive description of the three modules.

3.1 Data Collection

For our research on classifying personality traits, we used the MBTI dataset which has been taken from the Kaggle [16]. There are 8600 rows of data in the MBTI dataset,

type	posts
INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||http://41.media.tumblr.con
ENTP	'I'm finding the lack of me in these posts very alarming.|||Sex can be boring if it's i
INTP	'Good one _____ https://www.youtube.com/watch?v=fHiGboIFFGw|||Of cours
INTJ	'Dear INTP, I enjoyed our conversation the other day. Esoteric gabbing about the

Fig. 1 MBTI dataset

each of which contains information about a person's personality type and a section of the previous 50 posts they have made. And the four personality type classification categories: E (Extroversion)-I (Introversion) [E-I], S (Sensing)-N (Intuition) [S-N], F (Feeling)-T (Thinking) [F-T], and P (Perception)-J (Judging) [P-J], each review in the data is tagged with a distinct class (see Fig. 1).

90% of the data is used as a training set to prepare the suggested model. Ten percent of the test data, separated from the training set, was employed. Once the model is fully trained, the train set is only utilized once. Finally, test data are used to evaluate the model. Data is sent to the pre-processing module after it has been acquired.

3.2 Data Pre-processing

Here firstly apply label encoding for handling categorical personality types. Then apply the to_categorical() method to transform it into a NumPy array or a matrix with binary values and it has the same number of columns as the number of unique personality types.

Then remove all non-alphabetic characters from input features using regex expressions and convert them into lowercase. The text is divided into tokens using the split function and then perform stemming operations and combine all the tokens to form a large text. Then apply one hot encoding to obtain a vector between index 0 to VOC-1 where VOC is vocabulary size. Then apply pad_sequences to ensure that all sequences in the list have the same length.

Then apply random oversampling on the entire dataset, i.e., select samples at random with replacement in order to over sample the minority class(es) so that it provides more measurement points that enable more samples to be averaged for greater precision.

3.3   Deep Learning Model

To classify personality qualities from the input text, a deep learning model is applied. The LSTM model is made up of three separate layers, which are the (i) Input Layer, (ii) Hidden Layer, and (iii) Output Layer.

Input Layer

Embedding Layer

With the Keras library, we had used the embedding layer to encode the words as real-valued vectors or numeric representations. The initial weights for the Embedding layer are random, and it will learn an embedding for each word in the training dataset. It consists of 3 arguments. Input_dim indicating vocabulary size, output_dim indicating dense embedding dimension, and input_length indicating length of input sequences.

Dropout Layer

One of the most widely used regularization methods to lessen overfitting in deep learning models is the dropout layer. When a model performs better on training data but worse on test data or unobserved data, it is said to have over fitted. Then we apply dropout layer which excludes input and recurrent connections to LSTM units probabilistically from weight and activation updates during network training. The dropout strategy randomly removes or drops some neurons from layers that are hidden or visible. And the dropout rate is between [0, 1].

Giving neurons a zero value, which deactivates some of the neurons, is the dropout layer's responsibility. This experiment demonstrates that the neural network model is regularized by this dropout strategy.

Hidden Layer

It consists of LSTM layer. The primary function of LSTM is that it uses previous knowledge to operate on current data. Cell states, the long-term memory of the LSTM units, is responsible for long-term storage of information without degradation. The three gates that make up the LSTM structure are utilized to place the weights between the memory cells and neural network.

The remember vector, often known as the initial forget gate, has a self-connection with value 1 and weight value 1, the memory cell will recall the data. If not, the memory cell will forget the information. The forget gate's equation is as follows:

$$\text{remember}_t = \text{sigmoid}\big(\text{weight}_r\big[\text{wgm}_{t-1},\, x_t\big] + \text{bias}_r\big) \tag{1}$$

The second input gate, also called as the "store vector", can write values into memory cells if it contains the value 1. The input gate's equation is as follows:

$$\text{save}_t = \text{sigmoid}\big(\text{weight}_s\big[\text{wgm}_{t-1},\, x_t\big] + \text{bias}_s\big) \tag{2}$$

The memory cell can be read by the network for values if the "focus vector", which is the third output gate, has a value of 1. The output gate's equation is as follows:

$$\text{focus}_t = \text{sigmoid}\big(\text{weight}_f\big[\text{wgm}_{t-1},\ x_t\big] + \text{bias}_f\big) \tag{3}$$

In Eqs. (1)–(3), remember$_t$, save$_t$, and focus$_t$ stand in for the forget, input, and output gates, respectively; weight$_r$, weight$_s$, and weight$_f$ stand in for weight matrices; x_t represents the present input; wgm$_{t-1}$ indicates the earlier concealed state, also referred to as "working memory"; and bias$_r$, bias$_s$, and bias$_f$ stand in for the forget, input, and output gates' bias values, respectively.

The cell state is then updated with a new candidate value. Using the following equation, the candidate values are computed.

$$\text{lgtm}'_t = \tanh\big(\text{weight}_l\big[\text{wgm}_{t-1},\ xt\big] + \text{bias}_l\big) \tag{4}$$

In Eq. (4), the lgtm$'_t$ denotes the candidate value and the hyperbolic tangent activation function is shown by tanh.

There are two outputs produced by each LSTM unit. The new concealed state comes in second, while the first is an updated or new cell state. Both outputs are now provided with the following formulae.

New Cell State

$$\text{lgtm}_t = \text{remember}_t * \text{lgm}_{t-1} + \big(\text{save}_t * \text{lgtm}'_t\big) \tag{5}$$

New Hidden State

$$\text{wgm}_t = \text{focus}_t * \tanh\big(\text{lgtm}_t\big) \tag{6}$$

In Eq. (5), lgtm$'_t$ denotes the updated cell state, lgtm$_{t-1}$ denotes the previous cell state, and wgm$_t$ denotes the new hidden state. These three gates guard against the problem of exploding and vanishing gradient for the LSTM units (Fig. 2).

Output Layer

For the feature classification, we chose the softmax activation function. The softmax function's objective is to categorize distinct classes using decimal probability, and the sum of all decimal probabilities should be 1. Using Eq. (7) [17], the softmax function normalizes each class output between 0 and 1. The probability is highest for the target class at last.

$$\text{softmax}\big(A_f\big) = \text{e}\big(A_f\big)(f = 0, 1, 2, \ldots, n)/\sum_{g=0}^{n}\text{e}\big(A_f\big) \tag{7}$$
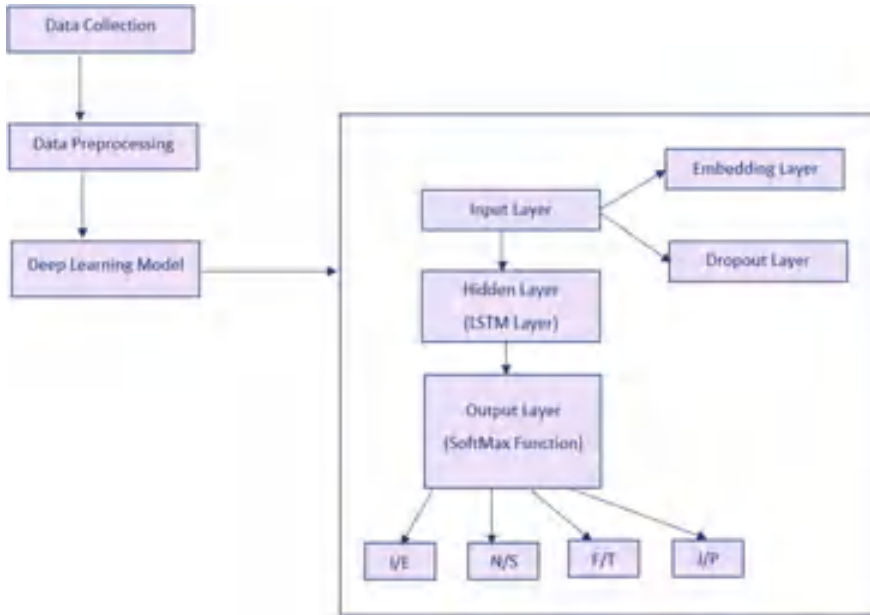
Fig. 2 Architecture of the proposed model

It is widely employed as the final activation function to normalize the output of a neural network to a probability distribution over the classifications of predicted output. Here we will compare probabilities for each trait and choose highest probability one.

Algorithm for Personality Classification Model

Step 1: Import the dataset as CSV file format.

Step 2: Sort posts and their labels into categories X and Y and apply all the pre-processing steps.

Step 3: Use random oversampling for balancing the class distribution of an imbalanced dataset.

$$ros = RandomOverSampler(random\_state = 42)$$

Step 4: LSTM Model Procedure.

Parameter Initialization
Voc_size=50000
embed_dim=100
seq_length=250
Build Deep Learning Model
Model=sequential ()
Embedding Layer

```
model.add(Embedding(voc_size,embed_dim,input_length = seq_length))
```
Dropout Layer
```
model.add(Dropout(0.25))
```
LSTM Layer
```
model.add(LSTM(100))
```
Softmax Layer
```
model.add(Dense(16,activation='softmax'))
```
Compilation Function
```
model.compile(loss='categorical_crossentropy',optimizer='adam',
metrics=['accuracy'])
```
Summary of the Model
```
Print(model.summary())
x_train,x_test,y_train,y_test=train_test_split(x_rus,  y_rus,  test_size=0.1,
random_state=42)
history=model.fit(x_train,            y_train,validation_data=(x_test,y_
test),epochs=8,batch_size=64)
score = model.evaluate(x_test, y_test, verbose=1)
print ("Test Score:", score [0])
print ("Test Accuracy:", score [1])
```

4 Results and Discussion

Evaluation and analysis are performed on the test findings that were predicted. MBTI Dataset is analyzed as follows (Fig. 3).
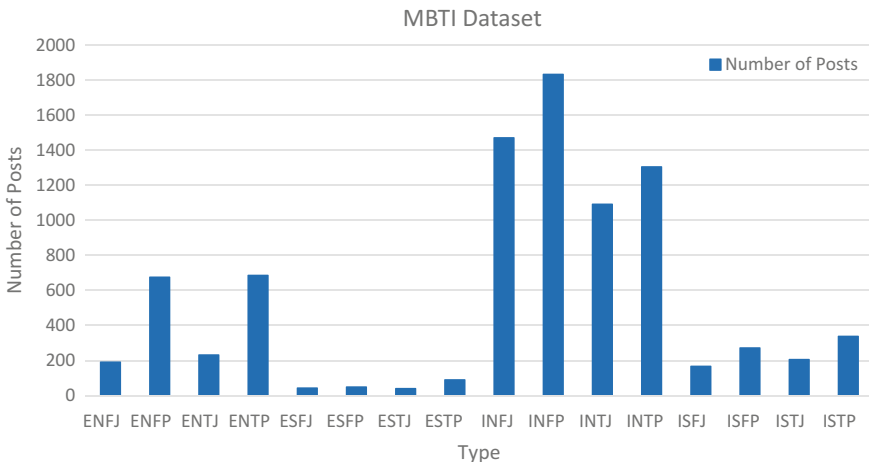
Fig. 3 Analysis of MBTI Dataset

It consists of 8675 rows and each row is classified into one of the 16 personality traits. And then LSTM model is used for evaluation and test score and accuracy is noted as follows:

Test Score: 0.4878.

Test Accuracy: 0.8578.

And also, graphs are plotted with respect to loss and accuracy with training and validation data (Figs. 4 and 5).
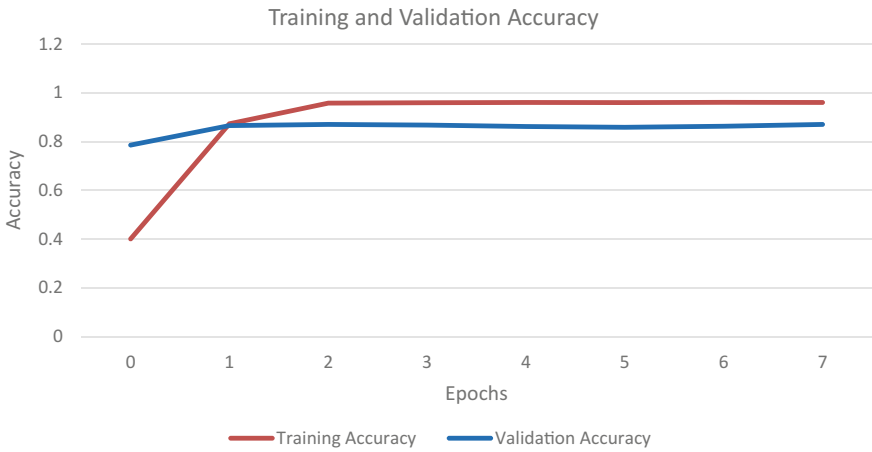
Fig. 4 Analysis of model accuracy

Fig. 5 Analysis of model loss

Input Text

French Text—"hello, re-joins moi sur skype", "bonjour comment ça va ?", "tu me donnes de l'argent"

English Text—"Hello join me on skype, hi how are you, you give me money".

Output: [[0.0058098 0.02121465 0.00825141 0.09037514 0.00110254 0.00261576

0.00137799 0.00274994 0.19313586 0.11524879 0.23059185 0.29645756 0.00463903 0.00693894 0.00856555 0.0109252]]

Your Personality is: INFP

So here when we input the text it gives you an array of probabilities with 16 personality traits. And based on the probabilities, it is classified into INFP for above input text which indicates introvert, intuition, feeling, and perception which is assigned to highest probability one.

And also, at last we plotted confusion matrix with 16 personality traits to get a better idea of model's performance. It indicates there are more correct predictions than incorrect predictions for this model's classification (Fig. 6).

Comparison With Other Models

In [18] they used traditional techniques to classify personality. Supervised learning methodology namely Naïve Bayes, SVM, LR, and Random Forest, are used for estimation. By using Naïve Bayes algorithm, they got an accuracy of 55.89%. And they got an accuracy of 65.44% by using SVM algorithm. Using logistic regression, they got 66.59% accuracy. They achieved lower accuracy because of traditional machine learning techniques. Our deep learning techniques has increased accuracy to a greater extent.

Fig. 6 Performance of model using confusion matrix

5 Conclusion and Future Scope

In this instance, we used the LSTM model to classify personality traits from textual content. Textual information is categorized according to various personality qualities including E-I, S-N, F-T, P-J. The LSTM model keeps the prior context information, which aids in utilizing important context details at the start of a sentence. It has the advantage of taking sequential information by looking at previous information. The findings demonstrate that the suggested LSTM model for categorization of character traits gave better outcomes in terms of increased accuracy (85.78%) and score (48.78%). Thus, the suggested LSTM model for classifying personality trait functions as best practices for policy, services and product choice, administration, and improvement.

6 Future Directions

The following are some potential future directions for the work: (i) By doing studies with a larger spectrum of personality qualities using non-textual elements like photos and videos, the research can be strengthened. (ii) It is possible to use additional datasets for personality trait categorization tasks besides the MBTI dataset. (iii) In order to improve the proposed work's overall accuracy, we can work on researching other parameters like adding more parameters in the algorithm and incorporating other base models like combining two or more base models to improve the accuracy.

References

1. Ahmad H, Asghar MZ, Khan AS, Habib A (2020) A systematic literature review of personality trait classification from textual content. Open Comput Sci 10(1):175–193
2. Osama M, El-Beltagy SR (2019) A transfer learning approach for emotion intensity prediction in microblog text. In: Proceedings of International conference on advance intelligent system informatics. Springer, Cham, Switzerland, pp 512–522
3. Asghar MZ, Sattar A, Khan A, Ali A, Masud Kundi F, Ahmad S (2019) Creating sentiment lexicon for sentiment analysis in Urdu: the case of a resource-poor language. Expert Syst 36(3):e12397
4. Khattak A, Habib A, Asghar MZ, Subhan F, Habib IAR (2021) Applying deep neural networks for user intention identification. Soft Comput 25(3):2191–2220
5. Ilimini K, Fernando TGI (2016) Persons' personality traits recognition using machine learning algorithms and image processing techniques. Adv Comput Sci Int J 5(1):40–44
6. Asghar MZ, Khan A, Zahra SR, Ahmad S, Kundi FM (2019) Aspect-based opinion mining framework using heuristic patterns. Cluster Comput 22(S3):7181–7199
7. Subramanian K, Radhakrishnan VB, Ramasamy S (2014) Database independent human emotion recognition with meta-cognitive neuro-fuzzy inference system. In: Proceedings of 9th International conference on intelligent sensors network and information processing (ISSNIP), pp 1–6

8. Shivakumar G, Vijaya PA (2011) Facial expression based human emotion recognition with live computer response. Int J Comput Sci Inf Technol 4:81–84
9. Celli F (2012) Unsupervised personality recognition for social network sites. In: Proceedings of 6th International conference on digital society, pp 59–62
10. Lestari NMA, Putra IKGD, Cahyawan AAKA (2013) Personality types classification for Indonesian text in partners searching website using Naive Bayes methods. Int J Softw Inform Issues 173
11. Navya Sri M, Ramakrishna Murty M et al (2017) Robust features for emotion recognition from speech by using Gaussian mixture model classification. In: International conference and published proceeding in SIST series. Springer, vol 2, pp 437–444
12. Arifin, Purnaha KE (2012) Classification of emotions in Indonesia text using K-NN method. Int J Inf Electron Eng 2(6)
13. Firoj A, Speanov EA, Riccardi G (2013) Personality traits recognition on social network—Facebook. In: ICWSM workshop on personality classification
14. Iacobelli F, Cullota A (2013) Too neurotic, not too friendly: structured personality classification on textual data. In: ICWSM workshop on personality classification
15. Celli F, Pianesi F, Stillwell D, Kosinski M (2013) Workshop on computational personality recognition (shared task). In: Proceedings of WCPR13, in conjunction with ICWSM-13, p 201
16. MBTI Dataset from Kaggle. https://www.kaggle.com/datasets/datasnaek/mbti-type
17. Gjurkovi M, Karan M, Vukojevi I, Boánjak M, Snajder J (2021) PANDORA talks: personality and demographics on Reddit. In: Proceedings of 9th International workshop natural language processing and social media, pp 138–152. https://doi.org/10.18653/v1/2021.socialnlp-1.12
18. Chaudhary S, Sing R, Hasan ST, Kaur I (2018) A comparative study of different classifiers for Myers-Brigg personality prediction model. IRJET 5:1410–1413

Art Generation Using Speech Emotions

Rishabh Patil, Bhagavatiraj Yadav, Omkar Narvekar, and Sanober Shaikh

Abstract Translation of speech to image directly without text is an interesting and useful topic due to the potential application in computer-aided design, human-to-computer interaction, creation of an art form, etc. So, we have focused on developing a novel deep learning and GANs-based model which will take speech as an input from the user, analyze the emotions associated with it and accordingly generate the artwork which has been demanded by the user which will in turn provide a personalized experience. Concept of convolutional VQGAN is used to explore codebook consisting of context-rich visual parts, and the entire composition is modeled with autoregressive transformer architecture. Concept of CLIP—Contrastive Language Image-Pre-Training—also uses transformers a model which is trained to find which particular caption from a collection of captions will best fit with the given image used in our project. The input speech is classified into eight different emotions using MLP classifier trained of RAVDESS emotional speech audio dataset, and this acts as a base filter for the VQGAN model. Text converted from speech plays an important role in producing the final output image using CLIP model. VQGAN + CLIP model together utilizes both emotions and text to generate more personalized artwork.

Keywords VQGAN · CLIP · Art generation · Speech emotions

1 Introduction

Deep learning basically uses supervised, unsupervised as well as semi-supervised learning models for the training purpose. It is quite evident that within a span of few years, there is a high possibility that deep learning tools, libraries and the languages can become the conventional component of each software development tool kit for

R. Patil (✉) · B. Yadav · O. Narvekar · S. Shaikh
Information Technology Department, Thadomal Shahani Engineering College, Mumbai, India
e-mail: rishabhpatil8108@gmail.com

S. Shaikh
e-mail: sanober.shaikh@thadomal.org

enhanced user experience which will give rise for easy building, configuration, and designing of models.

With the help of audio and video as a modality human can effortlessly identify other humans and sense their emotional state. When it comes to human perception, there is no doubt that vision is the advanced of all senses and hence images and visuals play a pivotal role. Coming to speech, infants when communicate with their parents they do not have knowledge about the linguistic details of that particular language but they are able to learn the co-relation between speech words and visual textures effortlessly. This shows us that how important role the comprehension of speech and actions plays in our life.

There are large numbers of applications of image generation in diverse spectrum of human activities; from the field of biomedicine to space technology, it has a vast scope and many applications. Image generation has abundance of scope which gives option to the researchers to go with the area of their interest, and it has also become a means for earning according to our interest. Lots of research findings are published, many communities are established, and lot of mentor support and guidance is available. Because of the abundance of technical gadgets and domains, digital image generation has become the most versatile method, but also economical. The best example is an NFT and is a blockchain-based tokenization of a collectible item which is perfect if you create limited edition digital artwork which has become an interesting method for earning among tech enthusiasts in today's world.

Today, translating data between different modalities like text to speech and speech to image is a cutting-edge area but more research is done only on text-to-image generation which does not provide a personalized experience. It is a difficult task to comprehend human emotions from an analytical perspective. It is easier to guess the emotional state of a person from their facial expression at the same time speech can also be used as another modality. Even for human beings, judging the emotions of the speaker from the speech at times might become tedious but at the same time multiple emotions could be well expressed through speech. So, we have focused on building a model which will take input as speech also will analyze the emotion and accordingly display the image asked for by incorporating emotions of the person recognized through the speech input, which will give a much more personalized experience for the users which is not possible by merely taking text input through a text prompt.

Almost half of the 700 languages in the world do not possess a written form, and it is not possible for such languages to benefit from the current text-to-image generation methodology because there is no clue about those words and sentence formation might become stressful. So, such languages require a technology which will overcome the constraint of this text and will help to map speech to images directly without the need of text format. We were fascinated by the image generation process and how it has been applied in the NFTs. Creation of image will immensely benefit the digital art producers, but to make the art livelier we will be incorporating emotions with it. For handling this we explored the possibility of using generative adversarial networks (GANs) for automatic generation on the art.

2 Literature Review

In [1], Lalitha and Madhavan have used pitch and prosody features and were successful in recognizing seven emotions. For the purpose of classification of emotions, we used SVM classifier. Idea was implemented using Berlin emotional database. 81% recognition rate was recorded. Suggested effort uses a condensed feature set with a decent precision rate for the seven emotions.

The paper proposed by Jaswinder Singh and Rajdeep Banerjee in [2] discusses the philosophy behind perceptron learning, how it operates, applications, in-depth principles, and a quick overview of multilayer perceptrons. The paper says that a perceptron has been the most fundamental artificial neural network models and was the one which began the on field of artificial nets with built-in learning and classification capabilities.

In [3], the article proposed by Tan Yuan, Xiaofeng Chen, and Sheng Wang mainly focuses on the favorable image generators accessible, VQGAN + CLIP, and how it can be combined with NLP and how it can be applied to spawn fresh clipart using a solo prompt with NLP and gradient. A discrete codebook has been used which provides interface among these designs and a patch-based discriminator and empowers strong density while ensuring the retention of high perceptual quality.

In [4], Zihan Chen and Lianghong Chen have made use of deep learning technology based on generative adversarial network to assist computer produce artworks from text. They have made use of the idea of loading the generation alongside the system, tally some classification information and by reconstructing data, they produced a better picture. An annotated Chinese image dataset was built first and then they continued to advance the ideal to form an end-to-end text-to-text image generation model.

In this survey [5], Licheng Jiao and Jin Jhao have concise a new group of deep learning techniques used in image processing. The article includes a section devoted to discussion of unresolved issues, difficulties, and potential developments. Since there is a diverse set of fields available in this domain, a comprehensive framework for comprehensive understanding has been provided toward critical aspects of this field. This survey tries to arouse researchers to catch a glimpse of various deep learning techniques in the ground of image processing and leverage these claims in their research tasks.

Training a text-to-image generator model involves gathering gigantic volumes of text-to-image records, which is not very practical to gather. In this article [6], Zihao Wang and Wei Liu have presented CLIPGEN, a self-supervised method for universal text-to-image generation that extracts language image priors using a trained CLIP model. In this approach, only a set of unlabeled images are required in the general domain to train the generator. Further, the image is converted into a series of distinct tokens in VQGAN codebook.

In [7], a speech-based technique for classifying emotions is presented by Tin Lay Nwe, Foo Say Wei. Around six basic emotions like happiness, surprise, and fear are probed. The presented recognizer uses a new feature vector based on mel frequency

short temporal voice power coefficients and a hidden Markov model as its foundation. The databases that were used contain about 90 emotional statements made by each of the two speakers. Numerous ungrouped and grouped emotions classifications were accompanied with an average accuracy of 72.22% and 60% in ungrouped and 94.4%, and 70% in grouped were achieved for two speakers, respectively.

In [8], Patrick Esser and Robin Rombach established how CNNs can adapt and make high-resolution images by combining the articulateness of transformers with their value in inductive preference. This demonstrates how to utilize CNNs to discover a context-rich vocabulary of picture elements and then use transformers to properly version their composition inside high-definition images. In particular, they gift the primary consequences on semantically steered fusion of megapixel pics through transformers and reach the artwork among autoregressive fashions on class-conditional ImageNet.

In [9], Alec Radford and Jong Wook Kim summarized the scalability of CLIP by educating a chain of eight models spanning nearly two orders of importance of computing and having a look at the overall transfer performance is an easily predictable characteristic. They find that during pre-training, CLIP, like the GPT family, learns to do a broad range of tasks, including OCR, the idea of geo-localization, action recognition, and many others. Benchmarking the zero-shot transfer results of CLIP on over 30 current datasets and discover it may be aggressive with earlier task unique supervised models.

In [10], Youssef Serrestou and Leila Kerkeni had compared various techniques of emotions recognition and had provided with the most accurate solution on arrangement of these techniques. The performance of the recurrent neural network classifier was compared to multivariate linear regression (MLR) and support vector machine for the seven emotions that were contained in the Berlin and Spanish databases (SVM). The results of the experiment performed on databases have the highest accuracy rate, i.e., 90.05% using RNN on Spanish database and 82.41% using MLR on Berlin database.

3   Methodology

This section of the paper explores various types of methods and algorithms used in the speech emotions to image generator model. When audio speech is input to the model, it converts it into text and emotions. The conversion of speech to text is done with the help of python "Speech Recognition" library. The input speech is also classified into one of the eight different emotions using the MLP classifier trained on "RAVDESS emotional speech audio" dataset. This emotion class is then used to set up a base filter/image for the VQGAN model and thus affects the entire structure and colors of the final image generated. The text generated from speech input plays a pivotal role in producing the final output image using the CLIP model. The VQGAN + CLIP model utilizes both emotions and text to generate more personalized artwork. The RAVDESS ("The Ryerson Audio-Visual Database of Emotional Speech and Song")
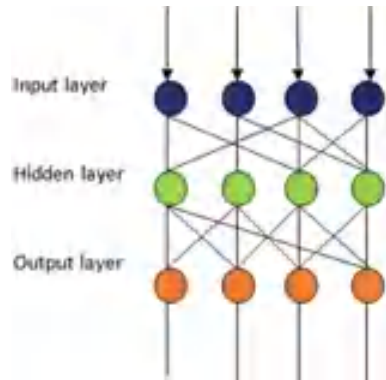
emotional speech audio dataset consists of different audio files from 24 actors which represent eight types of emotions in the form of ".wav" files. The eight emotions include calm, happy, sad, angry, fearful, disgust, surprised, shame.

From these eight emotions' types, the most observed emotions are "calm," "happy," "fearful," and "disgust." Features from these audio files were extracted using the librosa python library. These features were then fed into the MLP classifier from sklearn along with their corresponding emotion types with a test–train split of 25–75%, respectively. Multilayer perceptron classifier is a supervised machine learning algorithm also known as MLP classifier. Unlike other classifiers, it relies its classification task on an artificial neural network. Artificial neural networks (ANN) are widely used as they provide more accurate results for different machine learning operations. ANN can be considered as an attempt to simulate a human neural system. Figure 1 consists of multiple layers in an ANN which consists of input, output, and hidden layers between them. The input layer receives various information and passes it through all hidden layers. Eventually generating the expected output value based on training data from the output layer.

Each of these neural network layers is built with multiple perceptrons. As shown in Fig. 2, a perceptron is fed "n" number of features as input ($x_1, x_2, x_3,\ldots, x_n$). Each of these features is associated with weights ($w_1, w_2, w_3,\ldots, w_n$). These features must be in the numeric form. Then a function u calculates the weighted sum of all these given features along with their weights as shown in Eq. 1. Equation 1 shows weighted sum of input features.

$$u(x) = \sum_{i=1}^{n} w_i x_i.$$ (1)

As shown in Eq. 2, the obtained values are passed through an activation function f which provide final output y of the perceptron, where θ is a threshold parameter. Equation 2 shows activation function.
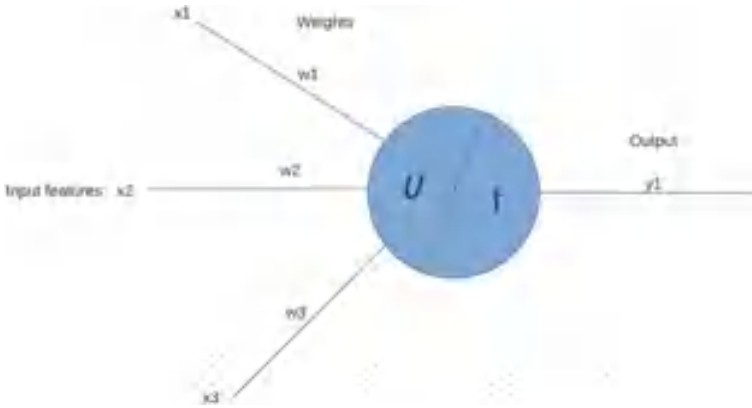
Fig. 2 Single perceptron

$$y = f(u(x)) = \begin{cases} 1, & \text{if } u(x) > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

Considering Eqs. 1 and 2, the final equation turns out to be (shown in Eq. 3).

$$W_1 X_1 + W_2 X_2 + \cdots + W_n X_n - \theta = 0 \tag{3}$$

Equation 3 is the equation of a hyperplane. If any input point is above this hyperplane the corresponding output of the perceptron will be 1, else the output of the perceptron will be 0. The neural network is improved by changing each of these weights and biases by comparing the perceptron output with the expected output.

Sklearn MLP classifier is based on such a neural network which helps to classify by just importing a few libraries. MLP classifier trains itself iteratively using backpropagation considering the losses and the model parameters. From a given set of attributes and target values, it can learn a nonlinear function approximator for classification as well as regression. MLP works as a universal approximator since it can approximate any continuous function (Fig. 3).

The MLP classifier trained with the RAVDESS dataset gave an accuracy of 76.56% on the test dataset which is sufficient for classifying majority of audio samples into the eight emotion types. These emotion types have a base image corresponding to each. These base images depict their corresponding emotion types by their textures and colors. For example as shown in Fig. 4, the base image of emotion type fear has a dark color and there is not much contrast within the image portraying fearfulness.

Similarly, each of the eight emotions is assigned base images with colors and textures depicting their corresponding emotions. All base images of the eight emotions are shown in Fig. 5.

For converting speech to text, python's speech recognition library is used. This library can recognize speech input from the microphone, transcribe an audio file, and process the audio signals to convert them into text format, making it suitable
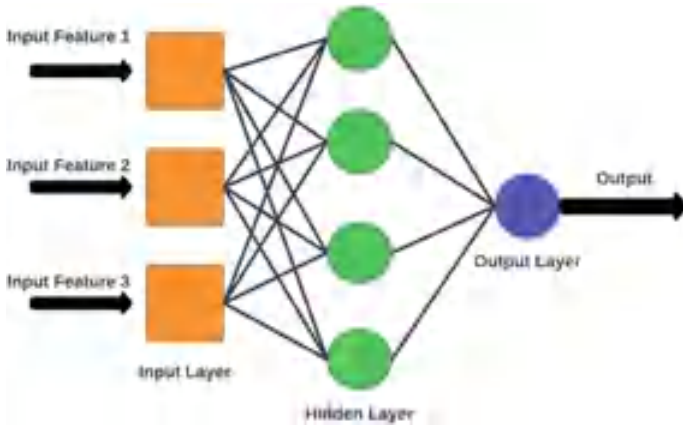
Fig. 3 One hidden layer MLP
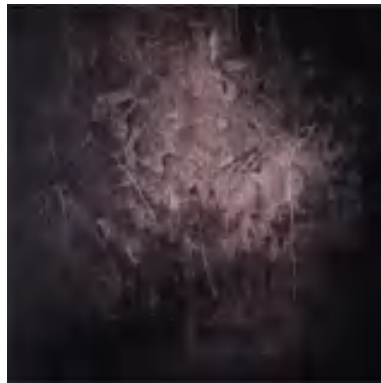
Fig. 4 Base image for emotion type—"fear"

for our model's use case. Another alternative for this library can be the Google Speech-To-Text API. By using MLP classifier and speech recognition library, we obtained the base image and text format conversion, respectively, of a given input audio file. These outputs are then passed to the VQGAN + CLIP model. VQGAN is a type of generative adversarial network (GAN) which involves transformers. Here the models are trained with only image data without the involvement of any text data. A GAN is a machine learning model which consists of two neural networks. These two neural networks compete with each other to improve the accuracy of the model. These models are unsupervised and can be used for the generation of new data based on input data. The two neural networks that compete with each other are called generator and discriminator. A generator is a convolutional neural network which creates outputs which are improved by using a feedback loop. Initially, the generator manufactures noise image, and with each iteration of the feedback
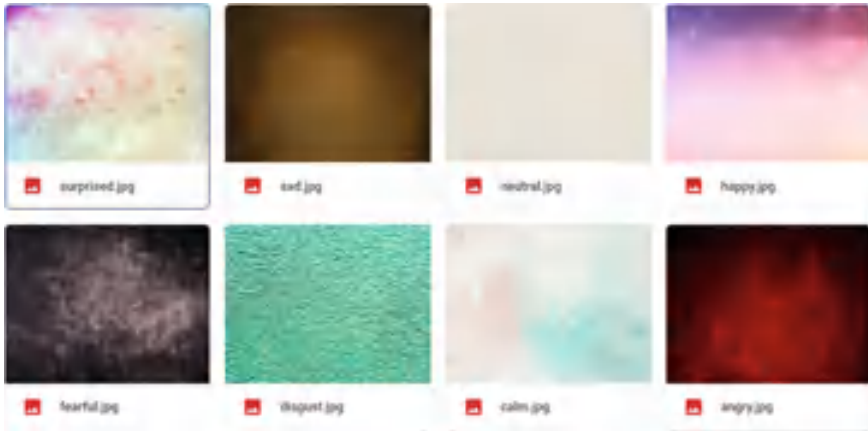
Fig. 5 Base images of eight emotions

loop, the noise image is converted to the expected output. A discriminator finds the difference between the generated image and the images from the dataset and forms loss functions. These loss functions are used to improve the images generated from the generator. This iterative process is shown in Fig. 6. On the basis of previously seen data, VQGAN can be used to generate images. The input image dataset is vector quantized in which clusters of different image vectors are encoded, and the cluster center is called a "codeword." All the image data from the dataset are stored in the form of a dictionary of codewords. This dictionary is called a codebook and is considered as a representation of the image data. This improves the efficiency and allows quicker processing of a GAN architecture for dealing with image data. Figure 7 shows the formation of the codebook and how it is used in a GAN model.
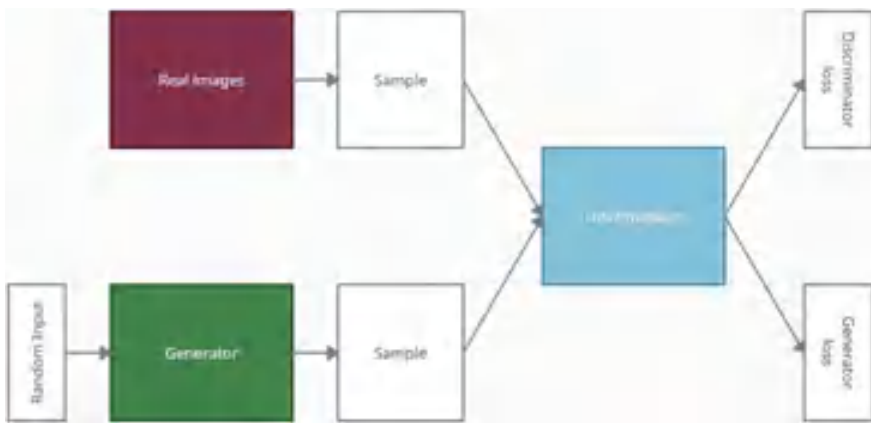
Fig. 6 GAN architecture
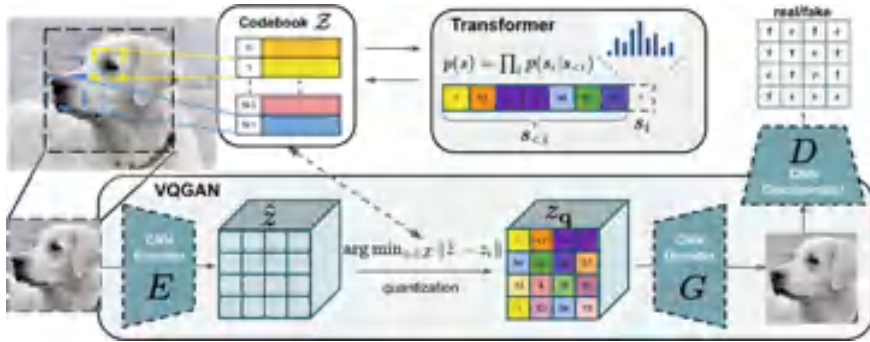
Fig. 7 VQGAN architecture [3]

CLIP is a model proposed by OpenAI that associates images from the dataset with captions. It is based on zero-shot learning, which enables it to provide improved outputs even on unknown data. The main goal of CLIP is to classify new unknown image data without any explicit labels. CLIP considered as a bridge between CV and NLP is trained on 400,000,000 (image, text) pairs. CLIP architecture is shown in Fig. 8.

VQGAN + CLIP when put together generates novel images based on the input text. First the VQGAN is supposed to generate a noise are iterate over it further, but in our model the base images shown in Fig. 6 are used as the initial images corresponding to their emotions. This helps for the VQGAN to adapt textures and colors corresponding to the speech emotion detected, and thus, the final output image also consists of such colors and textures giving a more personalized experience in the process of artwork generation. The base images are firstly vector quantized, and they are then encoded in a codebook. Further, this codebook is sent to the transformer which creates the artwork. The generated image is then used to assess accuracy. This process is repeated until we obtain a clear discernible artwork. Generally, 300 to 400 iterations are sufficient to create a discernible artwork. The iterative process of VQGAN + CLIP model is shown in Fig. 9.
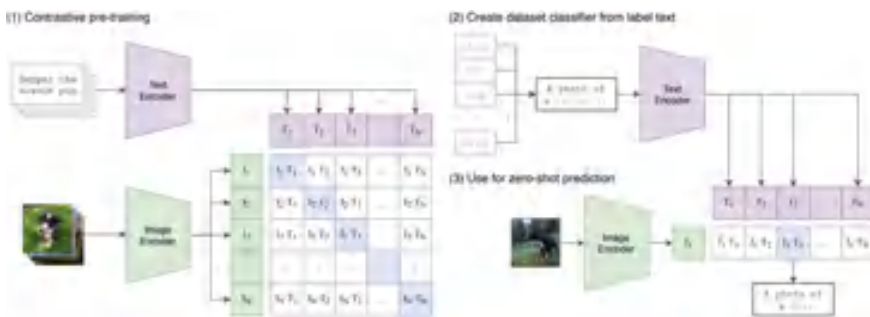
Fig. 8 CLIP architecture [3]
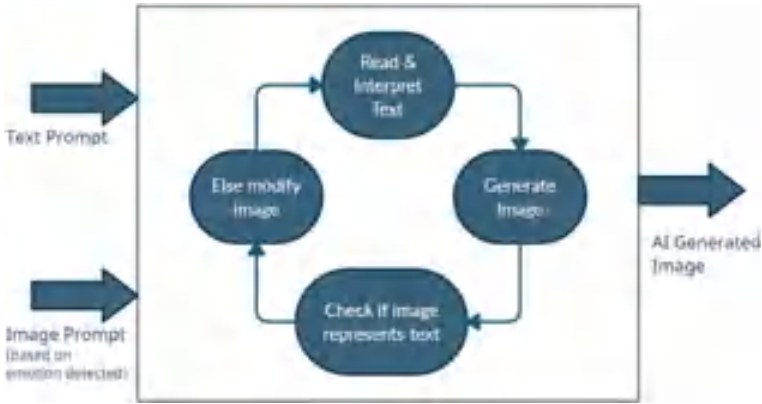
Fig. 9 VQGAN + CLIP architecture
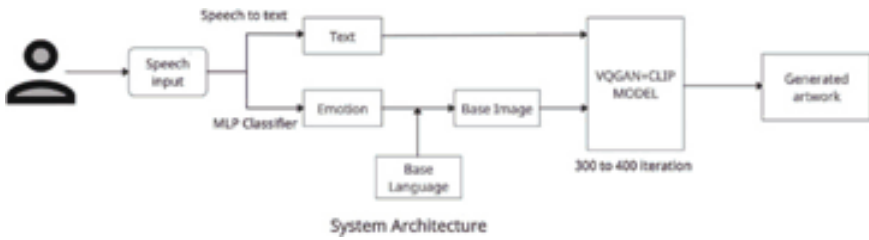
Fig. 10 System architecture

In the current state of the art, there is a wide range of models for generating artworks or images based on input text prompt. To achieve a more personalized artwork, there was a need to incorporate emotions of the person who wants to generate an artwork. In our research work, a user only needs to give speech input to generate a personalized artwork. This speech is then processed ahead, and final artwork output is generated based on the content the human speaks and with what emotions he/she speaks. This process is shown in Fig. 10.

4 Results

The model is tested on several speech inputs out of which two test cases are given below. In the first test case, a speech saying "Dark rainbow city" in a fearful manner was fed. The text detected from this speech prompt was "dark rainbow city," and the emotion detected was fearful. Thus, the base image was set to "fearful.jpg" from Fig. 5. This base image was iterated over for 245 iterations, and the final output that the model gave is shown in Fig. 12 (Fig. 11).

Fig. 11 Output after 15 iterations

Fig. 12 Final output after 245 iterations for speech input—"Dark rainbow city" with emotion fearful

As we can see in Fig. 12, the textures and colors of the final artwork are related to the term fearful, and according to the text prompt, the content of the image depicts different characteristics of a city having a rainbow at night time. This output image shows how novel images can be created using the VQGAN + CLIP model, since there are no chances of seeing a rainbow at night time in the dataset. Inputs and outputs of the given test case can be viewed here: https://drive.google.com/drive/folders/1OxBDfKDRp5lw6KSDCMLv6iSKIKCIT3wf?usp=sharing.

Fig. 13 Final output after 350 iterations for speech input—"sunrise" said with calm emotion

In the first test case, a speech saying "Sunrise" in a calm manner was fed. The text detected from this speech prompt was "sunrise," and the emotion detected was calm. Thus, the base image was set to "calm.jpg" from Fig. 5. This base image was iterated over for 350 iterations, and the final output that the model gave is shown in Fig. 13.

As we can see in Fig. 13, due to the emotion detected for the speech input as calm the colors and textures in the final output image are calm and soothing and content of the image is determined by the text prompt detected which is sunrise. Inputs and outputs of the given test case can be viewed here: https://drive.google.com/drive/folders/1M9a3K4GjuRz1zYqvAS798SZYlXmUUJ6k?usp=sharing.

5   Conclusion

In this paper, we have proposed an art generation model using VQGAN and CLIP. VQGAN which is used to generate novel images based on previously seen data. The input image dataset is vector quantized in which clusters of different image vectors are encoded, and all the image data from the dataset are stored in the form of dictionary of "codeword." This dictionary is called a codebook and is considered to be a representation of the image data which improves the efficiency and allows quicker processing of GAN architecture for dealing with image data. CLIP can be considered as a bridge between computer vision and natural language processing. In our research work, a user only needs to give speech input to generate a personalized artwork. This speech is then processed ahead and final artwork output is generated

based on the content the human speaks and with what emotions he/she speaks. In the implemented algorithm, the level of accuracy depends on the emotions which are associated with the utterances of the speaker. Depending on the clarity of the utterances and the emotions expressed, higher level of accuracy can be achieved for classification. Analysis of the relationship between image processing applications and deep learning will help us to improve our clarity in the deep learning domain, and we will also try to improve our knowledge and work more on the existing project to improve its accuracy.

References

1. Lalitha S, Madhavan A, Bhushan B, Saketh S (2014) Speech emotion recognition. In: 2014 International conference on advances in electronics computers and communications
2. Singh J, Banerjee R (2019) A study on single and multi-layer perceptron neural network. In: 2019 3rd International conference on computing methodologies and communication (ICCMC)
3. Yuan T, Chen X, Wang S (2022) Gorgeous pixel artwork generation with VQGAN-CLIP
4. Chen Z, Chen L, Zhao Z, Wang Y (2020) AI illustrator: art illustration generation based on generative adversial network. In: 2020 IEEE 5th International conference on image, vision and computing (ICIVC)
5. Jiao L, Zhao J (2019) A survey on the new generation of deep learning in image processing. IEEE Access 7
6. Wang Z, Liu W, He Q, Wu X, Yi Z. CLIP-GEN: language-free training of a text-to-image generator with CLIP
7. New TL, Wei FS, Tin LN, Foo SW, De Silva LC (2001) Speech based emotion classification. In: Proceedings of IEEE region 10 International conference on TENCON, vol 1, pp 297–301
8. Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 12873–12883 (2021)
9. PMLR (2021) Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International conference on machine learning, vol 139, pp 8748–8763
10. Speech emotion recognition: methods and cases study. Conference: 10th International conference on agents and artificial intelligence

Image Forgery Detection System Using Convolution Neural Networks

G. Sreenivasulu, B. Sujatha, K. Venu Madhav, Selvaraj Rajalakshmi, N. Thulasi Chitra, and B. Lingaswamy

Abstract In today's digital era, the power of images to convey our emotions and messages has reached new heights. Detecting close-to-identical images requires matching altered images to their original counterparts, a crucial task in recognizing manipulated visuals. Significant efforts have been devoted to developing pictorial applications that rely on efficient image similarity measurements and labeling Digital images can be easily edited and manipulated due to the advanced capabilities of image processing software. Consequently, this presents the challenge of accurately matching slightly altered images with their authentic sources, a task commonly known as "close-to-copy image discovery." This research paper delves into various image-matching algorithms that have been studied for this purpose. One prevalent image manipulation technique is image recoloring, which can alter image tones or subjects, making it difficult for the human eye to detect the changes. However, there is a lack of specialized methods designed to detect such manipulations. In response to this gap, our paper proposes an end-to-end system designed to detect recolored images among genuine ones. Our approach involves taking the original image and two auxiliary inputs based on brightness consistency and inter-channel correlation. The system then outputs the likelihood of the image being recolored. The core of our proposed method leverages a convolutional neural network (CNN) architecture,

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

B. Sujatha
Department of Computer Science Engineering, University College of Engineering, Osmania University, Hyderabad, India

K. Venu Madhav · S. Rajalakshmi
Department of Computer Science and Information Science, School of Sciences, BIUST, Palapye, Botswana, South Africa

N. Thulasi Chitra
Department of Computer Science Engineering, MLRIT, Secunderabad, India

B. Lingaswamy
Department of Master of Computer Application, Osmania University, Hyderabad, India

comprising three feature abstraction blocks and a feature fusion segment. To train the CNN, we assembled a dataset of recolored images and their corresponding ground truth, generated using various recoloring techniques. Through extensive experiments on a diverse set of recolored images, our proposed network demonstrates excellent generalization and robustness, making it highly effective for detecting image recoloring.

1 Introduction

Today, pictures have turned into an unavoidable piece of our existence with the normal use of shrewd procurement gadgets like cameras and cell phones, and the simplicity of sharing over the Web. In up with this reality, the quantity of picture-handling programming increments. They have become open, with the end goal that anybody may handily adjust and share online pictures. On equal, altering pictures has become means to hurt essentially our general public. A few methods have arisen, among them: joining, duplicate move, and expulsion are the most regularly utilized controls.

1. **Splicing**: This is a control method that duplicates one or numerous locales of a picture and glues them onto another picture. It tends to be utilized to add an extra component to a scene. The principal section of Fig. 1 shows a grafting example 1: unique picture on top and the controlled picture beneath with someone else.
2. **Copy-Move**: It very well may be utilized to add bogus data or concealing data. The next section of Fig. 1 depicts a unique picture on top and the controlled picture underneath with additional wellsprings.

Fig. 1 Applications of splicing, copy-move, and removal

3. **Removal**: It governor that replaces explicit pieces of a picture by, for instance, utilizing imprinting ways to deal with fill the unaccounted for pieces. It very well may be utilized to eliminate objects fully intent on concealing data. The third section of Fig. 1 shows an expulsion, the first picture on top and a missing angler on the controlled picture underneath.

Copy picture location is gotten by matching two unique pictures separately. This course of matching aids in the recognition of produced picture. There are a few visual applications which are devoted with part of endeavors. These visual applications need effective picture comparability marks and picture similitude measurements. In our current market, there are a few picture-handling programming which can without much of a stretch alter and control the first computerized picture. This tends us to move the matching changed pictures to their firsts, which is known as close-to-copy picture identification. Our framework examines the writing checked on the improvement of a few picture matching calculations.

A procedure which changes or changes the variety or topic of unique picture is known as picture recolouring which gets a subtle change in natural eyes. One of the most outstanding picture control processes is picture recolouring. For this sort of fraud, there is no extraordinary technique intended to recognize it. In this paper, we proposed a start to finish framework which recognizes the first picture from changed picture. In this manner, the recognization of recoloured pictures from regular pictures is drawn closer.

The proposed model takes the first picture and another picture, then, at that point, in light of the between channel connection and brightening consistency of the first picture is contrasted and another picture then the result likelihood is gotten. In this proposed framework, we utilized a calculation called CNN. Our calculation takes on CNN-based design, which comprises of three element extraction blocks and a component combination module.

2 Literature Survey

2.1 *Camera Method Identification (CMI)*

The interaction of picture which comes out of a camera manufacture or model has been utilized to catch a picture and is known as camera method identification (CMI). This cycle of CMI has gotten an enormous attention as it could be utilized as evidence in unambiguous lawful issues. Scientists from the picture scientific local area have proposed various ways to deal with distinguishing a CMI. Basically, they planned to separate the picture impressions port by the Cam advanced courses throughout the obtaining [1]. This cycle is separated into a few stages within the camera gadget leaving explicit highlights that can be taken advantage of during the ID interaction. Those finger impression highlights when assembled are interesting and the mark of a particular camera model. In this way, it permits to distinguish approximately

metadata like its starting point, the potential cycles applied and the uprightness of the first pictures.

Different works depended on advanced parts. In [2], they utilized the data produced by the gif, jpeg pressure development and furthermore impressions port by demons [3–5]. As the picture obtaining process channel is hard to show, other CMI methods utilized highlights which consolidate picture measurement properties and administered AI methods. Specialists in [6] have introduced a nearby paired design procedure that locally catches adjoining pixel relations. Two or three examination works [1, 7] depended on the variety channel cluster (CFA). The course of interjection utilized for picture grouping is a connection arrangement existing in RGB variety band of the picture. The important supposition of the creators is the data gain given by utilizing CFA addition calculation. Their outcomes showed the force of CFA insertion to decide the camera model of the pictures.

Our inspiration is to set up a profound discriminative organization for concealing move location. In like manner, we discuss the main computations including fraud identification methodologies, concealing move approaches in this portion. A fabrication identification strategies imitation recognition methods expect to really look at the validness of pictures and can be broadly described into two classes: dynamic check and latent affirmation. Reinhard et al. propose a concealing move procedure by exhaustively moving shades. They apply a straightforward real assessment to constraining one picture and concealing characteristics on one more in the lab concealing space. The concealing moving can reasonably and capably make a convincing yield. Fined probabilistic model is used in furthermore work on this methodology. To all the more probable perform nonlinear concealing modifications. Beigpour et al. present an actual model of the picture improvement and apply to concealing moving, making the results progressively reasonable. Every one of the above procedures requires a model picture as data and we call this kind of methodologies model-based reshading. Pitie et al. utilize a N layered probability thickness limit and use a post-handling computation to keep up the point field of the main picture. Chang et al. separate a concealing range of a picture by gathering and make a supportive instrument for reshading by changing a concealing range. In spite of the way that these reshading computations might leave no visual clues, these methods might change the essential picture surfaces. In this work, we take focal points of two surfaces to perceive whether a picture is manufactured.

3 Proposed System

We will discuss our method in a great detail in this section. Its head objective is to give a powerful and viable system for CMI and picture fabrication location. It shows the worldwide structure of proposition. This structure contains four explicit parts. Initial segment worries all the picture prior-handling process. In this section, we additionally show the significance of considering the nature of the information to fortify the heartiness. Then, at that point, we make sense of the grouping method

utilizing convolutions brain organizations to distinguish camera models CMI. Next part features a top-to-bottom examination of our CNN which permits us to all the more likely comprehend and work on our structure. At long last, we test our proposed structure on a phony discovery application.

Existing forgery identification strategies take on some portrayal methods to join the data accomplished by proof assessors. Nonetheless, every depiction method has its own limits and downsides. As of late, CNNs [8] have shown a dangerous prevalence in picture order and other PC vision errands. Customary brain networks utilize the first picture in RGB networks as the contribution since it comprises data around the image like tone and primary elements. In this discussion, we utilize three-element extractions and a component combination module to learn phony significant highlights. The flow diagram of our planned methodology is We embrace the first picture as one of the information branches like customary brain organizations. Moreover, we determine as two bits of proof of picture recolored location in view of the perceptions that pictures may not keep up with the between channel relationship or illuminant consistency after the recolouring system.

These two bits of proof are utilized as two extra information branches along with the first picture. The organization engineering can be viewed as in. Since the learned elements depend on an information [9] driven approach, they can portray the characteristic properties of falsification development and help recognize the credibility of a picture. Subsequent to removing fabrication important elements, we utilize a component 13 combination organization to refine these highlights and result in the likelihood of credibility. In view of this reason, we assess the proposed calculation on manufactured pictures produced by different variety move strategies and the pictures gathered through the Web.

3.1 Image Processing

This section outlines the pre-design steps we have chosen to take into consideration (see Fig. 2). CNN needs a lot of time to prepare its data. Pre-handling methods add new types of current images to the current dataset, expanding it. This is done to discover how the mind links a large amount of info. Change models include scaling and tension, among others. We provide experiences with regard to each component employed for image pre-taking care of it throughout the accompanying subsections (Fig. 3).

Picture Changes: A crucial step in achieving high accuracy while developing a significant learning model is the quality and evaluation of the readiness dataset. Applying our method to the image impersonation field to electronically posted pics is one of our experiment's goals.

Patch Extraction: The photographs from the dataset are separated into 64×64 pixel blocks in the second stage of our construction. The following arguments unquestionably affect the decision to use small pictures rather than fully objective photos: (i) it
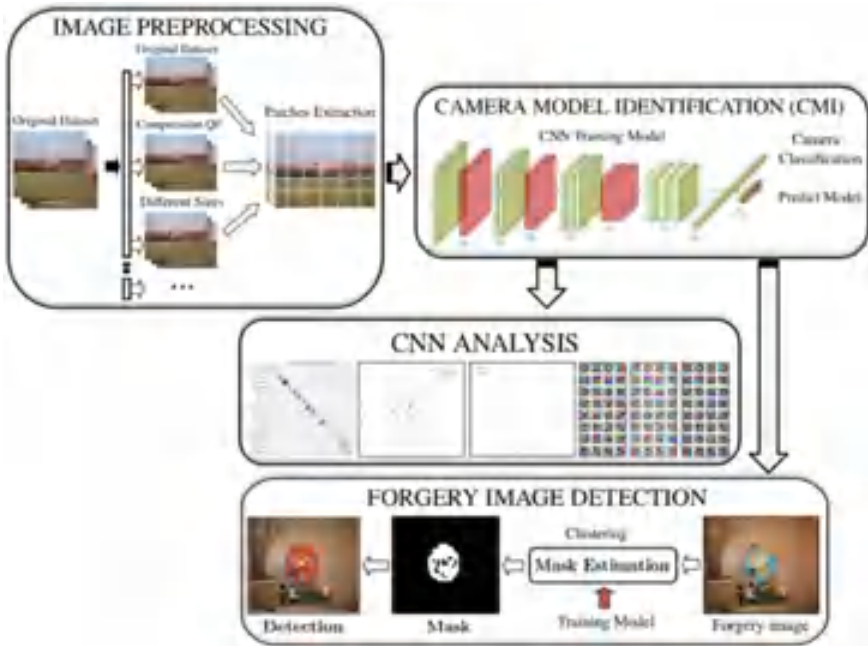
Fig. 2 Typical process of digital image forgery detection system (DIFS)
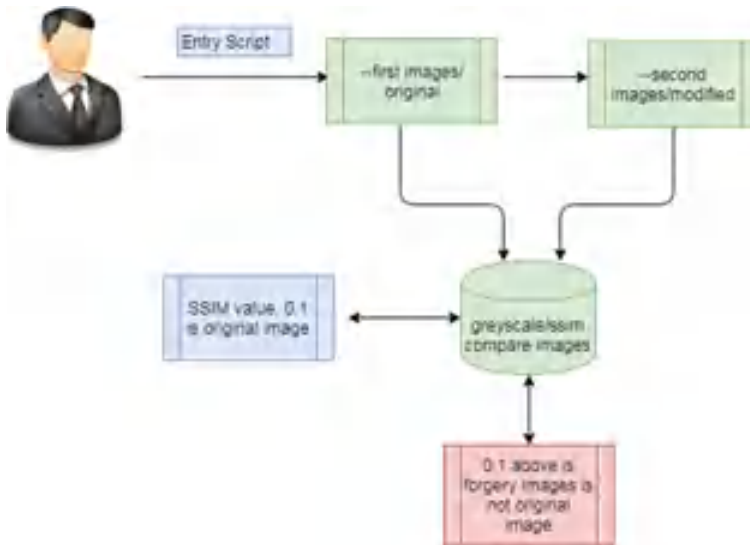
Fig. 3 Architecture of DIFS

reduces the amount of space needed to represent the camera model; (ii) the accuracy of the rough estimates used in the framework; and (iii) the data increase the probability of training the important learning model. Furthermore, according to [10], having to cut methods for portraying's evaluations has demonstrated better execution when using small picture patches. We choose the regions that have excessively boring or soggy terrain because the idea of the data is essential to building up the model.

Convolutional Brain Networks for Camera Model Recognizable Proof: Indeed, even its remarkable potential, significant learning has ended up being undeniable for camera method conspicuous verification. In this approach, we adventure convolutional mind associations (CNN) to remove camera method components from picture coverings. The primary CNN configuration expressly devoted to CMI has been projected in [10]**.**

In this work, we utilized a comparative organization. This decision is propelled with the plan to accomplish a high result camera model attribution precision with a minuscule organization design. The pre-owned network contains nine layers specifically four convolutional layers, three maxpooling layers, and two completely associated layers (1 ReLU layer and one softmax layer). Picture patches are taken care of in the CNN through an info layer, otherwise called the information layer.

3.2 CNN Analysis

CNN has become a known tool with great performance in a variety of application areas because of their significant capacity to create a supported part portrayal. Experts from these fields, however, need more significant interpretability from CNN's "disclosure" of response to the common criticism that the learnt qualities in a cerebrum association are not subject to interpretation; a few approaches for comprehending and envisioning convolutional associations have been made in the writing [4, 5]. These acquired characteristics are in fact difficult to decipher from a human expert perspective.

Convolutional layer (Conv1) has a kernel size of $4 \times 4 \times 3$ with 32 feature maps as output. Input is a set of patches of size $64 \times 64 \times 3$. The convolutional filter is applied with a stride of 1.
Convolutional layer (Conv2) contains 48 filters of size $5 \times 5 \times 32$ (stride = 1). It generates, as output, $28 \times 28 \times 48$ feature maps.
Convolutional layer (Conv3) contains 64 filters of size $5 \times 5 \times 48$ (stride = 1). It generates, as output, $10 \times 10 \times 64$ feature maps.
Convolutional layer (Conv4) contains 128 filters of size $5 \times 5 \times 64$ (stride = 1). It generates, as output, a vector of 128 feature maps.
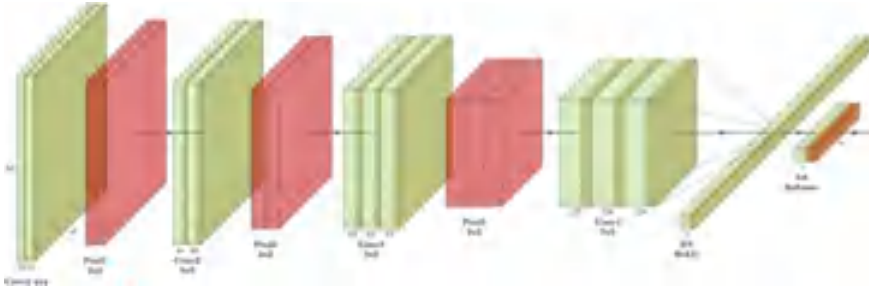
Fig. 4 CNN architecture [10]. Conv denotes a convolutional layer, pool, a maxpooling layer, and Fc a fully connected layer

3.3 Forgery Image Detection

Produced pictures are produced using patches of perfect pictures of various camera methods and sticking them collected. The objective of our impersonation distinguishing proof methodology is to measure expecting the image is totally made from one CMI patch which suggests the image is immaculate then again if a couple of districts of the pic are from different camera methods.

For example, the pictures are manufactured. In the last option case, deciding manufactured area in info picture is additionally important for the exhibition assessment of our methodology. Figure 4 presents the various strides of the fraud recognition calculation. To start with, the picture is separated into patches. The prepared method concentrates a component vector of N components addressing the quantity of camera models. A bunching calculation utilizes this data to gauge a double altering cover. In this section, dark locales show patches having a place with the perfect district, and white ones demonstrate fashioned squares. The picture is thought of as perfect in the event that there is no white district.

3.4 Forgery Image Detection Algorithm

Fundamental equivalence contrasts concerning various strategies referred to ahead of time; for instance, MSE or PSNR is that these techniques measure by and large mix-ups; on the other hand, SSIM is a wisdom-based model that contemplates picture degradation as seen change in essential information, while in like manner combining huge perceptual eccentricities, including both luminance covering and distinction hiding terms. Essential information is the likelihood that the pixels have strong between conditions especially when they are spatially close. These circumstances convey huge information about the plan of the things in the visual scene. Luminance

covering is an eccentricity by which picture distortions (in this particular circumstance) will commonly be less perceptible in splendid regions, while contrast veiling is an idiosyncrasy by which mutilations become less clear.

The SSIM record is determined on different windows of a picture. The action between two windows x and y of normal size N $\times$ N is [11]:

$$SSIM(x, y) = (2\mu x \times \mu y + C1)(2\sigma xy + C2)(\mu 2x + \mu 2y + C1)(\sigma 2x + \sigma 2y + C2)$$

where μx, μy, σx, σy and σxy are the local means, standard deviations and cross covariance for images x, y. If $\alpha = \beta = \gamma = 1$ (the default for Exponents), and $C3 = C2/2$ (default selection of $C3$) the index simplifies to.

4 Experiments

This part presents our comprehensive investigation results. Later itemizing the examination arrangement including picked datasets and assessment models, we will propose a starter concentrate on featuring the significance of pressure as a picture control procedure. It is a rundown of our work done in [9]. Then, at that point, we will detail the presentation of our structure for camera model recognizable proof. We will likewise examine our proposed model for a superior comprehension of CNN. Lastly, we will concentrate because of speculation for the organization layers.

4.1 Experiment Setup

Dresden dataset [10] is a freely accessible dataset reasonable for picture source attribution issues. Dresden contains in excess of 13,000 pictures of 18 distinct camera models. For each exact scene, a few cameras were organized in roughly a similar position and took practically a similar picture. This implies that every scene is addressed by a few (or every) current camera. The information base is made out of variety pictures of extremely top notch pictures (around 4000×3000 pixels). Note that we chose just regular JPEG photographs from camera models with more than one occurrence. For the motivations behind learning, we isolated the dataset into discrete preparation, approval, and assessment sets indicated.

4.2 Evaluation Criteria

CMI: To assess the camera method recognizable proof execution, we utilize the typical exactness got with a greater part casting a ballot.

Phony Discovery: We assess identification execution on both "known" and "obscure" datasets with regard to exactness, collector working trademark (ROC) bends and region under the ROC bend (AUC).

These measurements are regularly known and utilized; they recognize obviously the distinction between the presentation of concentrated on approaches DT, DV, and DE individually. This then, at that point, brings about 7938 pictures in the preparing set, 1353 pictures for approval and 5400 pictures in the assessment dataset [15].

Visual descriptors give statistics about an image. A good descriptor permits to discriminate between similar and dissimilar images. Note that the notion of similarity highly depends on the application. For instance, similarity means "visually consistent images" in the framework of image retrieval while it signifies "visually nearly identical" in duplicate detection. There exist many published surveys on image description; the reader can refer for surveys centered around image description for content-based image retrieval applications. In the following, four types of low-level image descriptors are presented.

4.3 Influence of Compression on Camera Model Identification

In this part, we propose our starter concentrate on that features the significance of control process on the camera model ID(CMI) exactness of our structure indicated CNNm contrasted with the one proposed by Bondi et al. [27]. To make this power appraisal, we train and test all prepared CNN beforehand subtleties with the four datasets of various quality variables. Note that Bondi et al. is the CNN model prepared on "Unique" pictures. CNN70, CNN80, and CNN90 are the CNN prepared on packed pictures with particular quality components 70, 80, and 90. CNNm is the CNN prepared on blended uncompressed and compacted pictures (Figs. 5 and 6).

5 Results

From our proposed system, we have multiple SSIM values observed. When same images are sent for execution, the output of SSIM value will be 1.0. Similarly when the images are different; i.e., modified images are sent, then the SSIM value will be less than 1.0 (Figs. 7 and 8).
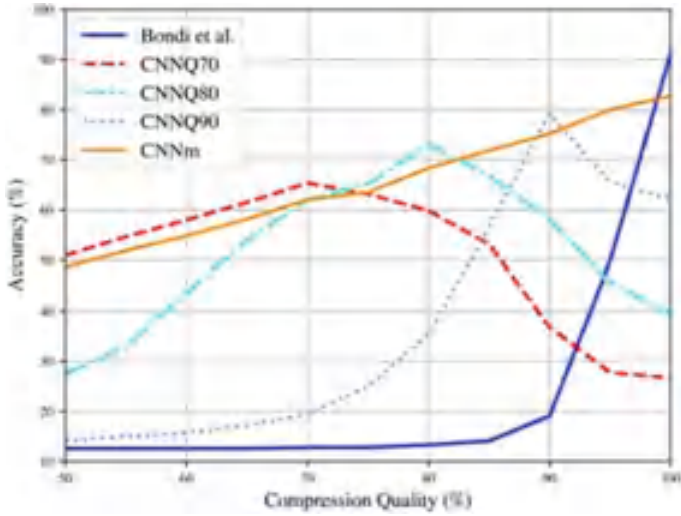
Fig. 5 Accuracy curves of CMI
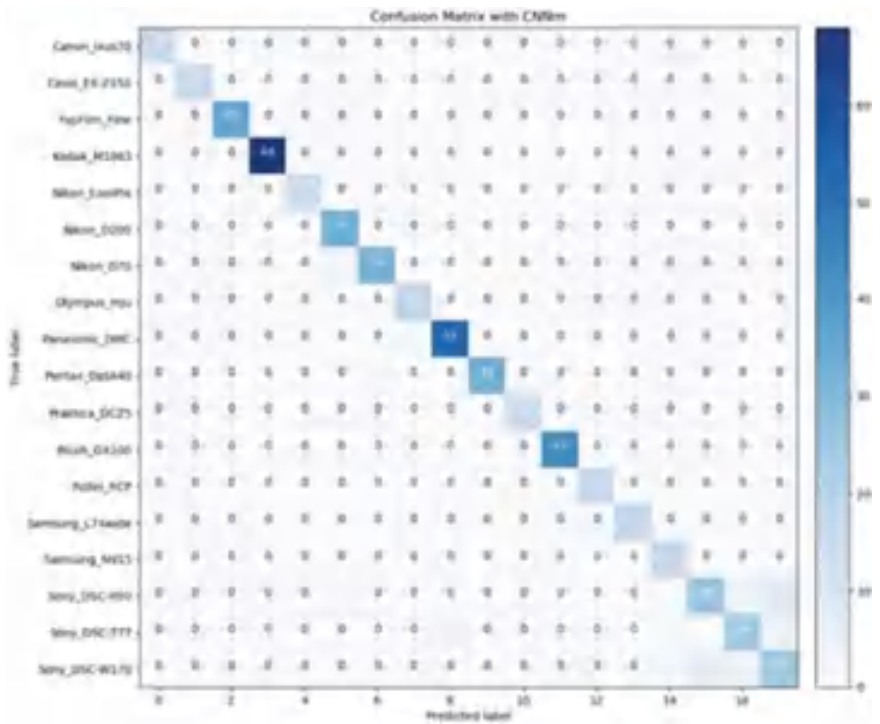
Fig. 6 Confusion matrix of CNN's model
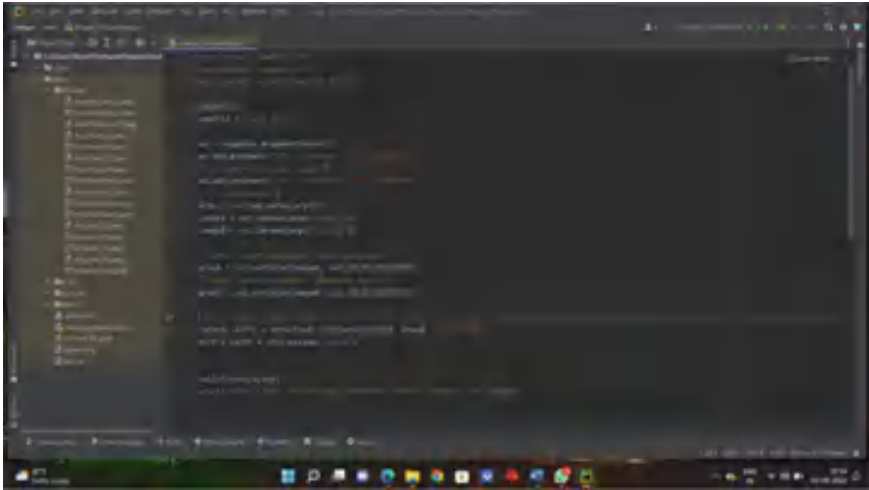
Fig. 7 Execution of an image code

Fig. 8 Considering an input image

6 Conclusion

In this work, we present an original profound learning approach for recolored picture location. Both the between channel connection and the enlightenment consistency are utilized to help the element extraction. We elaborate the planned standard of our new and deliberately approve the judiciousness by running various investigations. Moreover, two recolored datasets with various sources are made and the elite execution of our New exhibita the viability of the model. We trust our straightforward yet successful Late will act as a strong pattern and help future exploration in recoloured pictures location. Our future work will zero in on planning a more viable organization design and looking for a few undeniable level sings for better recognizing.

References

1. Sreenivasulu G et al (2011) Our-NIR: modified node importance representative for clustering of categorical data. Int J Comput Sci Inf Secur (IJCSIS) 9(4):146–150. ISSN 1947-5500
2. Marra F, Poggi G, Sansone C, Verdoliva L (2015) Evaluation of residual-based local features for camera model identification. In: International conference on image analysis and processing. Springer, pp 11–18
3. Sreenivsulu G, Viswanadha Raju S et al (2011) A threshold for clusting concept—drifting categorical data. In: IEEE 3rd International conference on machine learning and computing (ICMLC), vol 3, pp 383–387. ISBN: 978-1-4244-9253/11/$26,00@2011 IEEE
4. Tuama A, Comby F, Chaumont M (2016) Camera model identification with the use of deep convolutional neural networks. In: 2016 IEEE International workshop on information forensics and security (WIFS). IEEE, pp 1–6
5. Barni M, Bondi L, Bonettini N, Bestagini P, Costanzo A, Maggini M, Tondi B, Tubaro S (2017) Aligned and non-aligned double jpeg detection using convolutional neural networks. J Visual Commun Image Representation 49:153–163
6. Kharrazi M, Sencar HT, Memon N (2004) Blind source camera camera model identification. In: 2004 International conference on image processing, ICIP'04, vol 1. IEEE, pp 709–712
7. Sreenivasulu G, Viswanadha Raju S et al (2013) A comparative study of node importance in categorical clustering. Int J Adv Eng Global Technol (IJAEGT) 4(3):784–788. ISSN: 2309-4893 (print) | ISSN: 0975-397(online)
8. Amerini I, Uricchio T, Ballan L, Caldelli R (2017) Localization of jpeg double compression through multi domain convolutional neural networks. In: Proceedings of IEEE CVPR workshop on media forensics, vol 3
9. Filler T, Fridich J, Goljan M (2008) Using sensor pattern noise fore camera model identification. In: 15th IEEE International conference on image processing, ICIP 2008. IEEE, pp 1296–1299
10. Kee E, Johnson MK, Farid H (2011) Digital image authentication from jpeg headers. IEEE Trans Inform Forensics Secur 6(3–2):1066–1075
11. Bondi L, Baroffio L, G"uera D, Bestangini P, Delp EJ, Tubaro S (2017) First steps toward camera model identification with convolutional neural networks. IEEE Signal Process Lett 24(3):259–263
12. Cao H, Kot AC (2009) Accurate detection of demosaicing regularity for digital image forensics. IEEE Trans Inform Forensics Secur 4(4):899–910

13. Diallo B, Urruty T, Bourdon P, Fernandez-Maloigne C (2019) Improving robustness of image tampering detection for compression, pp 387–389
14. Sreenivasulu G, Viswanadha Raju S et al (2011) Our-NIR: node importance representative for clustering of categorical data. Int J Comput Sci Technol (IJCST) 2(2):80–82. ISSN: 2229-4333(print) | ISSN: 0976-8491 (onine)

Application of Machine Learning Algorithms for Power Theft Detection in Electrical Distribution System

P. Tejaswi and O. V. Gnana Swathika

Abstract The non-technical losses are prominently due to power theft in the distribution system, which results in profits' reduction, energy costs are also increased to other consumers, huge revenue loss to power utilities. Electric utility companies are facing issues in supplying electricity to their consumers in an effective manner. The precise detection of power theft is a challenging issue due to the overfitting issues, improper categorization on the imbalance power consumption data, and the increase in False-Positive Rate of the conventional methods. Hence, machine learning algorithms play a crucial role to detect the power theft and to restore huge revenue loss for utility companies. Random forest, support vector machine learning algorithms are applied to detect power theft. The performance metrics of this machine learning algorithms are evaluated and compared for accurate prediction.

Keywords Machine learning · Random forest · Support vector machine · Train data · Test data

1 Introduction

Nowadays, the serious issues faced by the developing countries are due to electricity theft which results in large amount of revenue loss to the economy of the country. Power loss is one of the significant problems that needs attention in the distribution system. Power theft is done in different ways which is not an easy task to find how a power theft has occurred. The issues of power theft should be solved as quickly as possible. The electricity theft is committed by consumers in multiple ways that

P. Tejaswi
School of Electrical Engineering, Vellore Institute of Technology, Chennai, India
e-mail: tejaswi.2019@vitstudent.ac.in

O. V. Gnana Swathika (✉)
Centre for Smart Grid Technologies, School of Electrical Engineering, Vellore Institute of Technology, Chennai, India
e-mail: gnanaswathika.ov@vit.ac.in

include bypassing and tampering the energy meter and, hooking of direct line in [1]. The connection of distribution network to power supply in shunt to the meter results in bypassing the meter. Village and suburban areas experience this type of theft where the insertion of magnetic interfering materials or terminals of the energy meter is shortened results in mal operation of the energy meter [2]. During power theft, there arises a situation like insufficient total load flow data, and compensation of static reactive power and power factor maintenance is a challenging issue in [3]. These losses affect electrical load on the distribution system, power supply quality, and electricity tariff's imposed on genuine customers.

Power theft is detected automatically with the implementation of IoT-based energy meter [4]. Microcontroller Atmega328P is used to identify and control power theft remotely by disconnecting and reconnecting a specific consumer's service [5]. IoT-based early warning system is designed to display the status and online theft monitoring of all equipment in the distribution network [6–9]. Machine learning algorithms plays a vital role in the prediction of power consumption and detection of electricity theft. Random forest (RF) and support vector machine (SVM) learning algorithms are explained in Sects. 2.1 and 2.2.

2 Methodology

Data collection and preprocessing, building algorithm, training and testing data of the algorithm, and evaluation of performance metrics are the basic steps involved in machine learning algorithms as shown in Fig. 1.

The data set is collected and preprocessed for accurate prediction of power theft using the machine learning algorithms. Preprocessing the data improves prediction accuracy of machine learning algorithm [10, 11]. On a big complex dataset, random forest and Support vector machine learning algorithms are employed to train an ideal system for ML prediction. A sample load data set of size 5000×13 is taken from Kaggle website which contains the consumption of the various electrical appliances. The whole data set is split into train data and test data in which 80% of the data set is chosen as training data and 20% of the dataset is chosen as testing data. Random forest and support vector classification machine learning algorithms are applied for detection of power theft to commercial loads. The performance metrics of this machine learning algorithms is evaluated and compared for commercial load data set.

2.1 Random Forest Machine Learning Algorithm

The following steps illustrates the application of random forest algorithm as in Fig. 2 to resolve the power theft issue in the distribution system in [10].
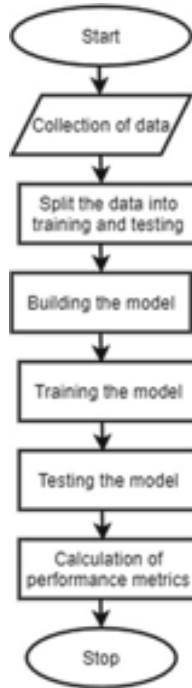
Fig. 1 Basic machine learning model
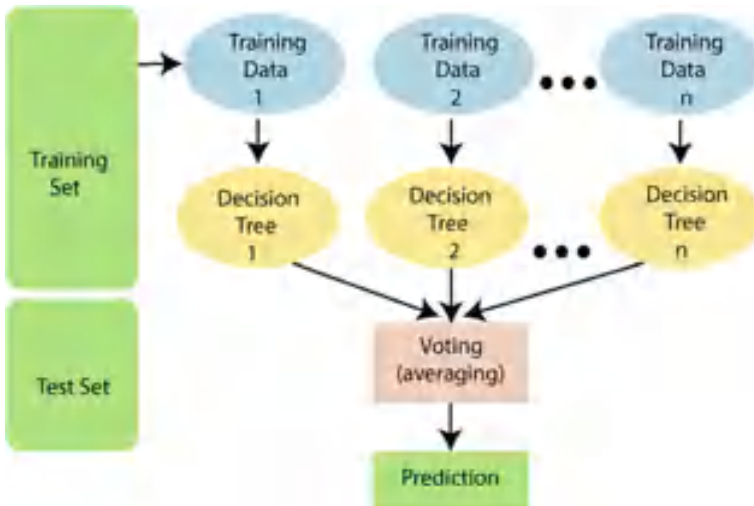
Fig. 2 Random forest machine learning algorithm

Step 1: Random samples are selected from a commercial load dataset.

Step 2: For every sample of training set. A decision tree is constructed. Predicted output is acquired from each and every decision tree.

Step 3: Voting will be accomplished on each and every predicted output of power theft.

Step 4: The most voted predicted output is chosen as the final prediction output of power theft.

2.2 Support Vector Machine Learning Classification Algorithm

In this algorithm, commercial load data set is collected from Kaggle website. The data are preprocessed, trained, and tested for accurate detection of electricity theft [10, 11]. Figure 3 illustrates the flowchart of support vector machine learning algorithm.

Fig. 3 Support vector machine learning algorithm
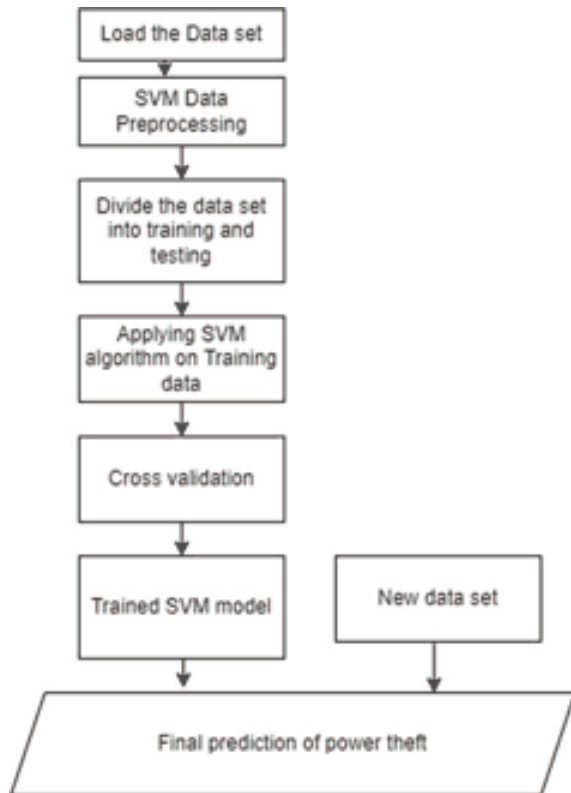
3   Results and Discussion

In random forest (RF) and support vector machine (SVM) learning algorithms, seventy-five percent data are chosen as training data and twenty-five percent data are chosen as testing data. The input variables in the data set are Fans, Interior lights, Water heater, Gas facility, Heating facility, and cooling facility. The target variable in this machine learning algorithms is power theft detected which is a Boolean variable. Accuracy metrics is calculated for the test commercial load data set of this algorithms which is tabulated in Table 1. Figures 4 and 5 depict the training set size versus accuracy score in which training score (training data) and cross-validation score (test data) learning curves are evaluated for this algorithms. The accuracy score of training score is equal to 1 and the accuracy score of cross-validation score is approximately 0.98 in RF machine learning algorithm. By increasing the number of trees in RF machine learning algorithm, accuracy score is increased. The accuracy score of training score is nearly equal to 0.98 and the accuracy score of cross-validation score is approximately 0.97 in SVM learning algorithm.

Table 1 Comparison of performance metrics of RF and SVM machine learning algorithms

Machine learning algorithm	Accuracy score	Absolute mean error	Mean-squared error
Random forest	0.98	0.988	0.009
Support vector machine	0.97	0.983	0.011
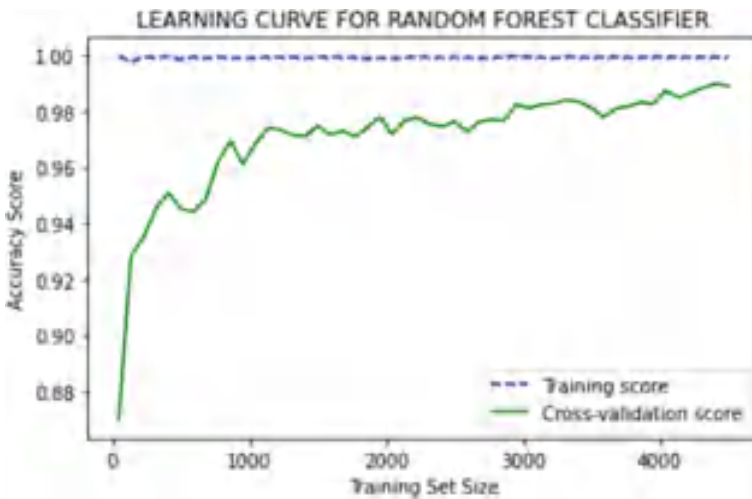
Fig. 4 Learning curve of random forest classification algorithm for detection of power theft
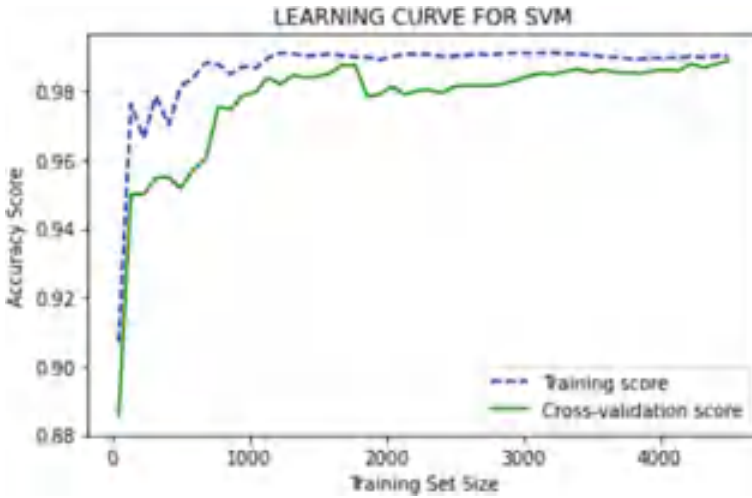
Fig. 5 Learning curve of support vector classification algorithm for detection of power theft

The accuracy score, bsolute mean error, Mean squared error are compared for the random forest and support vector machine learning algorithms in Table 1. The accuracy of random forest machine learning algorithm is more accurate when compared to support vector machine learning algorithm.

4   Conclusion

Power-theft is one of the major challenges which results in data manipulation in the distribution system. Authentication, Confidentiality of critical data, and power quality of data are the crucial factors for distribution system reliability and efficiency. RF machine learning algorithm is more accurate when compared to SVM learning algorithms for detection of power theft. The performance metrics of RF and SVM learning algorithms are evaluated and compared for commercial load dataset of the distribution system. RF and SVM algorithms can be implemented for real-time hardware models to detect power theft and alert the user.

References

1. Amin M (2010) Toward self-healing infrastructure systems. Computer 33(8):44–53
2. Hasan MM, Mouftah HT (2015) Encryption as a service for smart grid advanced metering infrastructure. In: IEEE symposium on computers and communication (ISCC), pp 216–221
3. ECI Telecom Ltd. (2011) Fighting electricity theft with advanced metering infrastructure

4. Uddanti S, Joseph C, Kishoreraja PC (2017) IoT based energy metering and theft detection. Int J Pure Appl Math 117(9):47–51
5. Metering AS, Visalatchi S, Sandeep (2017) Smart energy metering and power theft control using Arduino & GSM. In: 2nd international conference for convergence in technology (I2CT). IEEE, pp 858–961
6. Qing-Hai O, Zheng W, Yan Z, Xiang-Zhen L, Si Z (2013) Status monitoring and early warning system for power distribution network based on IoT technology. In: 3rd International conference on computer science and network technology. IEEE, pp 641–645
7. Bhargav P, Nishat F, Ansari UA, Gnana Swathika OV (2022) IoT based smart metering, smart buildings digitalization, 1st edn. CRC
8. Majee A, Bhatia M, Swathika OVG (2018) IoT based microgrid automation for optimizing energy usage and controllability. In: Second International conference on electronics, communication and aerospace technology (ICECA), pp 685–689
9. Naveen Kumar G, Gnana Swathika OV (2022) AI applications to renewable energy–an analysis. In: Smart buildings digitalization, 1st edn. CRC
10. Li S, Han Y, Yao X, Yingchen S, Wang J, Zhao Q (2019) Electricity theft detection in power grids with deep learning and random forests. J Electr Comput Eng 2019:12. Article ID 4136874
11. Depuru SSSR, Wang L, Devabhaktuni V (2011) Support vector machine based data classification for detection of electricity theft. In: 2011 IEEE/PES power systems conference and exposition, pp 1–8

Comparative Analysis of ML Algorithms' Application in SAPV Generation Systems

Aadyasha Patel and **O. V. Gnana Swathika**

Abstract Electricity is in excessive demand in both the developing and developed countries. To meet this ever-increasing demand, the supply of electricity needs to be sufficient. For meeting this requirement, prior knowledge of monthly/daily/hourly energy consumption data is necessary so that adequate energy can be generated. ML algorithms are used to forecast energy usage data. Forecast is obtained using pre-recording, training and testing the data using ML algorithms. Sometimes there are discrepancies in logging the data, like erroneous or missing values. Those can also be filled-in using various ML techniques. In this chapter, a SAPV system is set up whose input and output parameters are recorded. Any missing values here are forecasted using three algorithms—Multiple Imputation using Chained Equations, Case-Based Reasoning with and without interpolation. The output is analyzed and compared using error metrics such as MAD, MSE, RMSE and MAPE. The %MAPE of Case-I, Case-II and Case-III is 10.065%, 13.979% and 15.472%, respectively. On comparing all the three cases, the values forecasted by Case-I yield better results.

Nomenclature

ML	Machine learning
SAPV	Stand-alone photovoltaic
MAD	Mean absolute deviation

A. Patel
School of Electrical Engineering, Vellore Institute of Technology, Chennai, India
e-mail: aadyasha.patel2019@vitstudent.ac.in

O. V. Gnana Swathika (✉)
Centre for Smart Grid Technologies, School of Electrical Engineering, Vellore Institute of Technology, Chennai, India
e-mail: gnanaswathika.ov@vit.ac.in

MSE   Mean square error
RMSE   Root mean square error
MAPE   Mean absolute percentage error
CBR   Case-based reasoning
MICE   Multiple imputation using chained equations
PLX-DAQ   Parallax data acquisition
p, q Euclidean space points
q_i, p_i Euclidean vectors
n Number of values
A_t Actual value
F_t Forecasted value

1 Introduction

While recording any kind of parameters from a stand-alone photovoltaic system, certain information may not get saved or an erroneous value may get recorded. To fill-in the flawed and missed values with the appropriate data, ML algorithms are used. One such algorithm is CBR as explained by [1] which stores its problem-solving experiences in a case base. For every new problem to solve, like in [2–4], it takes the reference from the case base. It finds similar instances to find solutions for new problems. According to [5], this method is applicable in all missing data conditions of the categorical and numerical type. Another algorithm discussed in [6, 7] is a variation of CBR which is implemented with interpolation. It finds two approximately similar solutions to the problem. The principle of interpolation from [6] is being to find out the solution of the problem from between the two approximately similar solutions. MICE is the third algorithm to be discussed in this chapter for filling-in the missing values of the dataset. As reported in [8], MICE works by running multiple regression models to fill-in for the missing data. The authors in [9, 10] have talked about few areas where MICE is implemented. The results obtained were cross verified with very low percentage error.

2 ML Forecasting Techniques

There are various machine learning algorithms available for forecasting the data. In this chapter, the data will be compared and analyzed using the following algorithms:

- Multiple Imputation by Chained Equations.
- Case-Based Reasoning with interpolation.
- Case-Based Reasoning without interpolation.
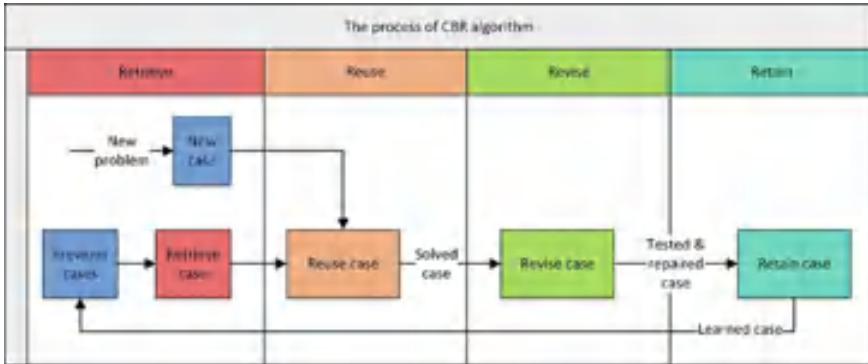
Fig. 1 Process of CBR algorithm

2.1 CBR Algorithm

CBR algorithm works by using past experiences as reference to resolve present conflicts. It is said that the CBR algorithm behaves the same way a human does while making a decision, i.e., by recollecting, comparing, modifying and implementing the result (see Fig. 1).

Euclidean Distance Equation. The CBR algorithm finds out the most relevant solution by implementing the Euclidean Distance Equation. Equation (1) finds the shortest distance between the points (new case and previous case) by drawing a straight line between them. The equation is as follows:

$$d(p, q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}. \tag{1}$$

Interpolation Technique. Construction of the new dataset from already saved dataset is called interpolation mathematically. A new data model is created using (1), then interpolation of already saved dataset takes place. The data model and the dataset are compared for missing values to fill-in the lost data.

2.2 MICE

The corresponding neighboring columns of a missing value in a dataset are compared with other columns. This way of comparing and analysis of datasets to forecast the missing values takes place as iterations. The procedure continues until convergence is reached and all the missing values are filled (see Fig. 2).
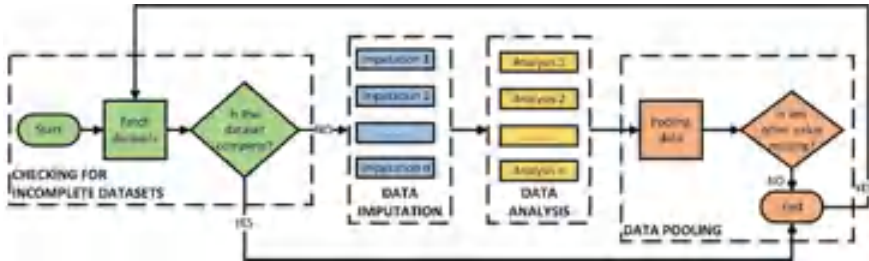
Fig. 2 Process of MICE algorithm

2.3 Results and Discussion

Data Collection. The data are collected using open-source hardware and software. The list of components used is listed in Table 1. The panel of the stand-alone photovoltaic system is mounted on a rooftop. Its sensors are connected to the panel to sense the incoming voltage and current. These values are sent to be logged in PLX-DAQ Excel sheet via an Arduino Uno board. The data are time-stamped and logged from 12 Noon to 2 PM for over a week. The time gap between two successive entries being 5 s is logged and saved.

Comparison of Algorithms. The three cases are executed and their forecasted values obtained via the ML algorithms are analyzed and compared using the error metrics listed in Table 2. Here, Case-I represents MICE-forecasted voltage, Case-II represents forecast using CBR with interpolation and Case-III represents forecast using CBR without interpolation. Figures 3, 4 and 5 depict the comparison between predicted and actual voltages, comparison between predicted and actual currents and comparison between predicted and actual powers, respectively.

Table 1 Component specifications

S. No.	Component	Specification
1	Solar panel	Rated power $= 10$ W Rated voltage $= 17$ V Current at maximum power $= 0.57$A Open-circuit voltage $= 21$ V Short-circuit current $= 0.69$ Tolerance $= \pm 5\%$ Irradiation $= 1000$ W/m^2 Cell temperature $= 25$ °C
2	F031-06 voltage sensor	(0–25)V
3	ACS 712 current sensor	(0–30)A
4	Arduino Uno Board	
5	PLX-DAQ	MS Excel Add-on

Table 2 Error metrics

S. No.	Error metrics	Formulae
1	MAD	$\dfrac{\sum_{t=1}^{n}\lvert A_t - F_t\rvert}{n}$
2	MSE	$\dfrac{\sum_{t=1}^{n}(A_t - F_t)}{n}$
3	RMSE	$\sqrt{\dfrac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}}$
4	MAPE	$\dfrac{\sum_{t=1}^{n}\left\lvert \frac{A_t - F_t}{A_t}\right\rvert}{n} * 100$
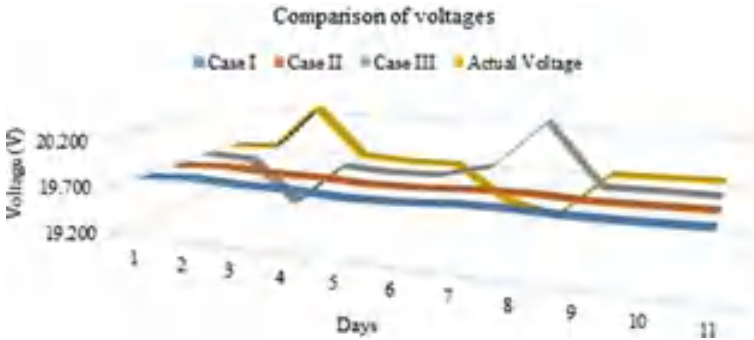
Fig. 3 Comparison between predicted and actual voltages
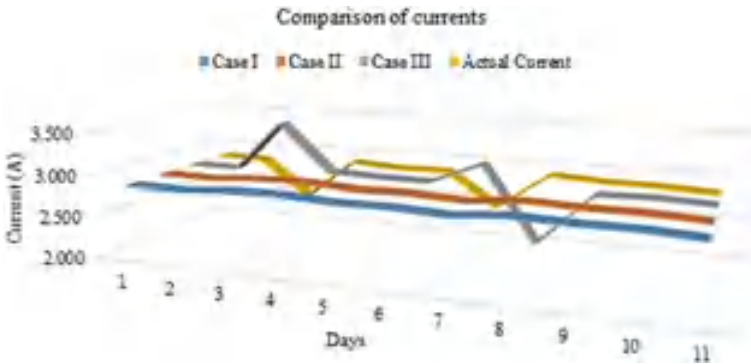
Fig. 4 Comparison between predicted and actual currents

Lower error metrics imply that the working system is efficient. From Table 3, it is observed that the MAD, MSE and RMSE of Case-I for voltage are 0.261, 0.107 and 0.304, respectively. The MAD, MSE and RMSE of Case-I for current are 0.263, 0.129 and 0.322, respectively. Similarly, the MAD, MSE and RMSE of Case-I for power are 5.641, 51.779 and 6.530, respectively. These values are lower compared with the values corresponding to Case-II and Case-III. The respective %MAPE of
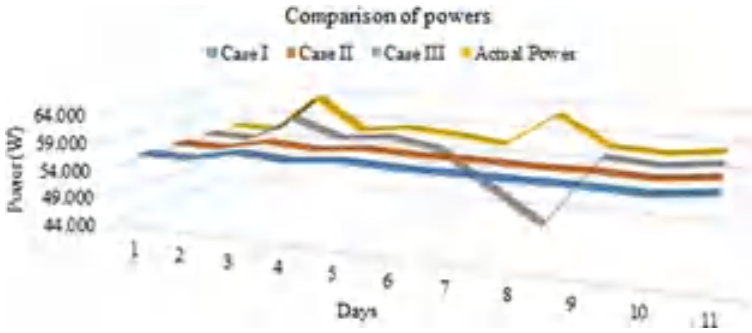
Fig. 5 Comparison between predicted and actual powers

Table 3 Comparison of error metrics

		Voltage	Current	Power
MAD	Case-I	0.261	0.263	5.641
	Case-II	0.367	0.366	7.850
	Case-III	0.420	0.394	8.707
MSE	Case-I	0.107	0.129	51.779
	Case-II	0.198	0.199	90.722
	Case-III	0.259	0.260	114.886
RMSE	Case-I	0.304	0.322	6.530
	Case-II	0.444	0.445	9.494
	Case-III	0.508	0.499	10.675
% MAPE	Case-I	1.321	9.444	10.065
	Case-II	1.856	13.073	13.979
	Case-III	2.122	14.062	15.472

Case-I, Case-II and Case-III is 10.065, 13.979 and 15.472%. On comparing all the three cases, the values forecasted by Case-I yield better results. Hence, the forecasts made for the stand-alone photovoltaic system by MICE are much better than the other two algorithms.

3   Applications

The MICE algorithm can be applied to a dataset whose particular group of data is not available for a certain period of time. During the process of data collection, multiple features of data can be contemplated.

4 Summary

This chapter discusses about filling the missing data recorded by a stand-alone photovoltaic system using three different ML forecasting methods. Apart from recording, the saved data are analyzed and compared to obtain the optimum result. The results show that MICE is a better option for forecasting of lost values compared to CBR with and without interpolation. CBR requires similar reference cases for giving out results, but MICE does not need so. The error metrics from Table 3 prove that MICE provides lower errors than the other two algorithms.

References

1. Kolodner JL (1992) An introduction to case-based reasoning. Artif Intell Rev 6(1):3–34. https://doi.org/10.1007/BF00155578
2. Alazzam MB, Tayyib N, Alshawwa SZ, Ahmed MK (2022) Nursing care systematization with case-based reasoning and artificial intelligence. J Healthc Eng 2022:9. Article ID 1959371. https://doi.org/10.1155/2022/1959371
3. Anthony Jnr B (2021) A case-based reasoning recommender system for sustainable smart city development. AI Soc 36(1):159–183. https://doi.org/10.1007/s00146-020-00984-2
4. Seo SY, Kim SD, Jo SC (2020) Utilizing case-based reasoning for consumer choice prediction based on the similarity of compared alternative sets. J Asian Financ Econ Bus 7(2):221–228. https://doi.org/10.13106/jafeb.2020.vol7.no2.221
5. Löw N, Hesser J, Blessing M (2019) Multiple retrieval case-based reasoning for incomplete datasets. J Biomed Inform 92:103127. https://doi.org/10.1016/j.jbi.2019.103127
6. Patel A, Swathika OVG, Subramaniam U, Babu S, Tripathi A, Nag S, Karthick A, Muhibbullah M (2022) A practical approach for predicting power in a small-scale off-grid photovoltaic system using machine learning algorithms. Int J Photoenergy 2022:21. Article ID 9194537. https://doi.org/10.1155/2022/9194537
7. Swathika OVG, Kalyanasundaram K, Elavarasan RM, Khahro SH, Subramaniam U, Pugazhendhi R, Ramesh M, Gopalakrishnan RM (2021) A review on effective use of daylight harvesting using intelligent lighting control systems for sustainable office buildings in India. Sustainability 13(9):4973. https://doi.org/10.3390/su13094973
8. Patel A, Swathika OVG (2022) Photovoltaic system-integrated smart buildings: a mini review. In: Smart buildings digitalization: IoT and energy efficient smart buildings architecture and applications, pp 255–274. https://doi.org/10.1201/9781003201069
9. Ganapathy S, Bhaskarapillai B, Dandge S (2021) The effect of multiple imputations by chained equations on the factors associated with immunization coverage in India. Int J Heal Sci Res 11(6):249–262. https://doi.org/10.52403/ijhsr.20210638
10. Ruggles TH, Farnham DJ, Tong D, Caldeira K (2020) Developing reliable hourly electricity demand data through screening and imputation. Sci Data 7(1):155. https://doi.org/10.1038/s41597-020-0483-x

IoT-Based Protection of PV-Wind Integrated Microgrid System Fault Analysis Using Wavelet Approach

K. V. Dhana Lakshmi, P. K. Panigrahi, and G. Ravi Kumar

Abstract Nowadays, microgrid comprising of wind and solar is utmost extensively used in power sector to decrease structure losses and also to improve the dependability in the arena of electrical systems. Combination of generation missions enhances innovative energy sources to a prevailing power system network. So, there is a need to develop innovative safety pattern owing to variations in the topology and vibrant actions of the system. Profligate error recognition algorithmic tactics are essential to merge diverse kinds of producing causes and loads under smart environment. The safety order should allow bodily checking as well as values measured with the support of innovative skills. Internet of things (IoT) is an important foundation to observe electrical systems in numerous environmental settings of the system. Wavelet (WT) fundamentally explores the fault transitory signals of dissimilar occurrence and splits the waveform into diverse estimated and thorough measurement values, which offers the significant information regarding the sorting and position of fault. The recognition of faulty line by execution of wavelet thorough measurements of Bior1.5 mother wavelet and for position of transmission line fault using artificial neural network technique. The suggested way provides condition monitoring analysis of IoT-based protection of microgrid with grid connected and islanded mode using wavelet approach under various types of faults.

Keywords Wavelet transform · PV-wind source · Fault detection · Idle mode · Internet of things (IoT)

K. V. Dhana Lakshmi (✉)
Department of Electrical and Electronics Engineering, GNITS, Hyderabad, India
e-mail: dhanalakshmi.lakshmi61@gmail.com

GIET University, Gunupur, Odisha, India

P. K. Panigrahi
Department of Electrical and Electronics Engineering, GIET University, Gunupur, Odisha, India

G. Ravi Kumar
Department of Electrical and Electronics Engineering, Bapatla Engineering College, Bapatla, India

1 Introduction

The methods used previously for transmission line protection [1] are calculated by the fraction of the potential difference and current at specific relay power point [2] by measuring the impedance in comparators circuit. However, the above-stated protection method may not be apt because of huge recognition time and the separation of faulty element and also due to numerous disadvantages [3, 4] like the undesirable working due to power fluctuations and hefty stuffing situations that lead to falling stream and extent shutdowns [2]. Therefore, it is essential to consider about another safety as an alternative of distance safety scheme [2]. In [5], now numerical statement created relays intended aimed at diverse working form of the scheme is suggested to sense the error in the station and regions [2].

As there is a need for the mechanical protection of the system, Internet of things (IoT) has come out as unique imminent skills for a smart grid network [6]. The IoT-coupled power sector permits to improve extra projecting safety order not only electrical but also the machine-driven tasks in the current grid network [2] such as in strategy, assembly ordering, action and care [7]. Cyber-security [8, 9] elicits number of difficulties in the smart grid system that has been transferred and conferred in [10, 11] about accessibility, integrity and confidentiality [12].

A microgrid consists of three key modules; those are renewable energy sources, distribution and diverse loads. Microgrid protection has challenges which are complicated, and the strategy of safety scheme has to operate successfully to grid connected and idle mode. The safety scheme should be developed in such a way to work successfully for various accountability current extents in the system [13]. The safety scheme should be successful in solving the problem as compared with the ancient model power system where the current flow is unidirectional for radial system; then again in the case of microgrid, the current flow is bidirectional [14].

The PV-wind renewable energy stays earliest in power generation [2]. The choice of defensive component is difficult because of synchronization necessary amongst grid side over-current shield and distance shield at transmission line discussed in [15]; however, these kinds of schemes can be proficient of suppressing tasks of electrical defence system [2].

The advocated scheme desires a sooner reaction of energy system parameters and decreases strength loss and energy enhancement [2]. The protection machine can accomplish two essential chores more often than not fault detection and discrimination of fault vicinity beneath grid connected and isolation mode [2]. This allows for safety degree of the associated apparatus and additionally operating employees with immediate manipulation of terminated power loss [2]. In general, asymmetrical and balanced forms of faults befall in transmission referred to as LG, LL, LLG, LLLG and 3-segment ground (TPF). The fault detection and region are critical in protecting power region modules for continuing ordinary power float. A microgrid security set up a set of rules which is described in [16] by the assistance of transitory modern wave shape using wavelet precise measurements [2]. The counselled studies effort focusses safety observes of microgrid below grid-based and isolated mode with the

help of IoT tracking, and wavelet-based multi-resolution-analysis (MRA) is used with the standardization of measurements of Bior-1 five mother wavelet [2].

2 Calculation of Fault Using Symmetrical Components

Usually, power scheme grid might drop below 1-ϕ, 2-ϕ, 3-ϕ short-circuit faults. The investigation of faulted system can be accepted as a result of making usage of positive, negative and zero sequence components and their relations built with a kind of error in the structure [2].

The different types of symmetrical and unsymmetrical fault representations are shown in Fig. 1.

The successive currents are represented as follows:

$$
\begin{bmatrix} I_0 \\ I_1 \\ I_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix}
\tag{1}
$$

(a) Line to Ground Fault   (b) Line to Line Fault

(c) Double Line to Ground Fault   (d) Three Phase Fault
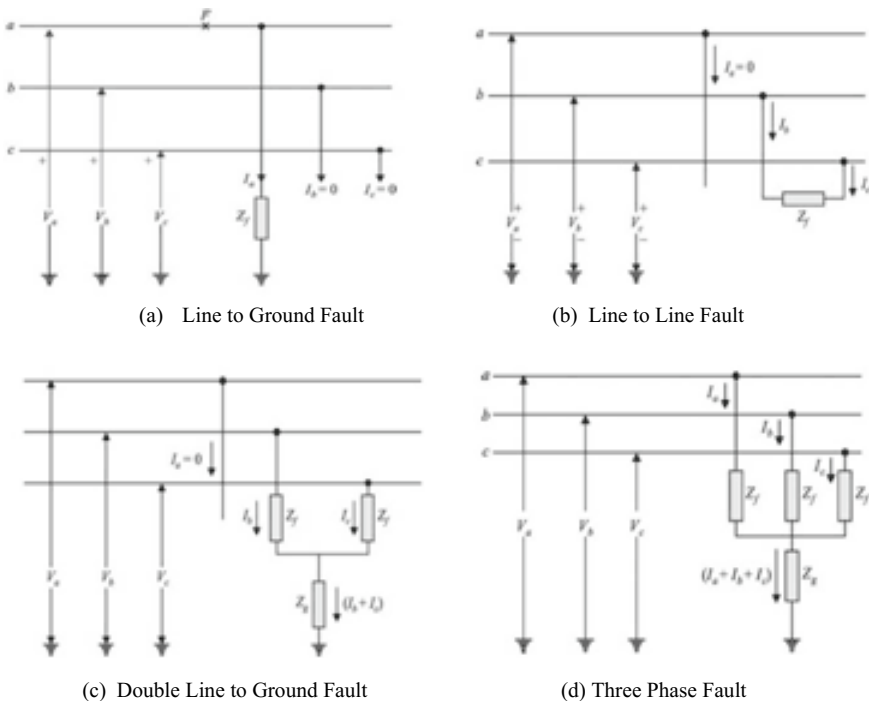
Fig. 1 Fault analysis using symmetrical components

the single line fault with suitable tagging is exemplified in Fig. 1a. All primary situations are presumed as below:

$$I_b = I_c = 0 \tag{2}$$

$$V_a = Z_f - I_a \tag{3}$$

$$I_0 = I_1 = I_2 = \frac{1}{3} I_a \tag{4}$$

$$I_0 = \frac{V_a}{Z_0 + Z_1 + Z_2 + 3Z_f} \tag{5}$$

The 1-ϕ-to-gnd fault current I_a designed as

$$I_a = 3I_0 = \frac{3V_a}{Z_0 + Z_1 + Z_2 + 3Z_f} \tag{6}$$

The two-phase fault is displayed in Fig. 1b.

$I_b = -I_c$ and $I_a = 0$; $V_b - V_c = I_b Z_f$. The sequence current of 2-phase-to-ground fault computations follows:

$$I_{a1} = \frac{V_a}{Z_1 + Z_2 + Z_f} \tag{7}$$

The double line fault current is given by

$$I_b = -I_c = \frac{-j\sqrt{3}V_a}{Z_1 + Z_2 + Z_f} \tag{8}$$

The LLG error is presented in Fig. 1c. The current is assumed as $I_a = 0$ and $V_b = V_c = (I_b + I_c)Z_f$.

The double line fault current is designed using Eq. 12.

$$I_1 = \frac{V_a}{Z_1 + \frac{Z_2(Z_0 + 3Z_f)}{Z_2 + Z_f + Z_0}} \tag{9}$$

As the fault is symmetrical,

$$\begin{bmatrix} V_0 \\ V_b \\ V_c \end{bmatrix} = \frac{1}{3} \begin{bmatrix} Z_f & 0 & 0 \\ 0 & Z_f & 0 \\ 0 & 0 & Z_f \end{bmatrix} \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \tag{10}$$

The error currents are designed as follows:

$$I_a = \frac{V_a}{Z_1 + Z_f}; \; I_b = a^2 I_1; \; I_c = a I_1 \tag{11}$$

3 Mathematical Modelling and Analysis System Components

The combination of solar and wind energy sources has become utmost prevalent than former foundations for the production of electrical power to the consumers [17]. The main benefit of alternative sources is accessible instantaneously, no shipping cost, lack of fossil fuels with little price dissimilarity and enhanced financial effectiveness. The planned power sector includes combination of photovoltaic and wind energy homes, linked to prevailing web. Altogether, the alternative and grid-related homes connected to AC transmission line over bus arrangement [17].

3.1 Solar PV System Modelling

Sun supremacy is the verve that is produced through the photovoltaic and is changed into electric verve using solar cell [17]. An approached electric path for PV display indicates I_{ph} characterizes the cellular photo-cutting-edge, shunt and series resistances are R_{sh} and R_{se} of solar cell [17]. PV modules are framed with the aid of amassing the series and shunt mixture module agencies. Those units remain related in series Ns for growing the voltage range, and Np modules are related in equivalent to surge concurrently receiving necessary energy technology [18]. The corresponding path of a PV cellular and PV group is shown in Fig. 2. PV exemplary mandatory apparatuses are buck–boost converter [17] and DC to AC converter as shown in discern-2. Mathematical equal circuit for picture-voltaic array as defined as follows:

$$I_{ph} = [k_i(T - 298) + I_{sc}]I_r/100 \tag{12}$$

Where I_{sc} in Amps and K_i is a constant, functioning hotness in Kelvin. Solar irradiance, I_r in W/m^2 [17]. The segment saturation current:

$$I_0 = I_{rs}\left[\frac{T}{I_r}\right]\exp\left[\frac{q * E_{g0}}{nk}\left(\frac{1}{T} - \frac{1}{T_r}\right)\right] \tag{13}$$

$T_r = 298.15$ K; $E_{g0} = 1.1$ eV;

$$V_t = \frac{k * T}{q} \tag{14}$$
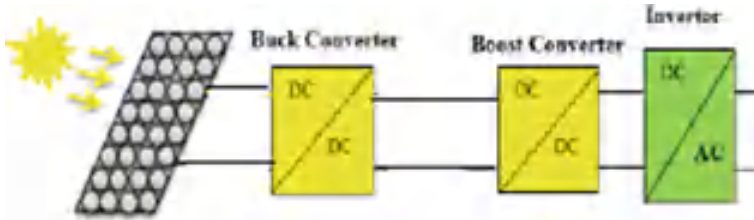
Fig. 2 Solar PV model representation

$$I_{sh} = \frac{V * \frac{N_p}{N_s} + 1 * R_s}{R_{sh}} \tag{15}$$

Solar energy segments can transform glowing solar energy straight away to small voltage to surge near to standard value with DC–DC converters and switch-mode regulators and now adjust it with a controlled DC production voltage [17].

3.2 Formation of Wind Energy Transformation

An airstream turbine yields little AC output voltage by altering airstream energy into electrical energy [17]. The produced small production voltage is improved to transmission voltage with combination of buck and boost converter [19]. The suggested arrangement contains doubly fed-induction-generator (DFIG) wind turbine with important modules [17]. The mathematical displaying and investigation have been labelled as tracks:

The DFIG airstream turbine is common amongst the existing skill that delivers power from the energetic force of airstream thru turbines and mills. The mechanisms have rotor and stator windings, so they are gifted in electric transmission to application grid [17]. The doubly fed-induction-generator wonderful capabilities are the two-manner power converter powering rotor and grid-attached stator, which incorporates two IGBT bridge voltage supply converters which are connected to DC link as exemplified in Fig. 3 [17]. They regulate the frequency vigorously and allow diverse turbine rapidity. The turbine rotor takes care over the energy and rotational velocity of heavy winds with blade attitude instruction as explained in [20]. The top speed ratio A is referred as tip blade speed to wind velocity [17]. It is expressed in equation:

$$A = \frac{\omega * R}{V_m} \tag{16}$$

Where ω stands for mechanical angular velocity of wind turbine and V_m for velocity of wind.
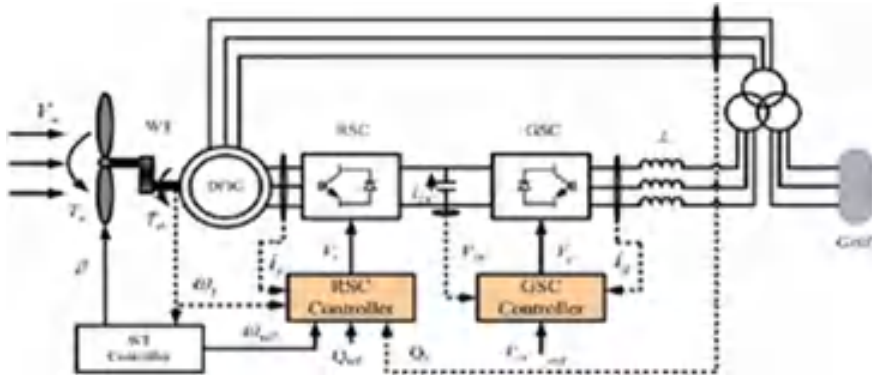
Fig. 3 Generation of energy using wind

$$T_m = \frac{1}{2} * \rho * A * v_u^2 * C_p \tag{17}$$

Where T_m mechanical torque. The direct and quadrature voltages of stator as per the synchronous reference frame theory are expressed as equation (3) and equation (4)

$$V_{ds} = R_{ds}i_{ds} + \frac{d\phi_{sd}}{dt} - \omega_{sd}\phi_{sd} \tag{18}$$

$$V_{qs} = R_{qs}i_{qs} + \frac{d\phi_{qd}}{dt} - \omega_{qd}\phi_{qd} \tag{19}$$

The direct and quadrature voltages of rotor of synchronous frame theory are represented as follows:

$$V_{rd} = R_{rd}i_{rd} + \frac{d\phi_{rd}}{dt} - \omega_r\phi_{rd} \tag{20}$$

$$V_{rq} = R_{rq}i_{rq} + \frac{d\phi_{rq}}{dt} + \omega_r\phi_{rd} \tag{21}$$

The real and reactive power controls are inscribed in expressions (7) and (8).

$$P_s = V_{sd}i_{sd} + V_{sq}i_{sq} \tag{22}$$

$$Q_s = V_{sq}i_{sd} - V_{sd}i_{sq} \tag{23}$$

$$P = \frac{V_s V_r}{X_L} \sin \delta \tag{24}$$

$$Q = \frac{V_s(V_s - V_r \cos \delta)}{X_L} \tag{25}$$

$$P^2 + \left(Q - \frac{V_s^2}{x_L}\right)^2 = \left(\frac{V_s V_r}{X_L}\right)^2 \tag{26}$$

Take on power due to reactive components as zero, so expression is changed as

$$P = \pm\sqrt{\left(\frac{V_s V_r}{X_L}\right)^2 - \left(\frac{V_s^2}{X_L}\right)^2} \tag{27}$$

A choice of real output provided by grid is known from expression (18) and then choice of 'Q' power is obtained by expression (20).

$$-\sqrt{\left(\frac{V_s V_r}{X_L}\right)^2 - \left(\frac{V_s^2}{X_L}\right)^2} \le P \le +\sqrt{\left(\frac{V_s V_r}{X_L}\right)^2 + \left(\frac{V_s^2}{X_L}\right)^2} \tag{28}$$

$$Q = \sqrt{\left(\frac{V_s V_r}{X_L}\right)^2 - P^2} - \frac{V_s^2}{X_L} \tag{29}$$

$$-\sqrt{\left(\frac{V_s V_r}{X_L}\right)^2 - P^2} - \frac{V_s^2}{X_L} \le Q \le \sqrt{\left(\frac{V_s V_r}{X_L}\right)^2 - P^2} - \frac{V_s^2}{X_L} \tag{30}$$

The K.E can be represented by the equation [17]

$$E = \frac{1}{2} * mv^2 \tag{31}$$

Where m is the air mass that is moving at velocity (v). The energy generated by wind i rate of change of kinetic energy is represented by equation [17]

$$P = \frac{dE}{dx} = \frac{1}{2}\frac{dm}{dx}V_w^2 \tag{32}$$

$$P = \frac{1}{2} * \rho * A * v_u^3 * C_p \tag{33}$$

V_u is denoted as upside wind rate on rotor blades in m/s, C_p the power coefficient of the rotor [17]

$$I_{ph} = [I_{sc} + K_i(T - 298)I_r]/1000 \tag{34}$$

The manual torque T_m is given by [17]

$$T_m = \frac{0.5\rho\pi R^2 C_p(\lambda, \beta) V_w^3}{\omega_t} \cdot \left[\frac{1}{T_{mBase}}\right] \tag{35}$$

The planned control system network [17] contains amalgam energy, i.e. mixture of solar and wind sources. All the alternative and net bases are linked to AC transmission line over bus arrangement [17].

4 Transmission System Protection Methods

Present-day energy systems require new methods to safeguard electric mundanely additionally electrically through the observation of machine execution. The protective system needs to notice ordinary disaster and electrical load variations through utilization of IoT application and fault recognition procedures.

IoT-based transmission system safety enables to offer system pushed and electric safety of power lines from the difficulties of usual adversities, unrefined dangers to shape, usual calamity and intensifying bushes as illustrated in Fig. 4a.

Wavelet rework (WT) is prominent technique to come across temporary faults by verifying distinct indicators and separating the predicted and thorough measurements using primary Bior1.five mom wavelet, giving vast data approximately error category and position [21]. A power device with microgrid defence procedure is defined in [22] multi-decision evaluation (MRA) with mother wavelets of defective indicators by bearing on their values with onset value. The counselled algorithm is defined in Fig. 4b.

(a) IoT protection   (b) wavelet based Algorithm
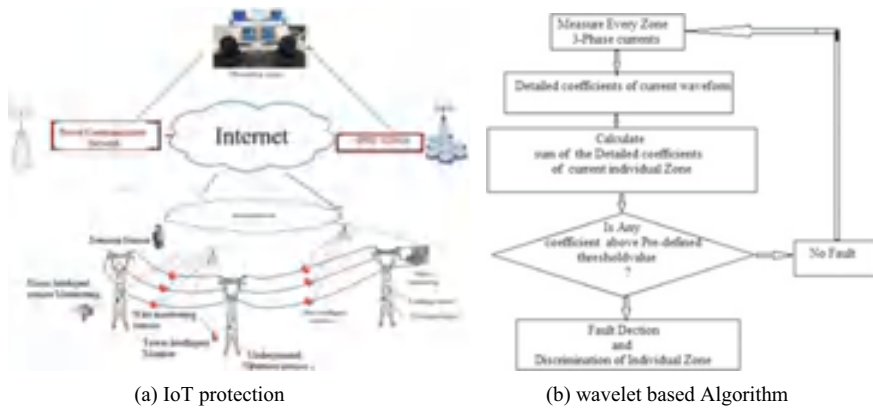
Fig. 4 Protection of microgrid using IoT
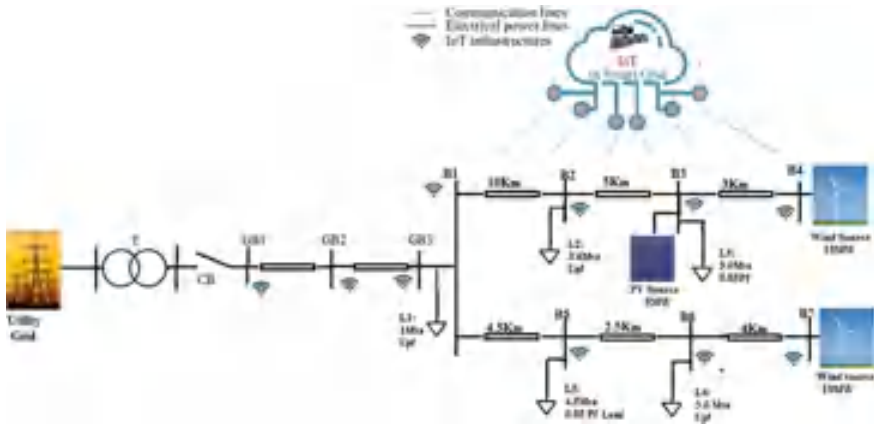
Fig. 5 System under consideration

5 Modelling and Implementation of Designed System

The arrangement consists of nine-bus microgrid associated system linked to grid and is divided into two sections:

One is utility grid side and other is microgrid side which contains 2-wind generation and 1-solar photo voltaic at bus numbers 4, 7 and 3. The proposed test scheme underneath learning is distribution network as shown in Fig. 5. The entire scheme is classified into eight zones as exemplified in Fig. 5, and arrangement of practical parameters is signified in Table 1.

6 Simulation Results

The suggested work reports detection and discrimination of error in numerous regions by making use of sum of the thorough measurements of current response of the scheme. The recognition of fault is noted through the scrutiny of fault table values. The phases are faulty or in good physical shape and are noticed by comparing their values with the predefined onset value which is described for the recognition and discernment of fault. Then the region's current signal has sampling rate of 144 kHz. For study of overall ten kinds of faults in each region (Fig. 6).

Actual current waveforms are used for discriminating grid connected and idling mode of system. It is seen that grid-based mode consumes maximum price when related to idle mode which indicates that the fault impression is greater at standard grid-based mode. The wavelet multi-resolution with sum of the thorough measurements procedure is able to detect the fault in less than 12 ms which can be seen in Fig. 8a, whereas as in the case of conventional approach, the period to detect the fault is more than 35 ms which is shown in Fig. 7. So it can be concluded that the developed

Table 1 System technical parameters of the proposed system

Location	Source	Load	Devices connected
Bus 1	Main Grid: 500 MVA 132 kV		Transformer: 100 MVA 132 kV/ 34.5 kV
Bus 2		3.6 MVA, UPF	
Bus 3	PV source: 5 MW 260 V	3.0 MVA, 0.85PF	Transformer: 20 MVA 260/34.5 kV
Bus 4	Wind source: 10 MW 575 V		Transformer: 20 MVA 575 V/34.5 kV
Bus 5		4.5 MVA, 0.85 PF	
Bus 6		3.0 MVA, UPF	
Bus 7	Wind source: 15 MW,575 V		
Line Parameters	c—R $= 0.1153$ Ω/KmR0 $= 0.4132$Ω/Km L $= 1.0524$mH/Km L0 $= 3.3251$mH/Km C $= 11.332$nF /Km C0 $= 5.0124$nF/Km		
Mother wavelet	Bior 1.5		
Sampling frequency	144 kHz		
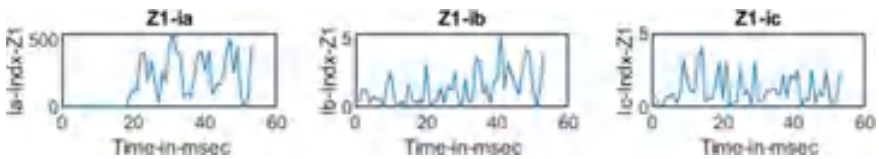
Fig. 6 Current waveform showing the fault in Phase A in region-1 indicating LG fault under grid-based mode

algorithm requires smaller time when likened to standard unoriginal process that can be recognized from Figs. 7 and 8a. The effect of fault inception angle is witnessed in Fig. 8b.

The numerous values are represented as indices of wavelet sum of detailed coefficients at various zones under dissimilar fault inception angle as illustrated in Tables 2 and 3. The tabular data advisable information regarding system behaviour by comparing grid connected, idle mode as well as healthy condition.
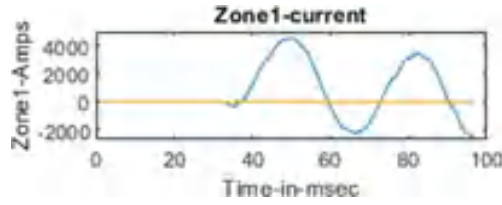
Fig. 7 Current response showing fault with conventional method of protection

(a)   (b)
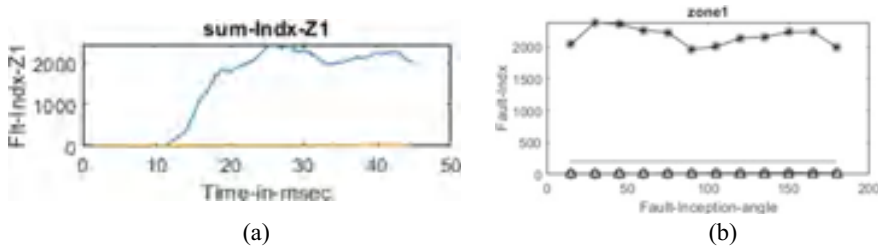
Fig. 8 **a** Sum of detailed coefficients; **b** fault index

Table 2 Comparative analysis of fault indices at zone 1

| | Zone1:1φ\phi G -BG Fault | | | | | | System Healthy condition | | |
| | Grid Connected Mode | | | Idle mode | | | Zone 1: Indices | | |
FIA	Index-A	Index-B	Index-C	Index-A	Index-B	Index-C	Index-A	Index-B	Index-C
0	10.623	2253.559	13.604	5.885	1600.134	8.098	16.693	10.267	17.797
15	11.956	2419.511	16.128	5.466	1486.975	8.681	15.454	12.951	22.168
30	8.893	2359.575	14.762	3.539	1535.319	10.044	6.408	12.064	16.568
45	8.628	2302.373	16.220	3.806	1546.218	9.603	7.974	12.751	17.106
60	9.308	2200.464	13.333	4.536	1536.247	10.180	6.441	12.735	16.477
75	7.686	2152.468	6.946	5.026	1597.545	8.113	9.270	10.973	12.732
90	9.767	2209.175	10.213	5.570	1512.698	7.276	13.317	10.329	12.951
105	12.004	2142.591	7.857	5.724	1520.968	7.209	11.670	10.842	13.318
120	10.022	2271.959	9.577	9.463	1583.546	7.162	14.597	8.630	12.752
135	11.907	2243.373	10.807	11.041	1721.713	6.857	14.450	8.314	12.722
150	9.752	2234.257	8.549	10.257	1729.520	6.954	11.393	12.228	14.817
165	6.875	1707.687	10.025	10.591	1202.963	7.161	11.894	12.848	15.376

Table 3 Comparative analysis of fault indices at zone 5

| | Zone5:1φ G-CAG Fault | | | | | | System healthy condition | | |
| | Grid Connected Mode | | | Idle mode | | | Zone 5: Indices | | |
FIA	Index-A	Index-B	Index-C	Index-A	Index-B	Index-C	Index-A	Index-B	Index-C
0	2789.441	4.521	2398.622	1448.560	5.594	1494.021	4.677	2.220	4.204
15	2805.220	5.271	2310.586	1987.190	7.015	2028.163	4.853	2.041	4.888
30	2517.177	5.121	1889.851	1881.483	7.544	2063.045	3.079	1.845	3.757
45	2441.731	6.253	1815.252	2100.788	10.162	2087.733	4.070	2.037	4.218
60	1935.176	4.453	1704.503	1746.940	10.100	2082.398	3.402	1.587	3.454
75	1269.698	5.323	1081.969	1343.782	10.596	1808.301	3.369	1.648	3.366
90	1283.634	5.963	1222.558	1421.875	11.241	1605.071	3.777	1.991	3.294
105	1375.280	5.794	1685.660	1830.269	10.458	1822.920	2.912	1.946	2.967
120	1226.926	6.112	1688.721	2154.156	10.308	2086.315	3.497	2.102	3.465
135	1482.406	4.782	1981.285	1793.343	8.933	1797.099	3.424	2.084	3.089
150	1718.513	6.784	2223.925	1736.364	7.774	1518.598	3.044	3.215	3.961
165	1477.317	7.209	1830.613	1584.244	7.614	1453.133	3.128	3.781	4.781

7 Conclusion

This broadsheet suggests safety arrangement of PV-wind incorporated network for power transmission with resource of IoT and wavelet primarily based set of rules, in order that it is able to provide strong safety scheme. IoT can crack problems related to machine pushed and bodily situations in real manner and also thereby promotes the plan of new technique for protecting the grid. WT is a distinguished research technique which scrutinizes the faults in temporary indicators at numerous frequencies by means of decaying the signals into measurements of bior1.5 mother wavelet, which offers extra important facts with recognition to the form of fault and position. The advised set of rules is verified and supplied intended for the recognition and differentiation of faults beneath numerous types of faults at diverse fault inception angles through wavelet multi-decision study with bios1.five mother wavelets targeted measurements.

References

1. IEEE Std C37.113-2015 (2015) IEEE guide for protective relay applications to transmission lines, pp 1–141
2. Dhanalakshmi KV, Ravi Kumar G, Subrahmanyam JBV (2021) IoT based protection of microgrid with grid-connected and islanded mode using wavelet approach. In: 2021 IEEE 4th International conference on computing, power and communication technologies (GUCON). 978-1-7281-9951-1/21. https://doi.org/10.1109/GUCON50781.2021.9573737

3. Aziz MMA, Zobaa AF, Ibrahim DK, Awad MM (2008) Transmission lines differential protection based on the energy conservation law. Electr Power Syst Res 78(11):1865–1872
4. Unde SV, Dambhare SS (2016) Differential protection of mutually coupled lines in modal domain using synchronized measurements. In: 2016 National power systems conference (NPSC). IEEE, pp 1–5
5. Lien K-Y, Bui DM, Chen S-L, Zhao W-X, Chang Y-R, Lee Y-D et al (2016) A novel fault protection system using communication-assisted digital relays for AC microgrids having a multiple grounding system. Int J Electr Power Energy Syst 78:600–625
6. Motlagh NS, Mohammadrezaei M, Hunt J, Zakeri B (2020) Internet of things (IoT) and the energy sector. Energies 13:494
7. Amin SM (2011) Smart grid: overview, issues and opportunities. Advances and challenges in sensing, modelling, simulation, optimization and control. Eur J Control 5–6:547–567
8. Hidayatullah NA, Kurniawan AC, Kalam A (2018) Power transmission and distribution monitoring using internet of things (IoT) for smart grid. IOP Conf Ser: Mater Sci Eng 384:012039
9. Habib HF, Lashway CR, Mohammed OA (2017) On the adaptive protection of microgrids: a review on how to mitigate cyber-attacks and communication failures. In: 2017 IEEE industry applications society annual meeting, 1–5 Oct 2017, pp 1–8
10. El-Hawary ME (2014) The smart grid-state-of-the-art and future trends. Electr Power Compon Syst 42(3–4):239–250
11. Ma R, Chen H-H, Huang Y-R, Meng W (2013) Smart grid communication: challenges and opportunities. IEEE Trans Smart Grid 4(1)
12. Fang X, Mishra S, Xue G, Yang D (2012) Smart grid—the new and improved power grid: a survey. IEEE Commun Surv Tutor 14(4)
13. Hooshyar A, Iravani R (2017) Microgrid protection. Proc IEEE 105(7):1332–1353
14. Memon AA, Kauhaniemi K (2015) A critical review of AC microgrid protection issues and available solutions. Electr Power Syst Res 129:23–31
15. Hessine MB, Jouini H, Chebbi S (2014) Fault detection and classification approaches in transmission lines using artificial neural networks. In: 17th IEEE Mediterranean electro technical conference MELECON 2014, pp 515–519
16. Sharafi A, Sanaye-Pasand M, Jafarian P (2011) Ultra-high-speed protection of parallel transmission lines using current travelling waves. IET Gen Trans Dist 5:656–666
17. Goli RK, Sree YM, Priya GV (2022) Neuro-wavelet protection scheme for PV-wind energy source integrated 6-terminal power network with IoT-based smart environment. J Inst Eng (India): Series B 103(6):2035–2048
18. AbdEl-Gawad H, Sood VK (2014) Overview of connection topologies for grid-connected, PV systems. IEEE, Toronto, pp 1–8
19. John AH (2005) Derivation of symmetrical component theory and symmetrical component networks. In: Georgia tech protective relaying conference, Atlanta, GA
20. Tourou P, Sourkounis C (2014) DFIG-based wind energy conversion systems under unbalanced voltage dips. In: 2014 International symposium on power electronics, electrical drives, automation and motion. https://doi.org/10.1109/speedam.2014.6872096
21. JayaBharata Reddy M, Venkata Rajesh D, Mohanta DK (2013) Robust transmission line fault classification using wavelet multiresolution analysis. Comput Electr Eng 39(4):1219–1247
22. Shekar SC, Ravi Kumar G, Lalitha SVNL (2019) A transient current based micro-grid connected power system protection scheme using wavelet approach. Int J Electr Comput Eng 9(1):14

Early Prediction of Healthcare Diseases Using Machine Learning and Deep Learning Techniques

O. Obulesu , N. Venkateswarulu , M. Sri Vidya, S. Manasa, K. Pranavi, and Ch. Brahmani

Abstract Making informed decisions and precise predictions is made possible by machine learning (ML). In order to detect and predict diseases including heart attacks, diabetes, breast cancer, chronic kidney disease, and COVID-19 in humans using numerous risk factors, classification models like Logistic Regression, Random Forest Classifier, Support Vector Machine, and Decision Tree Classifier are used. The datasets are categorized according to medical characteristics, and machine learning algorithms are utilized to process them. With the aid of conventional machine learning techniques, correlations between the various variables included in the dataset are discovered, and these correlations are then effectively utilized in the prediction of diseases. Using the patient's medical history, they can determine if the patient is likely to be diagnosed with a specific disease or not and anticipate the patient's health condition using training from natural events. The outcome of this prediction is whether the patient is likely to be diagnosed with any of the diseases mentioned above.

Keywords Diseases · Health · Machine learning · Prediction

1 Introduction

Heart attack has drawn a lot of interest in the field of medical research among various other life-threatening illnesses. Typically, face-to-face examination of patient, signs, and symptoms are used to make the diagnosis of a heart attack. The danger of heart disease is governed by a number of inconstant parameters, including smoking,

O. Obulesu · N. Venkateswarulu (✉) · M. Sri Vidya · S. Manasa · K. Pranavi · Ch. Brahmani
Department of Computer Science and Engineering, G. Narayanamma Institute
of Technology & Science, Hyderabad, Telangana, India
e-mail: venkateswarulu@gnits.ac.in

O. Obulesu
e-mail: obulesh194@gmail.com

cholesterol, hereditary and genetic factors of the disease, obesity, high blood pressure, and sluggishness.

The purveying of high-quality medical assistance at a reasonable amount is a notable problem for healthcare corporation like hospitals and clinics. High-quality medical assistance entails precise patient care and to deliver the successful therapies. The database on heart disease contains both numerical and category information. Before undergoing further processing, these records are cleaned and filtered to pull out any unnecessary data from the database. The recommended method may extract specific hidden information, such as patterns and connections associated to heart disease, from a historical database of heart illnesses. Results indicated that the suggested system has a special ability to accomplish the specified mining goals.

Background Study

Creating systems that can find out from experience and make predictions as a result is the main goal. A form of software called data mining helps computers to create and organize many qualities. For the purpose of predicting the aforementioned diseases, the prediction model uses categorization algorithms. The presentation of connected topics, such as machine learning and its methods, which are described together with brief definitions, data preprocessing, assessment metrics, and information about the dataset employed.

1.1 Problem Statement

Smallpox and polio have been eradicated in India because of its healthcare system, but other terrible diseases still affect 60% of the country's population every year. Breast cancer, heart attacks, lung illnesses, and COVID-19 have all recently been added to the list. Low- and middle-income nations frequently have a high death rate [1]. Numerous risky behaviors contribute to the increased risk of heart attacks, lung conditions, high cholesterol, obesity, hypertension, etc.

The conditions indicated above can be effectively managed by combining a change in lifestyle, medication, and, in some circumstances, surgery. With the proper care, symptoms can be lessened and the important organs' performance can be enhanced. The model's anticipated outcomes can be used to avoid costly operations like surgery and other invasive treatments altogether. The ultimate goal is to design an accurate machine learning model that uses data from primary medical tests to forecast if a user is likely to experience any of the five diseases listed above in the foreseeable future. It is possible to use many more input attributes to estimate the risk, but our objective is to predict with fewer attributes and more quickly.

1.2   Advantages and Drawbacks

Advantages

The suggested solution will combine computerized patient records with clinical decision assistance (dataset). This will improve patient safety, reduce unwelcome practice variance, reduce medical errors, and improve patient outcomes. This strategy has potential since data analysis and modeling tools, including data mining, can produce a knowledge-rich domain that can significantly improve the quality of healthcare options. It is increasingly essential to use data mining and machine learning approaches to help prediction and decision-making in the healthcare industry because the medical data domain comprises a significant number of records. As a result, business intelligence, which is useful for disease detection and prognosis, is influenced by medical data mining.

Drawbacks

Medical diagnosis is viewed as an essential yet challenging process that demands accuracy and efficiency. Clinical decisions are frequently made utilizing the doctor's experience and intuition rather than the fact-rich knowledge hidden in the database. This practice has an effect on the quality of treatment provided to patients by causing unintentional biases, errors, and astronomical medical expenses. The same might be automated, which would be quite advantageous. A knowledge-rich environment created by data mining has the potential to considerably raise the standard of therapeutic judgments.

1.3   Proposed System

In this, a system prototype is proposed that uses a binary classification model to forecast a person's risk factor based on his or her medical data. The system comes with a well-rounded graphic user interface that is simple to use and comprehend. The datasets used for the classification were obtained from Kaggle and were used in a supervised learning process. The datasets that were initially acquired as unstructured data in the form of medical reports and then transformed into structured datasets are the only ones covered by this reference. The datasets show an issue of binary classification. The use of machine learning algorithms like support vector machine, decision tree classifier, and K-neighbor classifier. The model uses the algorithm with the maximum accuracy. In the end, a user-friendly web interface was created to entirely abstract the system's essential functionality and implementation details, allowing anyone with little to no technical understanding to utilize it.

2   Literature Survey

This section discusses the number of researchers who have worked for disease diagnosis by applying various machine learning algorithms. They have acknowledged the effectiveness of machine learning algorithms for the diagnosis of various diseases.

In earlier research, methods for categorizing and predicting the aforementioned disorders were explored. These research, however, are more concerned with the unique effects of particular machine learning approaches than with their optimization through the use of optimal algorithms. Few researchers also try to apply hybrid optimization techniques for improved machine learning classification. On several medical data sets, experiments are carried out utilizing various classifiers and feature selection methods.

2.1   Heart Attack

Otooam et al. [2] introduced a modern framework for tracking and research purposes. This idea monitors and identifies coronary artery disease. UCI is taken from the Cleveland cardiac data collection. There are 304 examples in this dataset, along with 77 traits or attributes. Fourteen traits out of 76 attributes are utilized. Two tests using three algorithms—Bayes-Net, SVM, and FT—are run for detection purposes. The best FT.7 attributes are gathered using the Best First Selected algorithm, and cross-validation checks are employed for evaluation. By applying the test to the seven best features, Bayes-Net achieves 84.5% accuracy, SVM gives 85.1% accuracy, and FT correctly classifies 84.6%. Vembandasamy et al. [3] proposed a research which was employed using the Naïve-Bayes algorithm to note down the heart diseases. Bayes' theorem is contained in Naive Bayes. As a result, there is a significant assumption of freedom for the Naive Bayes.

2.2   Diabetes Disease

Iyer et al. [4] is using decisions trees and Naïve-Bayes, they experimented a job to predict diabetes disease. When insulin is produced insufficiently or is used excessively, diseases develop. The dataset used in this study is the Pima India diabetes dataset. In this data collection, the percentage division (71:31) predicts more accurately than cross-verification. J48 indicates 74.8698% and 76.9565% precision using cross-verification and percent splitting, respectively. Naive Bayes offers 79.5653% precise results while employing PS. A diabetes disease diagnosis method has been initially developed by Ephzibah et al. [5]. The suggested model combines fuzzy logic with GA. This is done to choose the best feature subsets and to increase classification accuracy. MLUCI Lab, which has 768 instances and 7 characteristics, compiles a dataset for analysis.

2.3 Breast Cancer

The most commonly occurring type of cancer is breast cancer. Over two million women are thought to be impacted annually. Five-year survival for breast cancer among women diagnosed between 2010 and 2014 varies greatly with changes in location. In most regions, it is generally recognized to be greater than fifty percent. The National Breast Cancer Foundation (NBCF) advises that women over the age of forty should obtain a mammography once a year.

Relevance Vector Machine [6] for the identification of breast cancer used the Wisconsin original dataset, that has 699 occurrences and 11 attributes, to achieve an accuracy of 97%, while [7] used the weighted Naive Bayes approach and assigned varying weights to various attributes based on their predictive power to achieve an accuracy of 92%. Sivakami [8] created a hybrid classifier in WEKA using SVMs and decision trees, and they achieved a 91% accuracy rate. [9] Fuzzy inference methods were employed on the dataset called Mamdani Fuzzy inference model and achieved 93% accuracy. This publication [10] provides various differences between numerous strategies employing Bayes Network, kNN algorithm, and Pruned Tree techniques on a breast cancer dataset with a total of 6291 points of data and a dimension of 9 attributes in 9 columns and 699 rows of data values.

2.4 Chronic Kidney Disease

A global public health issue, chronic kidney disease (CKD) affects 10% of the world's population [11, 12]. In China, the prevalence of CKD is 10.8% [13], but in the USA, it ranges from 10 to 15% [14]. In the overall adult population of Mexico, CKD affects that number has reportedly reached 14.7% [15]. Early on, CKD does not have noticeable symptoms. Because of this, the illness might not be discovered until the kidney has lost around 25% of its functionality [16]. Additionally, CKD has a significant global influence on the human body with substantial morbidity and mortality [17]. Cardiovascular disease may develop as a result [18, 19]. CKD is a pathologic illness that progresses and cannot be reversed [20]. Therefore, it is crucial to identify and treat CKD as soon as it manifests in order for patients to begin treatment and reduce the disease's progression.

2.5 COVID-19

Due to the illness, this virus has caused, including illness that has resulted in death and persistent person-to-person propagation in several countries, and the flare-up of novel coronavirus (COVID-19) infections has sparked worry globally [21, 22]. The Middle

East respiratory disease (MERS)-CoV, severe acute respiratory syndrome (SARS)-CoV, and the newly discovered SARS-CoV-2 virus are only a few of the many viruses that make up the CoV family. Al-Turaiki and his team claim that MERS-CoV symptoms include coughing, fever, nose congestion, shortening of breath, and occasionally diarrhea. Unfortunately, there is little knowledge about how the epidemic move of virus and the people getting impacted. Fifteen years after the SARS-CoV outbreak produced by the first highly pathogenic human CoV, another diarrhea syndrome considered acute and severe-CoV decimated animal productivity by infecting pigs with deadly infections. The subsequent illness was designated "COVID-19" by the World Health Organization on February 11, 2020 [21].

3 Methodology

The primary goal is to identify and diagnose diseases at an early stage and at a reasonable price. Data mining techniques allow us to identify diseases early on. With the right diagnosis, the condition can be totally cured. The healthcare business gathers a lot of data, but it is not mined to find confidential information. A technique for resolving this problem is data mining. It is an approach technique. Large amounts of data are examined in order to identify patterns that can be transformed into informative data. A standardized form must be used to collect the data. The datasets under consideration for this prediction include a variety of features, one of which is a binary answer variable. The datasets therefore show a binary categorization issue. Multiple machine learning methods are applied to the dataset in order to fulfill the goal of disease prediction, and dataset analysis is addressed. Additionally, it demonstrates that some traits are more crucial than others for higher-precision prediction. Eliminating numerous trials could result in financial savings due to the potential that not all of a patient's features will have a major impact on the outcome.

Objectives

The major aim is to use machine learning methods to create a model that can predict a person's likelihood of developing any of the five diseases indicated above in the near future.

Below, specific goals are listed.

a. To create a model that analyzes a patient's medical history to determine whether or not they are likely to be diagnosed with a certain disease.
b. To create and choose datasets that are processed using machine learning algorithms and are categorized according to medical parameters.
c. To effectively forecast diseases, correlations between various dataset properties can be discovered using common machine learning techniques.
d. To create a single GUI that incorporates all 5 diseases.

3.1 Proposed System

Training Data

Datasets can be gathered from a variety of public resources, including kaggle.com, data.gov, the UCI machine learning repository, etc**.**

Train the ML Algorithm

How well machine learning algorithms perform when used to label the predictions on the test data that was unutilized to train the model is determined by the process of training them. The method is quick and easy, and the outcomes allow us to evaluate the competent machine learning algorithms work when applied to address challenges involving predictive modeling (Fig. 1).

Evaluation

It is crucial to thoroughly verify how well our algorithm performs with fresh data (during training). The three main metrics for assessing a classification algorithm's effectiveness are accuracy, precision, and recall.

Input Data

We require a method of interacting with our model and providing it with problems. Typically, an API or user interface is used for this. It is necessary to develop an HTML page that will provide user input values to the flask application. Once more, it ought to remove the outcome from the flask and present it to the user.
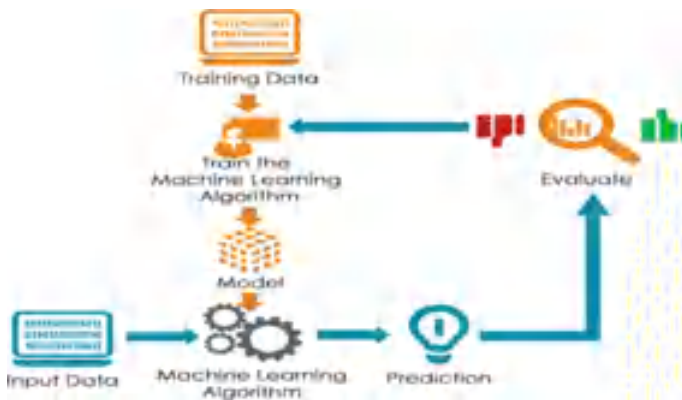
Fig. 1 Architecture of the proposed system

Fig. 2 Flowchart of the
proposed system

Flowchart

See (Fig. 2).

4 Implementation

4.1 Datasets Used

Without data, machine learning is impossible for AI systems to perform. It is the element that enables algorithm training the most. A training dataset is necessary for machine learning. It is the real dataset that was used to train the model to carry out different tasks. Datasets can be gathered from a variety of free resources, including kaggle.com, data.gov, the UCI repository, etc.

Dataset: heart.csv	Dataset: kidney_disease.csv	Dataset: diabetes.csv
Source: Kaggle	Source: Kaggle	Source: Kaggle
Year: 2018	Year: 2019	Year: 2018
Shape: 400 X 158	Shape: 303 X 14	Shape: 768 X 9

Dataset: covid.csv	Dataset: Breast Histopathology dataset
Source: Kaggle	Source: Kaggle
Year: 2021	Year: 2019
Shape: 5005 X 14	Shape: 848 images

4.2 Data Preprocessing

In the field of natural and material sciences, data collection, handling, and processing have shown to be the most dependable means of comprehending and learning new facts, information, and things. Therefore, it has become standard practice in the scientific and industrial sectors to identify the most effective strategies in statistical or bioinformatics circumstances.

Preprocessing includes:

i. Handling the null values.
ii. Handling the categorical values if any.
iii. Normalize the data if required.
iv. Identifying the dependent and independent variables.
v. Split the dataset into train and test sets.

4.3 Analyzing the Dataset

Renaming the Columns

Since the names of our columns do not accurately describe what they are, renaming them is a smart idea to make the columns easier to understand. Data columns will provide all of the column names that are present in your data, and in the next line of code, we are manually setting the column name.

Handling Null Values

The dataset may occasionally have some missing data. It is not always a good idea to remove rows or columns with null values because we could lose important data. Filling in the missing data with comparable values like mean, median, and mode is another technique to handle the missing values.

Use isnull to see how many values are missing from each column (). You can employ sum().

Handling Categorical Data

Label Encoding: Any structured dataset often has numerous columns that combine numerical and category variables. This is known as label encoding. A machine

cannot understand text; only numbers can. Without losing its categorical nature, the categorical data must be transformed into numerical data.

4.4 Splitting the Data

Splitting Data into Independent and Dependent Variables

The dataset's result is the only thing that counts as the dependent variable, while its inputs collectively count as the independent variable. The class is the dependent variable in our dataset, whereas all other columns are independent. However, we will only choose the independent columns that have a good correlation with our dependent column and are useful to it.

Splitting Data into Train and Test Set

The train–test split technique is used to gauge the performance measures of machine learning algorithms when they are used to generate forecasted values on data that were employed to train the model.

   The original data can be tested on when a separate test set is not available. The dataset is divided into sections for testing and the leftover portions are utilized for training. Utilizing the scikit-learn library's train test spilt, the splitting is carried out.

4.5 Model Building

Depending on the type of data being processed—such as images, sounds, text, and numerical values—different machine learning techniques can be applied. The algorithms can be selected based on the goal. In order to develop predictions and accomplish its objective, the machine learning model is built by learning from training data, generalizing that knowledge, and then applying that understanding to present-day data that it has unseen before.

K-Nearest Neighbor

Working of KNN Algorithm:

1. To begin with, we choose a K value for our KNN algorithm.
2. We will now measure the distance. Let us concentrate on the Euclidean distance here. Find the k-neighbors' Euclidean distance.
3. We now determine which of the new points we have provided is closest to our point by checking all of its neighbors. Here, we merely look for k-nearest.
4. Next, we look at the class with the highest score.

We select the maximum number, and then we give that class our new point.
We employ the KNN algorithm in this manner.

Decision Tree

Working of Decision Trees

The choice to use strategic splits has high impact on how accurate a tree is. The selection criteria for classification and regression trees differ. To decide whether or not to divide a node either into two or many sub-nodes, decision trees employ multiple techniques. Their homogeneity increases with the formation of more sub-nodes. In other words, we may claim that as the desired variable is increased, the purity of the node also grows. The decision tree splits the nodes based on all available parameters before choosing the division that produces the most similar sub-nodes.

Let us examine decision tree algorithms, and few of them are: ID3, C4.5, CART, CHAID, and MARS.

The ID3 method uses a top-down greedy search strategy to iteratively explore the universe of potential branches in order to construct decision trees by not going backward. A greedy algorithm always chooses what seems to be the best hit at the time. It uses all of the relevant variables to evaluate the nodes.

Support Vector Machine

Support vector machines, or SVMs, are a tool that may be used for both regression and classification applications. However, classification purposes typically make use of it.

Working of SVM

The support vector machine technique seeks a hyperplane in an N-dimensional space (N is the number of features) that accurately classifies the input points (Fig. 3).

The two groups of data points can be divided using a variety of different hyperplanes. Our objective is to find a plane with the largest margin, or the largest distance between data points from the two classes. Future data points can be classified with greater assurance by increasing the margin distance (Fig. 4).
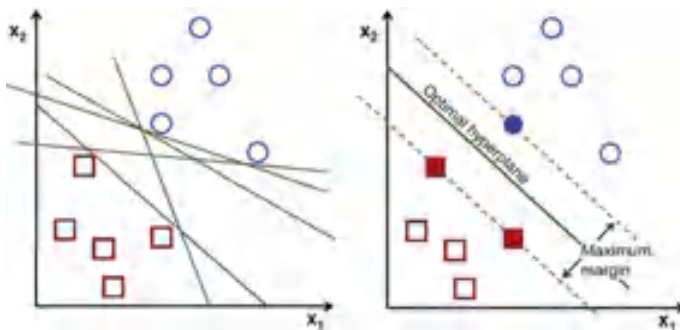
Fig. 3 Possible hyperplanes
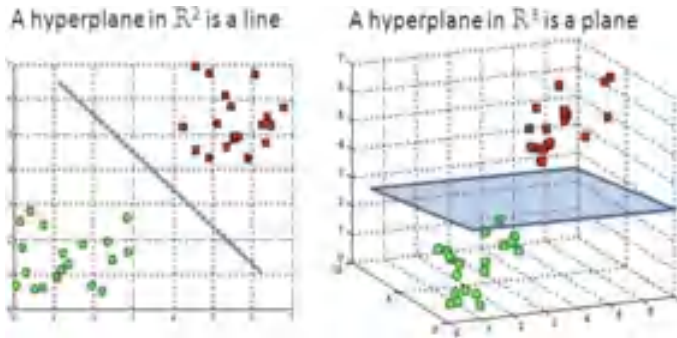
Fig. 4 Hyperplanes in 2D and 3D feature space

Hyperplanes and Support Vectors

Hyperplanes, which serve as decision boundaries, aid in classifying the data points. The data points that are located on either side of the hyperplane can be assigned to various classes. Additionally, the size of the hyperplane is influenced by the number of features. If there are only two input features, the hyperplane effectively looks like a line. If there are three input characteristics, the hyperplane collapses into a two-dimensional plane. When the quantity of features exceeds a certain point, it becomes impossible to imagine.

Convolutional Neural Network

CNNs, a subclass of Deep Neural Networks, are commonly applied to the processing of visual images. CNNs are able to recognize and classify particular aspects from photos. Their applications include image classification, image and video recognition, medical image analysis, and computer vision.

Basic Architecture

A CNN architecture primarily consists of two elements:

- A convolution tool that uses the feature extraction method to separate and classify the various image characteristics for analysis.
- A fully connected layer that predicts the class of the image based on the information gathered in earlier steps and uses the output of the convolution process.

Convolution Layers

Three main types of layers make up the CNN: fully-connected (FC), pooling, and convolutional layers. In addition to these three layers, two other critical factors are the dropout layer and the activation function.

1. Convolutional Layer
   A convolution operation is carried out between input image and an array filter having the dimensions M × M. With this layer as starting layer, different features are being extracted from input photographs.
2. Pooling Layer
   Max pooling—The largest contribution to this max pooling is the mapping of features.
   Average pooling—The calculation of average of components that are in the region of an image with known size is done.
   Sum Pooling—Predetermines the sum of the total elements in this section is calculated.
3. Fully Connected Layer
   Connecting neurons from one layer to the other layer that has weights and biases is what we call fully connected (FC) layer.
4. Dropout
   Training dataset might get overfitting because whole features of FC layer are tightly connected as a rule in general. Usually, a given model outperforms over a training data however, it has a harmful impact on the model's statistics when applied to a new random data.
5. Activation Mechanisms
   CNN model—activation function is the primary ingredient of this model. This function is used to choose and estimate any format of complex continuous connection between network variables. Following includes activation functions which are utilized normally and they are the Sigmoid functions, ReLU, tanH, and Softmax.

5 Results and Discussions

5.1 Comparative Analysis of Results

See (Fig. 5 and Table 1).

```
Epoch 1/10
23/23 [==============================] - 78s 3s/step - loss: 1.3227 - accuracy: 0.1588 - val_loss: 0.9963 - val_accuracy: 0.234
6
Epoch 2/10
23/23 [==============================] - 66s 3s/step - loss: 0.8136 - accuracy: 0.4434 - val_loss: 0.6276 - val_accuracy: 0.654
3
Epoch 3/10
23/23 [==============================] - 66s 3s/step - loss: 0.5780 - accuracy: 0.7017 - val_loss: 0.4121 - val_accuracy: 0.888
9
Epoch 4/10
23/23 [==============================] - 67s 3s/step - loss: 0.4618 - accuracy: 0.8356 - val_loss: 0.4829 - val_accuracy: 0.901
2
Epoch 5/10
23/23 [==============================] - 67s 3s/step - loss: 0.3943 - accuracy: 0.8867 - val_loss: 0.2941 - val_accuracy: 0.925
9
Epoch 6/10
23/23 [==============================] - 67s 3s/step - loss: 0.3551 - accuracy: 0.9019 - val_loss: 0.2971 - val_accuracy: 0.938
3
Epoch 7/10
23/23 [==============================] - 68s 3s/step - loss: 0.3539 - accuracy: 0.8978 - val_loss: 0.2872 - val_accuracy: 0.938
3
Epoch 8/10
23/23 [==============================] - 70s 3s/step - loss: 0.3265 - accuracy: 0.8950 - val_loss: 0.2638 - val_accuracy: 0.938
3
Epoch 9/10
23/23 [==============================] - 70s 3s/step - loss: 0.3093 - accuracy: 0.9019 - val_loss: 0.2383 - val_accuracy: 0.938
3
Epoch 10/10
23/23 [==============================] - 67s 3s/step - loss: 0.3094 - accuracy: 0.9033 - val_loss: 0.2259 - val_accuracy: 0.938
3
```

The Accuracy for Breast Cancer : 93.8

Fig. 5 Model results for breast cancer

Table 1 Results for the classifiers

	Heart attack	Chronic kidney disease	Diabetes	COVID
F1-score	0.82	1	0.84	1
	0.88	1	0.67	1
Precision	0.893	1	0.81	1
	0.83	1	0.72	1
Recall	0.76	1	0.87	1
	0.93	1	0.63	1
Support	0.33	30	100	670
	0.43	10	54	331
Accuracy	85.52	100	79	100

6 Conclusion and Future Enhancements

Multiple diseases can be predicted simultaneously using a multi-disease prediction model. Here, a disease prediction is made based on user input. The user will have the option. If the user wants to predict a specific disease, the associated disease model will be triggered and predicted based on the user's inputs. The benefit of using a

multi-disease prediction model in advance is that it can estimate the likelihood that numerous diseases will manifest themselves while also lowering the mortality rate.

This study can be expanded by including additional machine learning methods. Given the limits of this study, it is necessary to use more intricate models in combination in order to improve early prediction accuracy. Increasing the dataset size and a number of other enhancements can lead to more successful results. Machine learning and diverse different optimization strategies also can be used to similarly enhance the assessment results. For the convenience of patients and clinicians, there are more options for standardizing the data, comparing the results, and figuring out how to mix multi disease trained machine learning and deep learning models with particular multimedia.

References

1. World Health Organization (2020) Cardiovascular Diseases, WHO, Geneva, Switzerland. https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1
2. Otoom et al. (2015) Effective diagnosis and monitoring of heart diseases. Int J Softw Eng Appl 9:143–156
3. Vembandasamy et al. (2015) Heart disease detection using Naive Bayes algorithms. IJISET-Int J Innov Sc Eng Technol 2:441–444
4. Iyer A, Jeyalatha S, Sumbaly R (2015) Diagnosis of diabetes using classification mining technique. Int J Data Mining Knowl Manag Process (IJDKP) 5:1–14
5. Ephzibah EP (2011) Cost effective approach on feature selection using genetic algorithm and fuzzy logics for diabetes diagnosis. Int J Soft Comput (IJSC) 2:110. https://doi.org/10.5121/ijsc.2011.2101
6. Gayathri BM, Sumathi CP (2016) Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer
7. Kharya S, Soni S (2016) Weighted Naïve Bayes classifier—predictive model for breast cancer detection
8. Sivakami (2015) Mining big data: breast cancer prediction using DT-SVM hybrid model
9. Gayathri BM, Sumathi CP (2015) Mamdani fuzzy inference system for breast cancer risk detection
10. Mohd F, Thomas M (2007) Comparison of different classification techniques using WEKA for Breast cancer
11. Chen Z et al (2016) Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers. Chemometr Intell Lab 153:140–145
12. Zhang L et al (2012) Prevalence of chronic kidney disease in china: a crosssectional survey. Lancet 379:815–822
13. Polat H, Mehr HD, Cetin A (2017) Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. J Med Syst 41(4)
14. Papademetriou V et al (2017) Chronic kidney disease, basal insulin glargine, and health outcomes in people with dysglycemia: the origin study. Am J Med 130(12)
15. Hill NR et al (2016) Global prevalence of chronic kidney disease—a systematic review and meta-analysis. Plos One 11(7)
16. Hossain MM et al (2019) Mechanical anisotropy assessment in kidney cortex using ARFI peak displacement: preclinical validation and pilot in vivo clinical results in kidney allografts. IEEE Trans Ultrason Ferr 66(3):551–562
17. Alloghani M et al (2018) Applications of machine learning techniques for software engineering learning and early prediction of students' performance. Proc Int Conf Soft Comput Data Sci 246–258

18. Du L et al (2018) A machine learning based approach to identify protected health information in Chinese clinical text. Int J Med Inform 116:24–32
19. Abbas R et al (2018) Classification of foetal distress and hypoxia using machine learning approaches. Proc Int Conf Intell Comput 767–776
20. Mahyoub M, Randles M, Baker T, Yang P (2018) Comparison analysis of machine learning algorithms to rank alzheimer's disease risk factors by importance. In: Proceedings 11th international conference developments in systems engineering
21. Yu KH, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. Nat Biomed Eng 2(10):719–731
22. AlMoammar A, AlHenaki L, Kurdi H (2018) Selecting accurate classifier models for a MERS-CoV dataset. Adv Intell Syst Comput 868:1070–1084

PTS with Phase Factor-Based Reptile Search Algorithm and Hybrid Coding Approach for PAPR and BER Reduction in MIMO-OFDM

G. Krishna Reddy [ID] **and G. Merlin Sheeba** [ID]

Abstract Multiple input multiple output (MIMO) antennas are combined with the orthogonal frequency division multiplexing (OFDM), because of its potential decrease on interference and efficient data transmission. High peak to average power ratio (PAPR) and bit error rate (BER) are considered as important issue in MIMO-OFDM. Therefore, an efficient error correction code and partial transmit sequence (PTS) are required to be developed for improving the system performances. In this paper, the PTS with phase factor-based reptile search algorithm (PFRSA) is proposed to identify the optimal phase factor values for minimizing the PAPR. Moreover, the hybrid coding approach (HCA) includes polar coding and binary quasi cyclic low-density parity-check (QC-LDPC) code is developed for minimizing the inter-symbol interference (ISI) and inter-carrier interference (ICI) occurred in the communication. The minimization of PAPR and BER was achieved by using the turbo coding (TC) and enhanced switching differential (ESD)-based PTS. Therefore, the proposed HCA-PFRSA-PTS minimizes both the PAPR and BER in the MIMO-OFDM and hence increases its performance. The performance of the HCA-PFRSA-PTS is analyzed in terms of PAPR and BER. The MIMO-OFDM system with HCA-PFRSA-PTS is compared with OFDM, turbo coding enhanced switching differential-based PTS (TC-ESD-PTS), selective mapping PTS with artificial bee colony (SLM-PTS-ABC) and ant colony optimal PTS (ACO-PTS). The BER of the HCA-PFRSA-PTS for SNR 5 dB with 128 subcarriers is 0.0058 which is less when compared to the original OFDM, TC-ESD-PTS, SLM-PTS-ABC and ACO-PTS. The PAPR of the HCA-PFRSA-PTS with 128 subcarriers is 3.65 which is less when compared to the original OFDM, TC-ESD-PTS, SLM-PTS-ABC and ACO-PTS.

G. Krishna Reddy (✉) · G. Merlin Sheeba
Sathyabama Institute of Science and Technology, Chennai, India
e-mail: gkr999gkr@gmail.com

G. Krishna Reddy
Associate professor, ETE, G.Narayanamma Institute of Technology and Science, Hyderabad, India

G. Merlin Sheeba
Professor, ECE, Jerusalem College of Engineering, Chennai, India

1 Introduction

A multiple independent waveforms are broadcasted by using the numerous transmit antennas of MIMO which is different than the typical phased-array radar where it is deliberated as a single input single output radar [1]. The inter-symbol interference (ISI) becomes the issue, when the MIMO is operated in wideband channels. Therefore, the OFDM is integrated with the MIMO for improving the capacity and obtaining the elimination of ISI [2, 3]. Since, the OFDM is broadly used approach because of its high data rate transmission and robustness against the frequency selective fading channel, the integration of MIMO with OFDM is the alternative method for obtaining the high data throughput and spectrum efficiency in broadband wireless communication system without doing any increase in the bandwidth [4–6]. The important merits of the MIMO-OFDM than the single carrier system comprise of small nonlinear distortion, uniform average spectral density, ability of handling sturdy echoes, robustness to channel fading, resistance against impulse interference and utilization of smaller guard intervals [7, 8]. This MIMO-OFDM is ensured as a standard air interface approach in various real wireless communication systems such as Wi-Fi, LTE-A and Wi-Fi [9].

In an each transmit antenna of the MIMO, the inverse fast Fourier transform (IFFT) is resulted the time domain OFDM signal. This OFDM signal has sum of independent modulated data with orthogonal subcarrier frequencies which causes the huge peak amplitude at some time in MIMO-OFDM signal. Moreover, the obtained peak is larger than the average value that returns in higher PAPR. Accordingly, the nonlinear distortion is increased in the high power amplifiers because of the higher PAPR [4]. Therefore, the higher PAPR caused the degradation in BER, inter-modulation interference and so on [10, 11]. In general, PTS is an effective method to decrease the PAPR effect. However, the identification of optimal phase vector increases the system complexity.

Hence, the suboptimal search phase is designed to lessen the complexity. Specifically, the natural-inspired algorithms are considered as an effective approach for identifying the optimal solution, and the optimal phase vectors are generated in the search space [12]. Generally, PTS has an effective capacity over PAPR minimization without affecting the number of subcarriers [13]. Moreover, the error correction code is considered as one of the important approaches, because it minimizes the reception errors during communication which ensures high-quality communications [14].

The major contribution of the paper concise as follows:

- The PFRSA-based PTS is proposed for identifying the optimal phase factors to achieve the less PAPR. Since, the conventional reptile search algorithm (RSA)

is selected for phase factor optimization due to its effective equilibrium between exploitation and exploration.

- For minimizing the errors, i.e., ISI and ICI occurred in during the communication, the HCA is developed in MIMO-OFDM. Accordingly, the BER of MIMO-OFDM is reduced as well as BER is minimized by decreasing the PAPR.

The remaining of the paper arranged as follows: Sect. 2 provides the related work about the PTS and coding technique in MIMO-OFDM. The detailed explanation of the HCA-PFRSA-PTS is given in Sect. 3. The outcome of the HCA-PFRSA-PTS is provided in Sect. 4. Further, the Sect. 5 presents the conclusion.

2 Related Work

Agarwal and Mehta [15] investigated the possibility of combining the polar codes (PC) with convolutional codes (CC) to accomplish the error correction in the data communication. The developed PC-CC was examined under different modulation schemes such as quadrature amplitude modulation (QAM)-16, QAM-64, quadrature phase shift keying (QPSK) and binary phase shift keying (BPSK). This PC-CC was used to lower the inference over the MIMO-OFDM. However, this work was not concentrated over the PAPR reduction in MIMO-OFDM. Hadj Ali and Hamza [16] presented the teaching–learning-based optimization (TLBO) for selective mapping (SLM) to identify an appropriate phase factor for OFDM. The TLBO was used in SLM due to its less amount of calculations as well as it does not have any specific parameter condition. Therefore, this TLBO-SLM was achieved significant PAPR minimization over OFDM. Extra calculations were required for PAPR reduction when there was an increment in the amount of phase factors.

Rakshit et al. [17] developed the adaptive update lifting (AUL)-based wavelet transform for minimizing the side lobes in MIMO-OFDM. Next, the minimization of PAPR and BER was achieved by using the turbo coding (TC) and enhanced switching differential (ESD)-based PTS. Here, an effective error correction and identification of phase factor were supported by using the turbo decoder. However, the overall performances of TC-ESD-PTS were varied according to the number of subcarriers. Sidiq [18] presented the phase optimization approach by using the hybrid SLM-PTS with the artificial bee colony (ABC) algorithm. The modulation technique used in this filter-bank-multicarrier scheme was offset quadrature amplitude modulation. Here, the ABC algorithm was used to obtain the optimum phase patterns with lesser complexity. Hosseinzadeh Aghdam [19] developed the improved ACO-based PTS for minimizing the PAPR over OFDM systems. Here, the improved ACO was explored the values of phase rotation factors to identify optimal phase factor. These PTS designs were mainly concentrated only on the PAPR, but it failed to concentrate over the error correction which resulted in errors in the output.

3   HCA-PFRSA-PTS Method

In this HCA-PFRSA-PTS, the hybrid coding and PTS using PFRSA are accomplished to improve the performances of MIMO-OFDM.

The combination of polar and binary QC-LDPC codes used in the HCA is used to minimize the errors occurred during the data transmission. Moreover, the large PAPR occurred in the OFDM signal is minimized by selecting the optimal phase factor from the PFRSA. The architecture of MIMO-OFDM using HCA and PTS using PFRSA is shown in Fig. 1.

3.1   PAPR of MIMO-OFDM

The signal broadcasting via the OFDM-based transceiver divides the spectrum channel to the orthogonal subchannels. Here, an each subchannel autonomously controls its own data by using the individual subcarrier according to the bandwidth, whereas OFDM signal is the addition of all independent subcarriers. The binary sequences, i.e., input data, are mapped into the symbols by using a QAM at the transmitter.

Next, the N amount of symbols $X = \begin{bmatrix} X_0, X_1, \ldots, X_{N-1} \end{bmatrix}^T$ are included in the IFFT for modulating the signal in time domain $x = \begin{bmatrix} x_0, x_1, \ldots, x_{N-1} \end{bmatrix}^T$.

Equation (1) shows the OFDM signal in the discrete time domain along with the oversampled factor L.

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{(j2\pi nk/LN)}, 0 \le n \le LN - 1 \tag{1}$$

Where an amount of subcarriers is denoted as N and the complex signal n broadcasted using the subcarrier k is denoted as X_k.

Eq. (1) shows that the time domain OFDM signal has N amount of independently modulated and orthogonal subcarriers with huge peak values. The PAPR of the OFDM signal at discrete time is determined as the proportion among the maximum

Fig. 1 Architecture of MIMO-OFDM using HCA-PFRSA-PTS

power and the average power of the complex OFDM signal which is expressed in Eq. (2).

$$\text{PAPR}\{x[n]\} = \frac{\max\{|x[n]|^2\}}{E\{|x[n]|^2\}}, 0 \le n \le LN - 1 \tag{2}$$

Where the average power is denoted as $E\{.\}$ Eq. (3) specifies the PAPR of an each antenna of MIMO-OFDM.

$$\text{PAPR}(x_i) = \frac{\max\{|x_i[n]|^2\}}{E\{|x_i[n]|^2\}}, 0 \le n \le LN - 1 \tag{3}$$

Where the amount of transmitted antennas is specified as $i = 1, 2, \ldots, N_T$. The time domain signal of an each transmitting antenna is expressed in Eq. (4).

$$x_i[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k^i e^{(j2\pi nk/LN)} \tag{4}$$

Hence, the derived expression for the PAPR of MIMO-OFDM system is denoted as shown in Eq. (5).

$$\text{PAPR}_{\text{MIMO–OFDM}} = \max\{\text{PAPR}(x_i)\}, i = 1, \ldots, N_T \tag{5}$$

The peak to average power ratio (PAPR) of signal is compared by using the complementary cumulative distribution function (CCDF). The CCDF is the probability of PAPR that is higher than the threshold level PAPR_0. This condition is expressed in Eq. (6).

$$\text{CCDF} = \text{Probability}(\text{PAPR} > \text{PAPR}_0) \tag{6}$$

For huge number of subcarriers (i.e., $N > 10$), the OFDM signal in time domain is approximately zero-mean Gaussian distributed by considering the central limit theorem. Accordingly, the distribution of power is a central chi-square distribution with 2 degrees of freedom; therefore, the CCDF is rewritten as shown in Eq. (7).

$$\text{CCDF} = (1 - (1 - \exp(-\text{PAPR}_0))^{NL} \tag{7}$$

3.2 Encoding Using Hybrid Coding Approach (HCA)

In transmitter side, the hybrid encoding, i.e., the combination of polar encoding and QC-LDPC encoding, is accomplished for improving the PAPR performances. The HCA-based encoding process is explained as follows:

The $(N : K : \mathcal{A})$ is used to characterize the polar codes that used to obtain the channel capacity though the channel polarization. The code length $N = 2^m$ is assigned without generality loss, where m is an integer. Polar codes are used to achieve the best channel capacity by two different ways: (1) information bits transmission over noiseless subchannels and (2) frozen bits transmission that are known by both the transmitter and receiver. Therefore, the development of polar code is equal to identify K most effective subchannels in that data bits are broadcasted and the location of subchannel is denoted by $\mathcal{A}$. Moreover, the complement of $\mathcal{A}$ is represented as $\mathcal{A}^c$ in that bits are determined as frozen bots which always fixed as 0. After choosing the information subchannels, the polar code's encoding process is denoted with the matrix multiplication as shown in Eq. (8).

$$x_N = u_N G_N, \ G_N = B \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m} \tag{8}$$

Where the information bits and frozen bits are kept in the vector u_N which represents the codeword required to be encoded; the generator matrix is denoted as G_N; the bit-reversal permutation matrix is denoted as B; the Kronecker product is denoted as $\otimes$ and the encoded codeword is denoted as x_N.

After performing the polar encoding, the binary QC-LDPC-based encoding process takes place to encode the x_N. The null space of sparse circulant array of the same size is used to characterize binary QC-LDPC code. Accordingly, the parity-check matrix H of a QC-LDPC is determined based on the base graph and shift coefficients $(SC_{i,j})$. The circulant permutation matrix and zero matrix of size $Z \times Z$ are used to replace the elements 1s and 0s which exist in the base graph. Consider the code of QC-LDPC is expressed by the following $c_b \times d_b$ array of $Z \times Z$ circulants for two positive integers c_b and d_b with $c_b \leq d_b$ which is expressed in Eq. (9).

$$H = \begin{bmatrix} Q(P_{1,1}) & Q(P_{1,2}) & \cdots & Q(P_{1,d_b}) \\ Q(P_{2,1}) & Q(P_{2,2}) & \cdots & Q(P_{2,d_b}) \\ \vdots & \vdots & \ddots & \vdots \\ Q(P_{c_b,1}) & Q(P_{c_b,2}) & \cdots & Q(P_{c_b,d_b}) \end{bmatrix} \tag{9}$$

Equation (10) shows the exponent matrix of the H given in Eq. (9).

$$\text{Exp}(H) = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,d_b} \\ P_{2,1} & P_{2,2} & \dots & P_{2,d_b} \\ \vdots & \vdots & \ddots & \vdots \\ P_{c_b,1} & P_{c_b,2} & \dots & P_{c_b,d_b} \end{bmatrix} \tag{10}$$

Where $\text{Exp}(H)$ represents the exponent of H, and each entry of matrix E is denoted as shift value. After performing the encoding process, encoded input bit (EX) is further processed with pilot insertion and IFFT where the phase factor optimization is accomplished using the PFRSA.

3.3 Phase Factor-Based RSA (PFRSA)-Based PTS

The PFRSA-based PTS provides the optimal phase factor for minimizing the PAPR during the communication over the MIMO-OFDM. The conventional RSA is turned into PFRSA by considering the complex phase factor as the objective during the searching process which leads to improve the communication. In PFRSA-PTS, the encoded input data block $EX = [EX_1, EX_2, \dots, EX_N]^T$ is separated into V amount of disjoint subblocks which is expressed in the following Eq. (8).

$$EX = \sum_{v=1}^{V} EX_v \tag{11}$$

Next, the subblocks are oversampled and converted by using LN-point IFFT followed by an each subblock which is multiplied with the phase factor obtained from PFRSA-PTS as shown in Eq. (9), and also, it is illustrated in Fig. 2.

$$ex = \text{IFFT}\left\{\sum_{v=1}^{V} b_v.EX_v\right\} = \sum_{v=1}^{V} b_v.\text{IFFT}\{EX_v\} = \sum_{v=1}^{V} b_v.ex_v \tag{12}$$

Where an each subblock's time domain signal is represented as ex_v and b_v represents the complex phase factor which is selected in the range of $[0, 2\pi]$. The fitness function used to identify the optimal phase factor is expressed in Eq. (10).

$$\text{minimize fitness}(b, x) = \sum_{v=1}^{V} b_v.ex_v \tag{13}$$

Where the $b_v = \{e^{j\varphi_v}\}, v = 1, 2, \dots, V, 0 \leq \varphi_v \leq 2\pi$. The iterative process of PFRSA to identify the optimal phase factor is detailed in the following section.
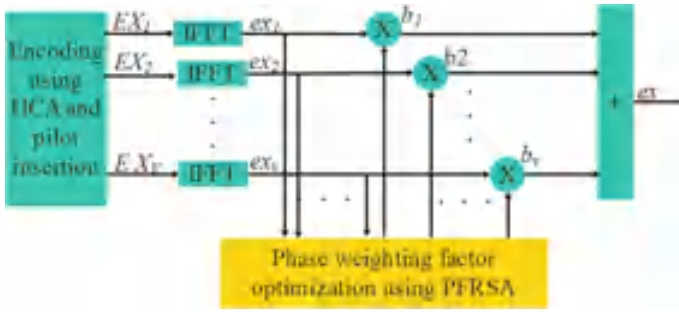
Fig. 2 Phase weighting factor optimization using PFRSA

3.3.1 Iterative Process of PFRSA

In PFRSA, the initial solutions are generated between the range of $[0, 2\pi]$. Both the exploration and exploitation principles are enabled by the motion of crocodile while encircling the target prey. In exploration phase (encircling), two movements such as high walking and belly walking are considered based on the encircling behavior. Equation (11) expresses the position-updating of PFRSA. If the current iteration (t) is lesser than the $T/4$, the high walking is enabled; otherwise, belly walking is enabled during the encircling process.

$$b_{(i,j)}(t+1) = \begin{cases} \text{Best}_j(t) \times -\eta_{(i,j)}(t) \times \beta - R_{(i,j)}(t) \times \text{rand}, t \leq \frac{T}{4} \\ \text{Best}_j(t) \times b_{(r_1,j)} \times \text{ES}(t) \times \text{rand}, t \leq 2\frac{T}{4} \text{ and } t > \frac{T}{4} \end{cases} \quad (14)$$

Where $b_{(i,j)}$ is the j th location of the solution i; the best solution achieved from the whole population is $\text{Best}_j(t)$; random number is denoted as rand; maximum number of iterations is denoted as T; the hunting parameter is represented as $\eta_{(i,j)}(t)$ that is expressed in Eq. (12); β is a fixed parameter whose value is 0.1; $R_{(i,j)}$ denotes the reduce function which is expressed in Eq. (13); $r_1 - r_4$ and $b_{(r_1,j)}$ represent the random number and location, respectively. Moreover, the evolutionary sense ($\text{ES}(t)$) is expressed in Eq. (14).

$$\eta_{(i,j)} = \text{Best}_j(t) \times P_{(i,j)} \quad (15)$$

$$R_{(i,j)} = \frac{Best_j(t) - b_{(r_2,j)}}{Best_j(t) + \epsilon} \quad (16)$$

$$\text{ES}(t) = 2 \times r_3 \times \left(1 - \frac{1}{T}\right) \quad (17)$$

Where the ϵ is a small value and Eq. (15) provides the difference value $\left(P_{(i,j)}\right)$.

$$P_{(i,j)} = \alpha + \frac{b_{(i,j)} - M(b_i)}{\text{Best}_j(t) + (\text{UB}_j - \text{LB}_j) + \int'} \tag{18}$$

Where the average locations are denoted as $M(b_i)$ which is expressed in Eq. (16); the lower and upper limits of the PFRSA are LB_j and UB_j, respectively, and α is set as 0.1.

$$M(b_i) = \frac{1}{n} \sum_{j=1}^{n} b_{(i,j)} \tag{19}$$

Subsequently, the second principle, i.e., exploitation (hunting) is accomplished where it utilizes two approaches such as hunting coordination and hunting collaboration. The hunting coordination takes place when the condition of $t \leq 3\frac{T}{4}$ and $t > 2\frac{T}{4}$ are satisfied; otherwise hunting cooperation takes place in the exploitation as shown in Eq. (17).

$$b_{(i,j)}(t+1) = \begin{cases} \text{Best}_j(t) \times P_{(i,j)}(t) \times \text{rand}, \ t \leq 3\dfrac{T}{4} \ \text{and} t > 2\dfrac{T}{4} \\ \text{Best}_j(t) - \eta_{(i,j)}(t) \times \int -R_{(i,j)}(t) \times \text{rand}, \ t \leq T \ \text{and} t > 3\dfrac{T}{4} \end{cases} \tag{20}$$

After obtaining the phase factor, it is multiplied with the time domain signal of each subblock to accomplish the independent sequence rotation which used to achieve the OFDM signal with small PAPR.

4 Results and Discussion

In this section, the performance of the proposed HCA-PFRSA-PTS and existing researches is evaluated by using the MATLAB R2018a software. The basic system configuration used for the analysis is 6 GB RAM and i5 processer. Here, the MIMO system is considered with two antennas along with two different number of subcarriers such as $N = 128, 512$. The modulation technique used for this HCA-PFRSA-PTS is 16 QAM. The simulation parameters of this HCA-PFRSA-PTS are mentioned in Table 1.

Table 1 Simulation parameters

Parameter	Value
Modulation type	16 QAM
Subcarriers number (N)	128, 256, 512
Oversampling parameter (L)	4
Number of frames	10,000

The performance of the HCA-PFRSA-PTS method is analyzed by means of PAPR versus CCDF and BER versus SNR. Here, the performances are analyzed by means of original OFDM, TC-ESD-PTS [17], SLM-PTS-ABC [18], ACO-PTS [19] and HCA-PFRSA-PTS.

4.1 CCDF of PAPR Evaluation for HCA-PFRSA-PTS

The evaluation of PAPR for HCA-PFRSA-PTS with original OFDM, TC-ESD-PTS [17], SLM-PTS-ABC [18] and ACO-PTS [19] is shown in Fig. 3. The PAPR evaluation of 128 subcarriers is shown here. Here, in Fig. 3, the PAPR is analyzed for by means of CCDF. From the evaluation, it is known that the HCA-PFRSA-PTS achieved a lesser PAPR than the original OFDM, TC-ESD-PTS [17], SLM-PTS-ABC [18] and ACO-PTS [19] for both the subcarrier lengths.

The comparison of PAPR for different subcarrier lengths is given in Table.2. The PAPR for subcarrier length $N = 128, 256, 512$ is given in Table.2. Specifically, the HCA-PFRSA-PTS with subcarrier length of 128 provides lesser PAPR than the HCA-PFRSA-PTS with 256,512 subcarriers. The percentage reduction in PAPR for HCA-PFRSA-PTS is 60.55 over OFDM, 58.5 for ACO-PTS [19], 26.4 for SLM-PTS-ABC [18] and 20 for TC-ESD-PTS [17]. Clear difference of PAPR is observed between the systems with subcarriers 128 and 512.

The HCA-PFRSA-PTS achieves less PAPR, because of the optimal phase factor identification. Since, the derived phase factor is multiplied with time domain signal of each subblock for obtaining the independent sequence rotation that results in less PAPR. Moreover, the reason for HCA-PFRSA-PTS with lesser PAPR than the
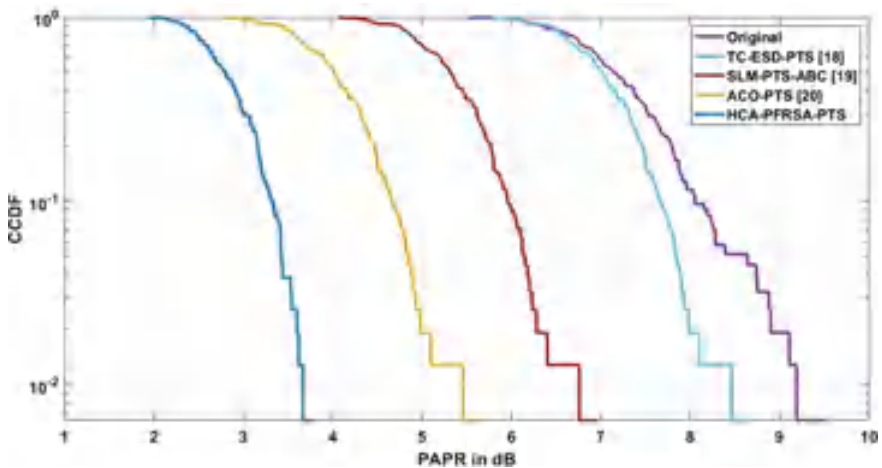
Fig. 3 CCDF of PAPR analysis for HCA-PFRSA-PTS with $N = 128$

Table 2 Comparative analysis of PAPR for HCA-PFRSA-PTS with $N = 256, 512$

Number of subcarriers	PAPR				
	Original OFDM	TC-ESD-PTS [17]	SLM-PTS-ABC [18]	ACO-PTS [19]	HCA-PFRSA-PTS
$N = 128$	9.25	8.45	6.81	5.41	3.65
$N = 256$	9.28	8.49	6.89	5.62	3.73
$N = 512$	9.43	8.59	7.23	6.21	3.97

other swarm algorithms is that it is an effective equilibrium among exploitation and explorations.

4.2 BER Evaluation for HCA-PFRSA-PTS

Figures 4 and 5 show the BER evaluation for HCA-PFRSA-PTS with original OFDM, TC-ESD-PTS [17], SLM-PTS-ABC [18] and ACO-PTS [19]. The (bit error rate) BER evaluation of 128 and 512 subcarriers are shown in Figs. 4 and 5 respectively. The BER is analyzed with respect to energy of bit per noise ratio in decibels. This BER evaluation shows that the HCA-PFRSA-PTS lesser BER for both the subcarrier lengths than the original OFDM, TC-ESD-PTS [17], SLM-PTS-ABC [18] and ACO-PTS [19]. More specifically, the HCA-PFRSA-PTS with $N = 128$ achieves lesser BER during the analysis. The hybrid coding of polar and binary QC-LDPC is used to minimize the errors occurred during the communication. Therefore, the developed HCA minimizes the ISI and ICI on MIMO-OFDM that results in lesser BER than the original OFDM, TC-ESD-PTS [17], SLM-PTS-ABC [18] and ACO-PTS [19].

The comparative analysis of BER for HCA-PFRSA-PTS with $N = 128$ and $N = 512$ is given in Tables 2 and 3, respectively. Here, the comparison of HCA-PFRSA-PTS is done with the existing researches original OFDM, TC-ESD-PTS [17], SLM-PTS-ABC [18] and ACO-PTS [19]. From the tables, it is concluded that the HCA-PFRSA-PTS greatly minimizes the BER for both the subcarrier sizes when compared to the existing researches. The BER of the HCA-PFRSA-PTS is minimized in two different ways: (1) the combination of polar and binary QC-LDPC used to minimize the ISI and ICI caused during the data transmission and (2) PAPR reduction achieved by the PFRSA-based PTS also used to minimize the BER in MIMO-OFDM. Therefore, the proposed HCA-PFRSA-PTS achieves less PAPR and BER when compared to the existing systems as shown in Tables 3 and 4. Table 3 shows the BER analysis for 128 number of subcarriers (N =128), Table 4 shows the BER analysis for 512 number of subcarriers (N =512) respectively.
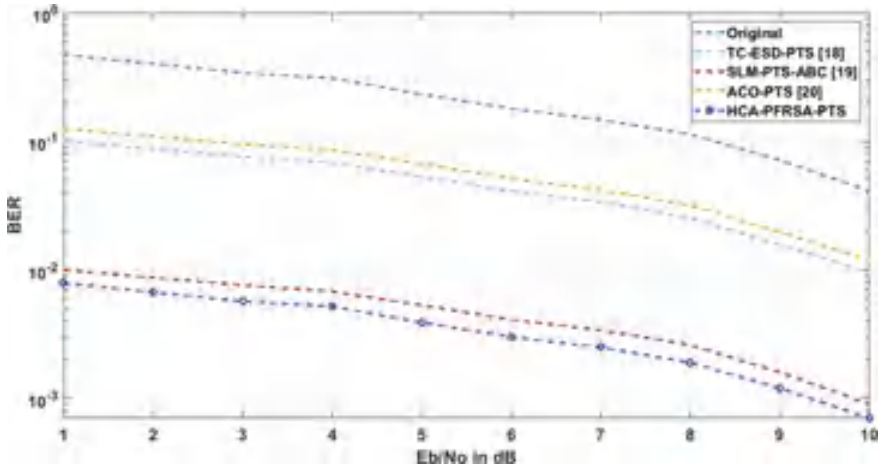
Fig. 4 BER analysis for HCA-PFRSA-PTS with $N = 128$
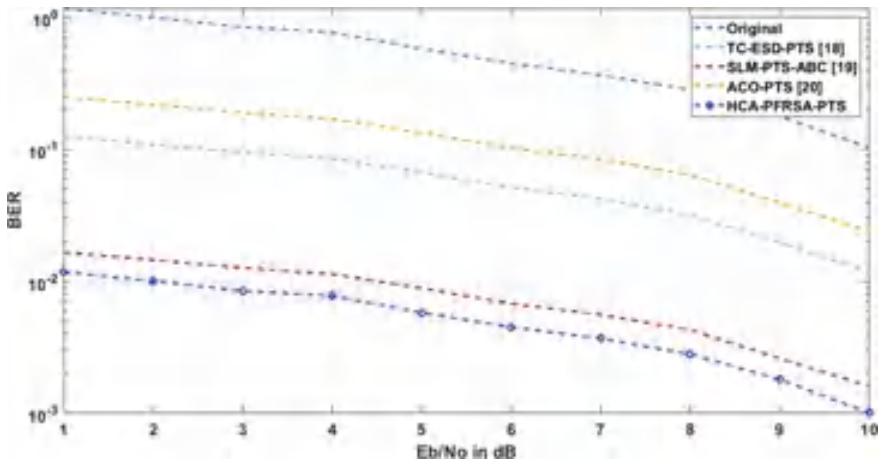
Fig. 5 BER analysis for HCA-PFRSA-PTS with $N = 512$

Table 3 Comparative analysis of BER for HCA-PFRSA-PTS with $N = 128$

SNR (dB)	BER				
	Original OFDM	TC-ESD-PTS [17]	SLM-PTS-ABC [18]	ACO-PTS [19]	HCA-PFRSA-PTS
1	0.4724	0.1003	0.0100	0.1254	0.0079
2	0.4022	0.0875	0.0087	0.1094	0.0067
3	0.3414	0.0762	0.0076	0.0952	0.0057
4	0.3106	0.0682	0.0068	0.0853	0.0052
5	0.2330	0.0533	0.0053	0.0667	0.0039
6	0.1804	0.0411	0.0041	0.0514	0.0030
7	0.1472	0.0337	0.0034	0.0421	0.0025
8	0.1136	0.0255	0.0026	0.0319	0.0019
9	0.0716	0.0158	0.0016	0.0197	0.0012
10	0.0412	0.0095	0.0009	0.0119	0.0007

Table 4 Comparative analysis of BER for HCA-PFRSA-PTS with $N = 512$

SNR (dB)	BER				
	Original OFDM	TC-ESD-PTS [17]	SLM-PTS-ABC [18]	ACO-PTS [19]	HCA-PFRSA-PTS
1	1.1810	0.1254	0.0167	0.2508	0.0118
2	1.0054	0.1094	0.0146	0.2187	0.0101
3	0.8535	0.0952	0.0127	0.1905	0.0085
4	0.7764	0.0853	0.0114	0.1706	0.0078
5	0.5826	0.0667	0.0089	0.1333	0.0058
6	0.4509	0.0514	0.0068	0.1027	0.0045
7	0.3679	0.0421	0.0056	0.0842	0.0037
8	0.2840	0.0319	0.0043	0.0638	0.0028
9	0.1790	0.0197	0.0026	0.0394	0.0018
10	0.1029	0.0119	0.0016	0.0237	0.0010

5 Conclusion

The rapid development of large data and its diverse characteristics demands the efficiency of error correction approach and PAPR reduction in MIMO-OFDM. In this paper, the PTS with PFRSA is proposed for identifying the optimal phase factors of MIMO-OFDM. This optimal phase factor identification results in lesser PAPR. Moreover, the HCA includes polar coding, and binary QC-LDPC is used to perform the error correction while broadcasting the data which resulted in less BER in MIMO-OFDM. On the other hand, the BER is also minimized by lessening the PAPR of the system. Therefore, the reduction in both the PAPR and BER ensures the high-quality

communication in MIMO-OFDM. From the results, it is concluded that the HCA-PFRSA-PTS achieves better performance than the original OFDM, TC-ESD-PTS, SLM-PTS-ABC and ACO-PTS. The BER of the HCA-PFRSA-PTS for SNR 5 dB and subcarriers 128 is 0.0058 which is less when compared to the original OFDM, TC-ESD-PTS, SLM-PTS-ABC and ACO-PTS. It is also observed from simulation results that the PAPR for HCA-PFRSA-PTS is 3.65, which is less when compared to other four methods. Hence, HCA-PFRSA-PTS-MIMO-OFDM with 128 subcarriers is the best performance system.

References

1. Wu W, Cao Y, Wang S, Wang Y (2018) Joint optimization of PAPR reduction based on modified TR scheme for MIMO-OFDM radar. Digital Sig Process 80:27–36
2. Sandoval F, Poitau G, Gagnon F (2019) On optimizing the PAPR of OFDM signals with coding, companding, and MIMO. IEEE Access 7:24132–24139
3. Bhandari R, Jadhav S (2019) Novel spectral efficient technique for MIMO-OFDM channel estimation with reference to PAPR and BER analysis. Wireless Pers Commun 104(4):1227–1242
4. Namitha AS, Sameer SM (2017) A bandwidth efficient selective mapping technique for the PAPR reduction in spatial multiplexing MIMO-OFDM wireless communication system. Phys Commun 25:128–138
5. Nandi S, Pathak NN, Nandi A (2020) A novel adaptive optimized fast blind channel estimation for cyclic prefix assisted space–time block coded MIMO-OFDM systems. Wirel Pers Commun 115(2):1317–1333
6. Sarkar M, Kumar A, Maji B (2021) PAPR reduction using twin symbol hybrid optimization-based PTS and multi-chaotic-DFT sequence-based encryption in CP-OFDM system. Photon Netw Commun 41(2):148–162
7. Vijayalakshmi M, Ramalinga Reddy K (2020) An effective hybrid approach for PAPR reduction in MIMO-OFDM. Analog Integr Circ Sig Process 102(1):145–153
8. Sharma P, Dhubkarya DC (2020) An energy efficient OFDM–MIMO systems for multimedia data transmission based on hybrid fuzzy approach. Wireless Pers Commun 112(3):1431–1450
9. Bao H, Fang J, Wan Q, Chen Z, Jiang T (2018) An ADMM approach for PAPR reduction for large-scale MIMO-OFDM systems. IEEE Trans Veh Technol 67(8):7407–7418
10. Singal A, Kedia D (2020) Performance analysis of MIMO-OFDM system using SLM with additive mapping and U2 phase sequence for PAPR reduction. Wireless Pers Commun 111(3):1377–1390
11. Kumutha D, Amutha Prabha N (2017) Hybrid STBC-PTS with enhanced artificial bee colony algorithm for PAPR reduction in MIMO-OFDM system. J Ambient Intell Humanized Comput 1–17
12. Kumar PR, Naganjaneyulu PV, Prasad KS (2019) Hybrid PS–GW optimised PTS scheme for PAPR reduction in OFDM system. IET Commun 13(18):2996–3002
13. Singh M, Patra SK (2018) On the PTS optimization using the firefly algorithm for PAPR reduction in OFDM systems. IETE Tech Rev 35(5):441–455
14. Watanabe K, Kojima S, Akao T, Katsuno M, Maruta K, Ahn CJ (2019) Modified pilot selection for channel estimation of systematic polar coded MIMO-OFDM. ICT Express 5(4):276–279
15. Agarwal A, Mehta SN (2020) PC-CC: An advancement in forward error correction using polar and convolutional codes for MIMO-OFDM system. J King Saud Univ-Comput Inf Sci 32(8):917–927
16. Hadj Ali T, Hamza A (2021) Low-complexity PAPR reduction method based on the TLBO algorithm for an OFDM signal. Ann Telecommun 76(1):19–26

17. Rakshit M, Bhattacharjee S, Sanyal J, Chakrabarti A (2020) Hybrid turbo coding PTS with enhanced switching algorithm employing DE to carry out reduction in PAPR in AUL-based MIMO-OFDM. Arab J Sci Eng 45(3):1821–1839
18. Sidiq S, Sheikh JA, Mustafa F, Malik BA, Sofi IB (2021) PAPR minimization of FBMC/OQAM scheme by hybrid SLM and PTS using artificial: bee-colony phase—optimization. Arab J Sci Eng 46(10):9925–9934
19. Hosseinzadeh Aghdam M, Sharifi AA (2021) A novel ant colony optimization algorithm for PAPR reduction of OFDM signals. Int J Commun Syst 34(1):e4648

G-EYE: Smartphone Compatible Portable Indirect Ophthalmoscope for Generating Quality Fundus Images

N. Kalyani, M. Seetha, Y. Sravanidevi, and M. S. V. L. Sasirekha

Abstract Retinal imaging remains as challenge using a Smartphone. Fundus imaging through portable device enables fundus examination affordable, which can bring a revolutionary change in eye diagnosing system. This key feature attracts attention of many researchers to explore and design portable devices to capture the best quality of image. This makes it simple to gather a dataset of fundus images that can then be used to train various models linked to eye ailments. The dataset is generated with possible maximum resolution depending on smartphone camera and to train and test deep learning models for the prediction of eye diseases. The portable device can be used to record eye fundus in rural areas where there aren't any slit lamp devices or expensive indirect ophthalmoscopes. The inability to successfully complete computer-aided diagnosis has been hampered by the lack of extensive publicly available information. The contributions of this paper is the design details of the portable device G-EYE that is compatible to smart phone and methods adopted to collect quality retinal images for research.

Keywords Portable fundus camera · 20D lens · Ophthalmoscope · Raspberry pi 3 · Retinal imaging

1 Introduction

The human eye is an organ of vision and one of the most important parts of the body. Unfortunately, there are many factors that can cause problems with eyesight. However, other eye conditions and illnesses are more severe and can cause blindness that lasts a lifetime. Age-related macular degeneration, glaucoma, and other late-onset ocular diseases are becoming more common in wealthy countries and are one of the main causes of blindness. Treatment for visual impairments has undergone a

N. Kalyani (✉) · M. Seetha · Y. Sravanidevi · M. S. V. L. Sasirekha
Computer Science and Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad, India
e-mail: nara.kalyani@gnits.ac.in

revolutionary transformation as a result of the introduction of computer-aided eye ailment identification. It has several characteristics and advantages that are highly beneficial to doctors. This ground-breaking technique is now the focus of the attention of many researchers.

A classification model is created for the categorization of illnesses, and it requires a huge collection of fundus pictures that are clinically recorded using a high-end ophthalmoscope. The main obstacle to the effective completion of this computer-aided diagnostic has been the lack of a substantial publically available dataset. Although a dilated pupil generally provides a superior field of vision than an undilated pupil, increasing the field from 40° to 90°, there are certain drawbacks. After the dilation, it takes the eye 15–20 min to prepare for the inspection, and the patient will suffer blurry vision for roughly 6 h.

The fundus-oculi is seen with a typical ophthalmoscope, which provides a crisp image but is extremely expensive to setup, has extensive maintenance requirements, and needs an expert to operate. These facilities are not available to those who live in rural or distant areas. Many rural residents are unable to access hospitals as a result of these restrictions, and as a result, they frequently get into risk of vision loss. The major goal of this experimentation is to create a portable, readily accessible device for imaging the fundus of the eye that can take high-quality fundus pictures anywhere, providing an algorithm with more data for accurate analysis.

There are various technologies now in use for seeing the fundus, which an ophthalmologist uses to diagnose the patient's eye condition. Below is a list of some of the market's current systems, along with some of its features and drawbacks.

D-EYE: The D-EYE device covers the smartphone's camera aperture and LED light source for use as a portable retinal imaging tool. The examiner uses the D-EYE App on the smartphone's display to enter patient information, focus the retinal camera, record, archive, view, and transmit photographs. The ability of this gadget to accurately observe the eye's optic disc and its ease of use for both specialists and normal people after proper training are its main benefits. Only accessible on a small number of smartphone models, including the iPhone 6S+, is a drawback. A constrained field of view, couldn't accurately capture the macula of the eye [1]. It is proved that 20D lens can be used to acquire good quality images as in clinical environment which is further enhanced with additional Koeppe lens to result in best quality image [2].

Peek Retina: The portable smartphone accessory Peek Retina, allows the user to view and record photographs of the optic nerve through a dilated pupil [3]. It was intended to be simple to use, straightforward, and adaptable to any location, including clinical settings and isolated rural places. It enables quick sharing of movies or photographs taken of the back of the eye with other experts, transfer to Electronic Patient Records (EPRs), or to show the patient what the user has seen and enables users to inspect the optic nerve to discover disorders such as glaucoma and bulging discs. Important restrictions are there is very little examination that does not involve dilating eye drops and Peek Retina cannot be utilized with confidence for disorders other than optic nerve evaluation, such as diabetic retinopathy.

ODOCS: A smartphone-based hardware ocular and retinal imaging system called ODocs was created to work with the cellphones [4]. They want to make ophthalmic imaging equipment as affordable as an iPhone. A 90% cost decrease over traditional retinal cameras. A portion of the purchase will help the Fred Hollows Foundation and the New Zealand Blind Foundation restore sight to the Pacific Islands. An iOS device may be used to control this device. ODOCS Pros are Usability for both experts and regular users and Capable of precisely seeing the eye's optic disc. ODOCS Cons are Due to the holder's inability to fit a 90D lens for experimenting, one kind of lens can only be used to examine objects clearly and if the camera and flash are not positioned correctly, it will be difficult to move the camera to observe the fundus.

Ultimate Benefits and Disadvantages: The fundamental benefit of all the eye examination systems already in use that were previously discussed is that ophthalmologists may utilize the video captured by these devices to determine the patient's eye's state. So when a doctor does an examination, the results are accurate, and if the patient has an eye condition or issue, the doctor can provide treatment options and even perform surgery and prescribing medication if necessary. Additionally, it is simple to access because they are portable. Only a few of the aforementioned devices are used to save fundus images, and all of them need dilating the pupil in order to view the retina. And just a few mobile phone versions are compatible with certain devices.

Mahmut Karakaya et al. made a study and compared different smartphone-based portable retinal imaging systems that are available in the marked and observed that using deep leaning framework would give good accuracy for DR when the quality of imaging is good [7]. His experiments resulted in accuracy of 61% for iExaminer, 62% for D-Eye, 69% Peek Retina, and 75% iNview.

Atalie et al. mentioned in a review paper that deep learning techniques are promising in screening of glaucoma [8]. The deep learning models could detect and quantify the glaucomatous damage using simple colour fundus images captured using portable devices. This gave scope for potential research in low-cost screening mechanisms. However, it is important to note that no matter how exciting AI technologies are it is important to validate new diagnostic tests to be rigorous with reference to standards that are defined by ophthalmologists.

Charlotte Laurent et al. conducted experiments on 272 eyes of 215 patients in capturing the video for examining the presence of Spontaneous venous pulsation (SVP) on slit lamp [9]. For observer 1, sensitivity was 84.77% with 128 out of 151 videos graded as positive for SVP. Similarly for observer 2 sensitivity was 76.82% with 116 videos correctly identified.

The challenge is not only on design of portable device but even with design of suitable model that can classify the image. Simi Sanya et al. [10] used ensemble CNN model to classify the diabetic retinopathy and achieved 74% accuracy. The review by Sravani et al. [11, 12] reported that the detection of retinal abnormalities using deep learning techniques has good potential nut this research requires huge data. Comparison of results obtained by methods like CNN and DCGAN and observed that the loss on general images is less and can be recommended for classification of eye diseases.
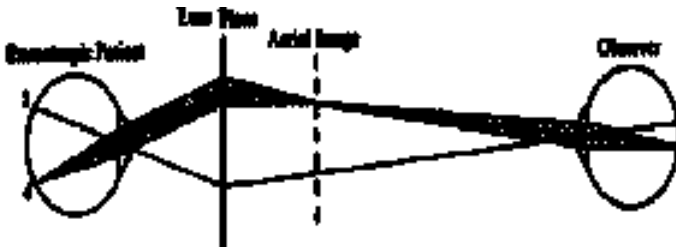
Fig. 1 Principle of ophthalmoscope

2 Proposed System G-EYE

Research has been done to develop a solution to the challenges described above. The major goal is to use the 20D lens and flash light in conjunction with the indirect ophthalmoscope concept. Using a device worn on the head, the doctor flashes an extremely strong light into the patients eye while holding the eye open. Through a lens held close to the eye, the healthcare professional can see the back of the eye. A tiny, blunt probe can be used to exert some pressure on the eye. The patient is instructed to look in different places. Typically, a detached retina is checked for during this examination.

2.1 Principle of Ophthalmoscope

In indirect ophthalmoscope, the ophthalmoscope lens's job is to refocus diverging light from the patient's pupil towards the observer's eye. The lens also concentrates parallel rays into an inverted aerial view of the patient's fundus as seen in Fig. 1 as a result of this action. One of the most obvious and inescapable features of indirect ophthalmoscope is the presence of an inverted aerial picture.

2.2 Application of the Concept

Here, an attempt is made to apply the same principle. When a customized smartphone flash at a distance of 25 cm from the patient and the 20D lens at a distance of 5 cm. The parallel light beams from the flash pass through the medium and converge at the 20D lens before diverging again and convergent at the eye lens to create the inverted picture of the retina (Fig. 2).

The fundamental premise behind this is that while eyes can sense a smartphone's light, using the phone's flash in a dark environment doesn't induce much pupil dilation. However, the eye cannot contract with an IR LED, which is beyond the range of
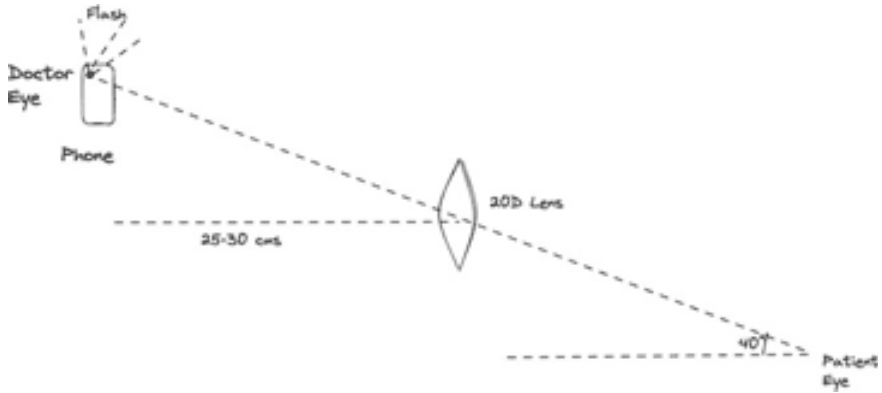
Fig. 2 Alignment of phone camera, light source, lens and patient eye for viewing fundus

human vision, and the pupil remains dilated in the dark. The black and white fundus may be clearly seen and captured by the night vision camera. As soon as the fundus is properly visible, we turn on the light so that we may record a colour image of the pupil constriction. The same indirect ophthalmology may now obtain a greater field of vision than the one with the 20D lens alone since the pupil is completely dilated.

3 Methodology

The study focuses on employing a portable instrument to capture a fundus. So let's start with the videos that were recorded using the 20D lens, Raspberry Pi, and connected night vision camera. A Python application receives this video as input and uses it to create picture frames. The picture frames are next combined into a single, high-quality image that may be utilized for deep learning techniques on eye illnesses, at least for glaucoma, age-related macular degeneration (ARMD), and diabetic retinopathy. This image is then stored in the database.

Primary Goals

- To use a portable equipment to take a picture of an eye's fundus.
- Lessen the use of mydriatics and extend the pupil size, which would aid in expanding the region that might be caught.
- Create an appropriate adaptor to handle the camera, light source, and lens all at once.
- To create a portable device by including the elements needed to use a smartphone to take retinal pictures.

3.1 Design of Fundus Camera

This non-mydriatic fundus camera's hardware configuration is based on the Raspberry Pi 3 Model B and the No IR camera board. The Raspberry Pi is an environment-friendly computing board the size of a credit card. With the exception of its fixed focus and sensitivity to infrared light [6], the No IR camera board's 5-megapixel camera is comparable to commonly used smartphone cameras. Micro-USB and HDMI cables are used to connect the Raspberry Pi to the LCD display [5]. Indirect ophthalmoscopy is performed with a 20-diopter condensing lens carried in the free hand in addition to the fundus camera. Software for the apparatus: Raspbian is the operating system for the Raspberry Pi, and it came pre-installed on an 8 GB NOOBS micro-SD card (Raspberry Pi Foundation). Through the LCD touchscreen and the free Florence virtual keyboard app, commands are entered into the Raspberry Pi. The Supplementary Instructions provide a straightforward Python programme for the fundus camera that is based on Dave Jones' picamera module.

3.2 Capturing Fundus Video

For this model, the following equipment is needed to capture the fundus:

- Raspberry pi 3
- IR led and NOIR camera
- Micro-USB cable
- LCD screen
- 20D lens

This module intends to record the fundus in a dim environment using an IR LED and NOIR camera, with the Raspberry Pi acting as the camera's operating system to start recording and save the footage. The following figure shows a representation of how this equipment is set up.

3.3 Extraction of Image from Video

In this process a.png or.jpg picture is made for each time-frame of 0.5 s till the film ends after image frames are extracted from the video utilizing the time-frame concept. A new folder with the same name as the video will be created automatically in the specified directory, where all the produced image frames will be placed. An intelligent system is available that can recognize the corners of a document in a reference frame, and the user can also choose the index of the reference frame. The
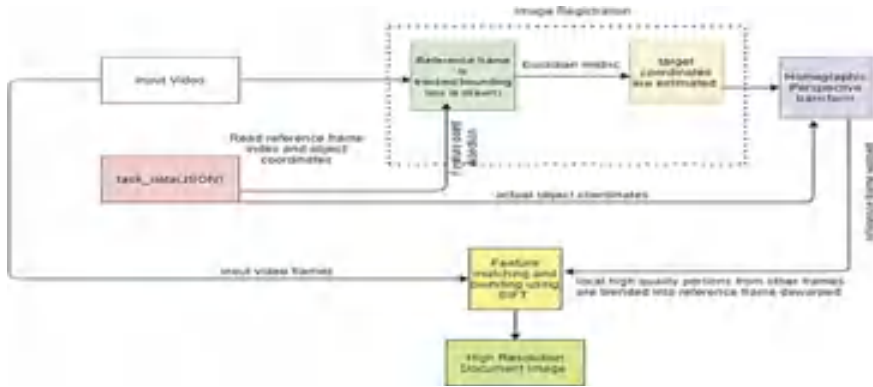
Fig. 3 Block diagram of the method

following inputs are required in order to obtain a high-resolution picture from a video:

1. A collection of frames that includes a reference frame that contains the whole picture of the document.
2. The reference frame index.
3. The coordinates of the corners of the document in the frame of reference.

The video frames contain top-notch data from various sections of the study. The fundamental concept of the approach is to extract this data from the frames and mix it with other data to get a final image. Determining where the content in a frame genuinely belongs in the main text is essential because of this we utilized sped-up robust features (SURF) to discover the patches in the final mosaic, which was a picture that had been tiled into sections and each area had a new image that matched the target image replaced with it (Fig. 3).

Patches are also de-warped to the output projection plane. The pixel values in the final output are weighted averages of all the pixels in the same location across the de-warped frames. The weighted de-warped frames vary in crispness. The task's data.json file and uploaded video are processed using the Open CV library as follows. An index is used to keep track of the reference frame, and a bounding box is produced using the top left, top right, bottom left, and top right coordinates of the object in the reference frame. This is known as feature point/key/interest point detection. The system's first stage, image registration, is completed by calculating the target image's height and width using Euclidian metrics. The image is now warped straight, the superfluous corners are eliminated, and the image is properly aligned using the target coordinates from the previous stage and the original object coordinates in the reference frame from the data.json file. The warped reference frame will be compared to the other frames in the movie using the scale-invariant feature transform (SIFT) technique.

A high-resolution document image is created by identifying, matching, and blending local high-quality patches from the frames with the distorted reference

frame sections. The produced image is compared to the original image using the mean structural similarity measure to assess its accuracy. From 0 (inaccurate) to 1, the output is available (stands for highly accurate).

4 Implementation

4.1 Architecture of Fundus Camera

The current equipment used for routine eye exams, such as the D-eye, ophthalmoscope, has a larger field of vision and also produces high-quality fundus pictures. However, many of these benefits can only be utilized in a clinical setting, which many rural residents cannot afford owing to a lack of resources (equipment and money). Only a small number of currently available portable devices are also compatible with as many mobile models as you can count on one hand.

The G-EYE model was developed to improve the use and compatibility of the fundus camera, which can record fundus and classify eye illnesses. Figure 4 depicts the architecture of G-EYE's.

4.2 Algorithm for Scale-Invariant Feature Transforms

We will be able to shoot a video using the aforementioned infrastructure, which will then need to be combined into a single image to provide the examiner a clear view of the whole eye in one image frame. To do that, we attempt to extract picture frames from the recorded video using the time-frame idea. With this idea,.png/.jpg images may be generated at intervals of 0.5 s till the film is finished. The same directory will automatically generate a new folder with the same name as the video. This will store every single photo frame that is produced.

We will now attempt to combine all of those photographs into a single image using an intelligent agent. Video footage and a json file with the object coordinates in the reference frame will be used as the inputs for this (top left, top right, bottom left, and bottom right). One can come close to the goal coordinates by building the bounding box using Euclidean geometry and the reference frame from the video. The local high-quality parts are combined with other frames to provide a reference frame. The blended frame and the picture frames created from the previous movie are entered into the scale-invariant feature transform (SIFT) algorithm. The outcome is a document picture with superb quality.

Fig. 4 Architecture of
G-EYE
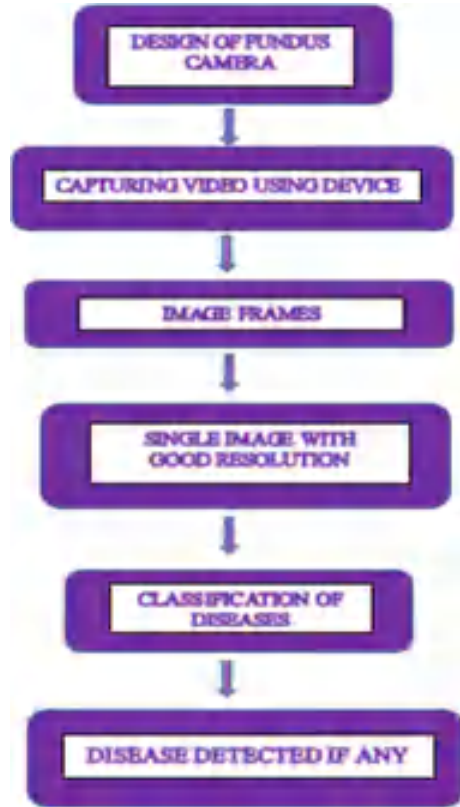
4.3 Camera Operation and Retinal Viewing Techniques

Without employing mydriatics, we need to have a suitably enlarged pupil size in order to see the retina. The pupil typically shrinks when the eye is exposed to light and can enlarge in darkness. Ophthalmologists find it challenging to inspect the eye as a result. Therefore, mydriatics are used to dilate the pupil, which also expands the field of view from 40° to 90°, in order to capture the fundus picture. It takes 15–20 min for this mydriatic to take action, and the patient continues to suffer blurring effects six hours after dilating. The patient is evaluated in a darkened setting to fulfil this requirement without the use of mydriatics. The light from the gadget causes the eyes to constrict once again even when we are in a darkened area. Therefore, we decided to utilize IR led and NOIR camera, which can take pictures in the dark, as an alternative (Fig. 5).

We employ a 20D lens to carry out this investigation. Despite the fact that there are many different types of lenses on the market, it is important to consider distance and magnification. Figure below shows that the 20D lens has a higher magnification than the 30D lens, and although having a higher magnification, we don't pick the
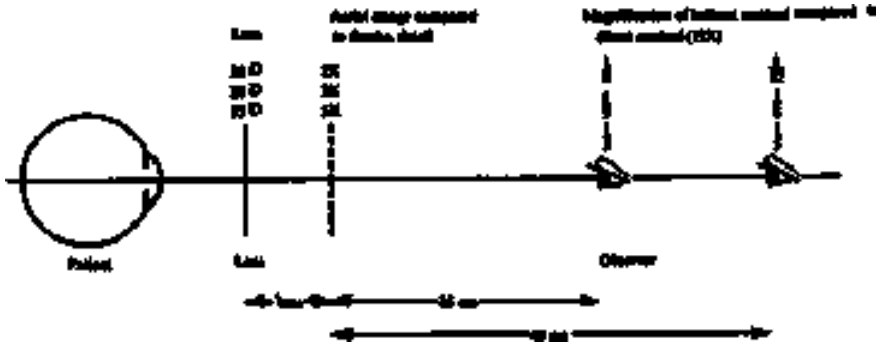
Fig. 5 Different lens along with their magnification

13D lens because of its availability and working distance. 13D lenses need a larger working distance to achieve that magnification (Fig. 6).

While camera operation when the micro-USB cable is inserted into the Raspberry Pi, the fundus camera is activated and the Python application is launched and turns on both the viewfinder and the infrared led light. The patient's pupils might naturally enlarge since the room is darkened by the fundus camera user. The user then conducts infrared indirect ophthalmoscopy while holding the condensing lens in one hand and the fundus camera in the other. The user will see a black and white image of the fundus when the condensing lens, fundus camera, patient's retina, and pupil are all correctly aligned. They take a colour snapshot of the fundus once they are pleased with the image. Utilizing the Raspberry Pi, a wireless dongle, or an Ethernet cable, the picture may then be accessed from the memory card and copied to other computers. The picture quality is enough for making broad diagnoses such optic disc sharpness.
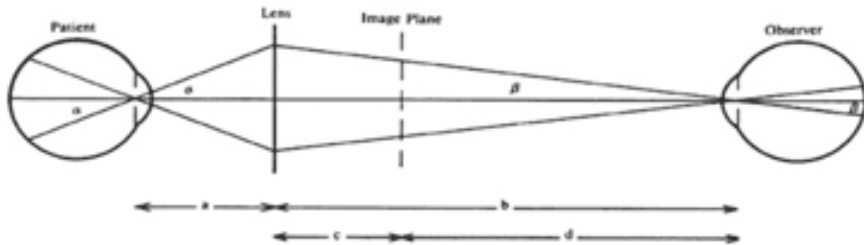
patient	lens	c	d	b	1/b	1/a	a	a/b	a+b
Emm.	13 D = 7.7 cm		25 cm	32.7 cm = 3.1 D		9.9 D = 10.1 cm		1/3	43 cm
			40 cm	47.7 cm = 2.1 D		10.9 D = 9.2 cm		1/5	58 cm
Emm.	20 D = 5 cm		25 cm	30 cm = 3.3 D		16.7 D = 6.0 cm		1/5	36 cm
			40 cm	45 cm = 2.2 D		17.8 D = 5.6 cm		1/8	51 cm
Emm.	30 D = 3.3 cm		25 cm	28.5 cm = 3.5 D		26.5 D = 3.8 cm		1/8	32 cm
			40 cm	43.3 cm = 2.3 D		27.7 D = 3.6 cm		1/12	47 cm

Fig. 6 Working distances with different types of lens

Fig. 7 Captured fundus image

5 Results and Discussion

Deep learning algorithms may use this high-resolution picture to forecast eye problems once the video from the 20D lens and Raspberry Pi computer has been correctly converted into image frames. In remote locations without access to expensive indirect ophthalmoscope equipment or tabletop slip lamp devices, the portable gadget can be utilized to take pictures of the eye fundus. A 4- to 5-disc diameter defines the video's field of view. The technology generates 120 frames on average for a minute of video. One frame from the recorded video shows the section of the fundus. The image created by this model contains the crucial fundus area necessary to correctly identify the target disorders when utilizing deep learning techniques after combining all pertinent frames. The algorithm used to rebuild the fundus has an accuracy grade of 0.64 (Fig. 7).

Using the built equipment, the aforementioned Fig. was taken. The area in the image above is known as the "Optic disc." Optic cup is the optical disc's primary feature. To determine if the eye is healthy or not, the optic cup ratio is essential. 0.3 Or 1/3 is the typical cup to disc ratio. Cup might get larger because of glaucoma. Therefore, the doctor must pay close attention to the temporal side nerve's rim.

6 Conclusion and Future Scope

The portable device's mobility, area of view, and magnification might all be enhanced and the resulting image quality can be further enhanced for a clearer understanding of eye issues. The Android software may be linked with machine learning models for anticipating eye problems including age-related macular degeneration, glaucoma, and diabetic retinopathy. Since there are now so few datasets available in the field of study, as soon as the image is taken and the Android app has processed the video, it can be analysed using machine learning models to make predictions more

quickly while also improving the models' accuracy. The picture must be obtained before using the convolutional neural networks (CNNs) model. Using generative deep learning techniques, a considerably larger dataset is created and diseases like age-related macular degeneration, cataract, glaucoma, diabetic retinopathy can be detected through generic deep learning algorithms and may produce new substantial datasets of fictitious retinal images from the pre-existing datasets. These pictures could improve the current dataset, improving fidelity. The new dataset—which contains both actual and synthetic images—could be beneficial when creating new deep learning diagnostic models, for which enormous datasets from major clinical trials of correctly graded photographs may not exist. In data synthesis techniques that make use of augmentation, the same images have been utilized again. This reduces the model's accuracy. Generative Adversarial Networks (GAN) should generate pictures that seem realistic and increase the model's precision. Given that the deep learning model's performance should be on par with that of genuine images, training using the synthetic dataset may be advantageous for both humans and robots.

References

1. Russo A, Morescalchi F, Costagliola C, Delcassi L, Semeraro F (2015) A novel device to exploit the smartphone camera for fundus photography. J Ophthalmol 2015, Article ID 823139. https://doi.org/10.1155/2015/823139. Open Access.
2. Haddock LJ, Kim DY, Mukai S (2013) Simple, inexpensive technique for high-quality smartphone fundus photography in human and animal eyes. J Ophthalmol 2013:5, Article ID 518479
3. Bastawrous A, Giardini ME, Bolster NM, Peto T, Shah N, Livingstone IA, Weiss HA, Hu S, Rono H, Kuper H, Burton M (2016) Clinical validation of smartphone based adapter: peek retina for optic disc imaging in Kenya. JAMA Ophthalmol 134(2):151–158. https://doi.org/10.1001/jamaophthalmol.2015.4625
4. Lockward JS, Arbaje NA, Méndez MG, Juan A, Peña JS, Pichardod JB (2022) Telemedicine in retinopathy of prematurity: crossing borders in pediatric vision health. J Ophthalmol 97(9)
5. Shen BY, Mukai S (2017) A portable, inexpensive, nonmydriatic fundus camera based on the Raspberry Pi® computer. J Ophthalmol 2017
6. Rossi L (2012) Pupil size under different lighting sources proceedings of CIE 2012. Lighting quality, vol CIE x037
7. Karakaya M, Hacisoftaoglu RE (2020) Comparison of smartphone-based retinal imaging systems for diabetic retinopathy detection using deep learning. BMC Bioinf 21(Suppl 4):259. Published online 2020 Jul 6. https://doi.org/10.1186/s12859-020-03587-2
8. Thompson AC, Jammal AA, Medeiros FA (2020) A review of deep learning for screening, diagnosis, and detection of glaucoma progression. Transl Vis Sci Technol 9(2):42. Published online 2020 Jul 22. https://doi.org/10.1167/tvst.9.2.42
9. Laurent C, Hong SC (2020) The detection of spontaneous venous pulsation with smartphone video ophthalmoscopy. Clin Ophthalmol 14:331–337. Published online 2020 Feb 3. https://doi.org/10.2147/OPTH.S238897

10. Sanya S, Maddala S (2021) An ensemble of convolutional neural network for diabetic retinopathy detection. Int J Eng Res Technol (IJERT), ICDML—2020 Conf Proc 9(2), special issue 2021. ISSN: 2278-0181
11. Sravani Devi Y, Phani Kumar S (2020) A Scoping review of diabetic retinopathy detection techniques using deep learning: taxonomy, methods, and recent developments. High Technol Lett. ISSN: 1006-6748
12. Sravani Devi Y, Phani Kumar S (2021) Retinal image synthesis for diabetic retinopathy assessment using dcgan and vae models. Int J Adv Electron Comput Sci 8(10). ISSN (p): 2394-2835

River Network Identification from Satellite Imagery Using Machine Learning Algorithms

M. Seetha , **D. V. Lalitha Parameswari** , **and G. Malini Devi**

Abstract Over 500 million people live near rivers, which are significant coastal depositional systems. River network identification is needed for hydro logical simulations to improve the understanding of hydro logical processes. They have a great significance to flood assessment and ship navigation. Rivers have a tight connection to ecological, socioeconomic environment, and agriculture. Mapping of rivers and canals networks is very important in applications related to water resources management. It helps us prevent significant loss of life and property by detecting natural disasters like floods. In this project the proposed model enhances and detects the complete river networks. Enhancement and feature extraction can be done using filters such as Gabor, PCA, and GLMC. Gabor filter helps in enhancing the river cross-sections and longitudinal continuity. A dataset's dimension (columns) is reduced through Principal Component Analysis (PCA). To acquire statistical texture features, Grey level co-occurrence matrix (GLCM) is developed. The number of pixels that change between two pixel values is counted by GLCM. In other words, the histogram's bin with indices equal to the values of the two pixels is increased. The strategies for classifying from a high-resolution multispectral satellite image, the river network is mapped using support vector machines and Random Forest. Support vector machine (SVM) and Random Forest are supervised machine learning models that uses classification algorithms for two-group classification problems. These algorithms can provide robust, accurate, and effective results for the provided image samples. A comparative study can be done on these classification techniques to evaluate the performance measures like accuracy and kappa coefficients.

M. Seetha (✉) · D. V. Lalitha Parameswari (✉) · G. Malini Devi
Department of CSE, G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India
e-mail: maddala.seetha@gnits.ac.in

D. V. Lalitha Parameswari
e-mail: dvlalitha@gnits.ac.in

G. Malini Devi
e-mail: gmalini12@gnits.ac.in

369

1 Introduction

Remote sensing is the process of gathering data over a distance. Remote sensors on satellites and aircraft are used to view Earth as well as other planetary bodies by detecting and monitoring reflected or emitted energy. Remote sensors provide a global view and a wealth of data in relation to Earth systems. For enabling the decision making these sensors are used based on its current and probable future status of our planet. Examples include the ability to view with substantial transparency than one can from the ground. The sophisticated cameras on satellite and aircraft captures the images of the earth's surface in a wide-ranging approach, which captures photos of vast portions of the Earth's surface.

- Images of the ocean floor can be acquired by using sonar systems on ships instead of delving all the way to the bottom.
- Images of variations in ocean temperature can be captured by satellite cameras.

   Remotely sensed photographs of the Earth can be used for a variety of purposes, some of which are as follows:
- The large forest fires giving overseers a far wider view than they would have on the ground can be mapped from space.
- Keeping an eye out for dust storms by monitoring the eruption of volcanoes, tracking clouds to help anticipate the weather.
- Helps in expansion of cities like monitoring phase-wise changes and transformations in woods or farms over a period.
- With its vast narrow valleys, mountain ranges, "magnetic striping" and charting of the arid topography of the ocean floor can be identified.

   River networks are dynamic in nature, changing over time due to a variety of natural and human-caused events. They may contract, grow, or alter their appearance or flow direction. Other natural resources, human resources, and the environment are all impacted by changes to water bodies. Surface water volume changes frequently have significant effects. In extreme cases, flooding can happen when surface water levels increase too quickly. Therefore, it's crucial to accurately detect the existence of surface water, extract its area, calculate its volume, and observe its dynamics.

   Over the past few decades, there has been a reduction in the monitoring of river discharge on a global scale. This data is almost accurate even though these data is essential for monitoring and predicting various levels related to river flow. Despite the potential strategic importance of the water resources in poorly gauged river basins, gathering data on them can be challenging for a number of reasons, such as limited accessibility, high seasonal variability, and, in some parts of the world, the presence of large wild animals. The need to update stage discharge relationships through

fieldwork more regularly than in other river systems is a result of changes in river geometry, which is another problem when it comes to data collection in places of this sort. River networks are vigorous in nature because they change over time due to a variety of natural and human-caused variables, which may cause them to contract, expand, or alter their appearance or path of flow.

Changes in water bodies have an impact on the overall ecosystems. Usually, changes in surface water volume have negative effects [2]. In extreme cases where surface water levels rise quickly, flooding may happen. Because of that, it is essential to accurately detect the presence of surface water, determine its extent, calculate its volume, and keep track of its dynamics [3].

The volume of water bodies can be mapped at the provincial or even global scale, and their dynamics can be observed at regular and frequent time intervals, using remote sensing data sets that provide spatially explicit and temporally frequent observational data about a variety of physical attributes about the Earth's surface. Such quick modifications may be found via remote sensing [4]. These issues focus on the requirements to examine the data collection methods which reduces dependance over pragmatic relationships among flows and perpetual flow proxy observations over typically water levels, based on these observations which reduces the costs [7].

2 Literature Survey

Over the past few decades, a great deal of research has gone into the identification of surface water using water indices, and it has been shown that this method can identify water from background features.

Different machine learning (ML) techniques are employed in the literature survey, and they are assessed according to how well they do classification. To achieve this goal, various papers are categorized into the three domains of science, business, and social science using machine learning (ML) approaches [12]. Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbour, Decision Tree, and Random Forest are the ML approaches used in this work. The approach and algorithms for picture recognition utilizing the aforementioned ML techniques are presented, in addition to a discussion of the ML algorithms that were used [10].

According to the comparative study using four distinct performance indicators, SVM and Random Forest are appropriate. The spectral and spatial properties were integrated by automatic multi-feature water body extraction (MFWE) approach, is proposed for water body extraction, i.e. unsupervised from GF-1 multispectral imagery [11].

The pixel region index (PRI) is discussed here, the extent of smoothness in a limited area surrounding a pixel is measured which is a spatial feature index. Particularly in metropolitan settings, PRI is useful for supporting the normalized difference water index in identifying significant water bodies [9]. On the other hand, since some of the water pixels close to the boundaries might not be covered by significant water bodies, k-means clustering is then used to group all the water pixels together as a

reference map [14]. The final water mask is created by merging the principal water bodies with the route map.

The current state of optical remote sensing for surface water detection, extraction, and monitoring, particularly the development during the past ten years. To address this conflict, methods including spatiotemporal fusion and pixel unmixing and reconstruction have been developed [6]. Remote sensing data and in-situ river flow have been combined to describe the spatiotemporal dynamics of surface water. The historical problem with optical sensors has been cloud and vegetarian. Combining synthetic aperture radar data with existing data is an efficient way to overcome this restriction. Cloud/terrain shadows have also been eliminated using digital elevation model data [5]. The emergence of big data and cloud computing techniques has made it simpler to meet the rising need for high-resolution global water dynamics monitoring.

Utilizing digital elevation models (DEMs) and Landsat imagery, a novel method is used to gather data on storage fluctuations in any inaccessible place [8]. The method is suited for monitoring tiny, frequently undocumented irrigation reservoirs and differs from existing approaches in that it does not call for any in-situ measurements [13].

3 Methodology

To solve these problems, a support vector machine-based river extraction method is put forth. These methods were used to automatically extract features from a high-quality multispectral satellite image and map the river network.

The five steps that make up the suggested procedure are data collection, feature extraction, preprocessing, training the model, accuracy evaluation and testing.

A. Objectives: To achieve the main objective, the supporting objectives are formulated as:

- To extract automatic feature for mapping of river network.
- The National Research Space Centre (NRSC) provided the photograph.
  Evaluate SVM and Random Forest classification techniques on detection of river networks.
- Evaluate these classification techniques by a comparative study with performance measures like accuracy and kappa coefficients.

B. Data Collection: Data were gathered by National Remote Sensing Centre (NRSC).

C. Preprocessing: Images have been preprocessed to encounter the codes. To distinguish between the river region and other regions the data augmentation is done to label the pixels.

D. Feature Extraction: The features from the river images are extracted using Gabor filters and canny edge recognition.

Fig. 1 Flow of river
segmentation
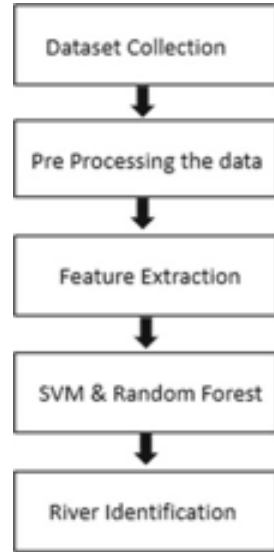
E. SVM: A supervised machine learning model called a support vector machine (SVM) employs classification techniques to solve two-group classification problems. SVM is regarded as one of the most familiar techniques for classifying patterns and images.

F. Random Forest: The supervised learning method includes the well-known machine learning algorithm like Random Forest. Random Forest is a classifier that functions by using various decision trees on several subsets of the input dataset and averages the results to enhance the dataset's predicted precision.

G. River Detection: A model is saved and used to predict rivers in additional images after training using SVM and Random Forest independently (Fig. 1).

4 Implementation

4.1 Datasets Used

Dataset is gathered from NRSC and it was a satellite image with 23 GB in size. ERADAS imagine is used to see the satellite image (Fig. 2).
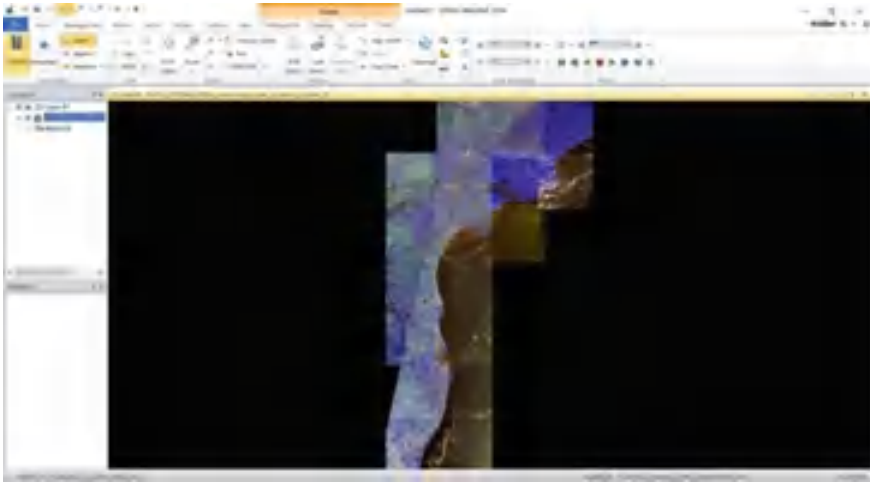
Fig. 2 Remote sensing image in ERADAS

4.2 Data Preparation

Subset of Image: A downloaded subset is a smaller version of a sizeable image. Due to the fact that satellite data downloads frequently include a larger area than what is desired, it is necessary to choose a particular section of the larger image to work with. The image is constructed using a scattering of satellite shots that contain both river and land regions. The 512×512 dimension is used to collect each image subset. Photos of rivers are manually taken for clear visibility and to avoid the blockage by clouds. Images from the subset are gathered using Eradas Software.

Figure 3 illustrates how the subcategory of photographs was gathered using the ERADAS software (Fig. 4).

Number of total subset images = 310.

4.3 Image Labelling

Data labelling requires a lot of manual labour and takes a lot of time. Due to the wide variety of ways that photographs can be annotated. These include polygonal segmentation, key-point, 3D-cuboids, Boundary box, semantic segmentation, and landmark. Semantic segmentation is used to identify the river and other classes in the image due to the fact that rivers lack a specific shape and are not regarded as objects.
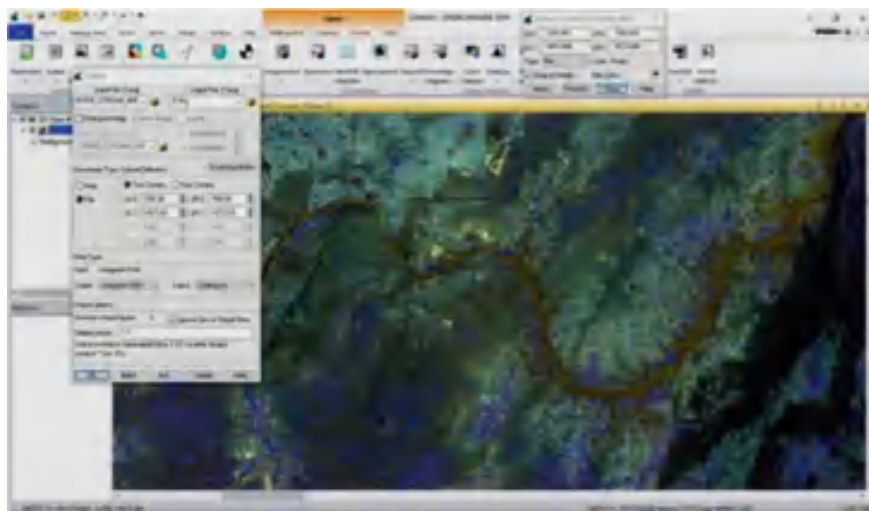
Fig. 3 Subset of image in remote sensing image
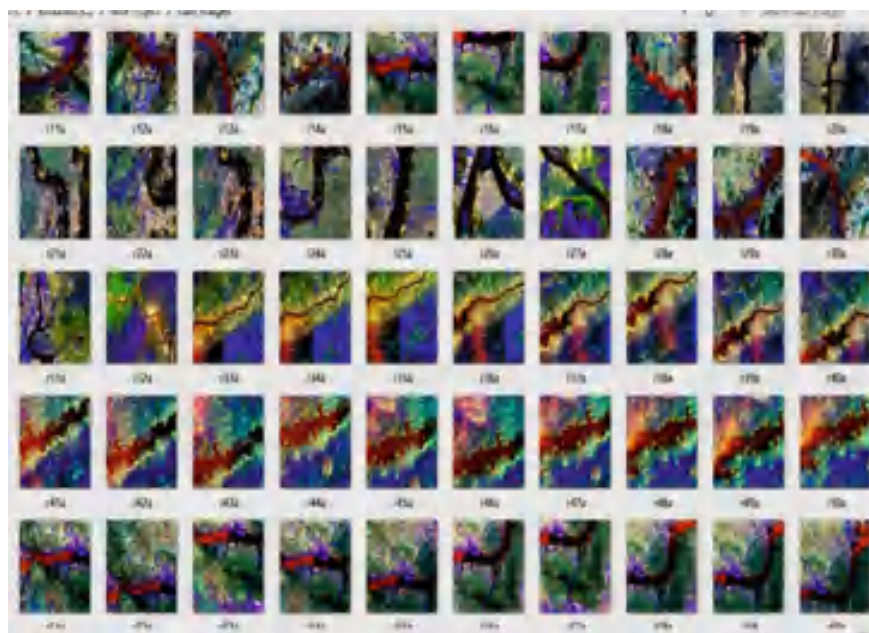
Fig. 4 Sample dataset
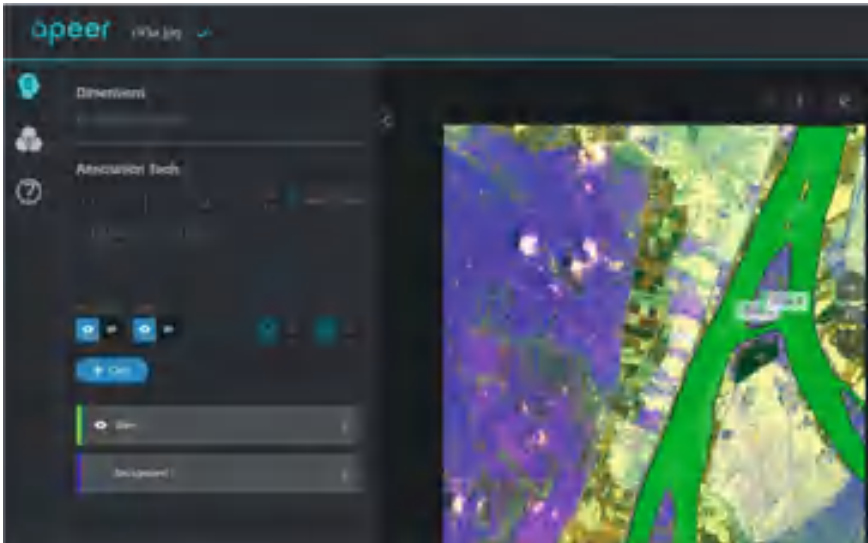
Fig. 5 Image labelling in APEER

4.3.1 Semantic Segmentation

Using semantic segmentation, related portions of the image that belong to the same class are grouped together. Each pixel in the image is given a class by a pixel-by-pixel annotation. Each pixel has a semantic value, and there are two possible classifications: river and non-river.

Users can add datasets to the APEER website and categories them using a variety of categories to facilitate semantic segmentation (Fig. 5).

4.3.2 Mask Images

Manual annotation is used for images. River class and non-river class are the two categories into which the images are divided (Fig. 6).

4.3.3 Image Segmentation Using SVM and Random Forest

A supervised machine learning model is called as support vector machine (SVM) that employs classification techniques to solve two-group classification problems. The common applications include tone recognition, voice recognition, text sorting, object identification, image classification, and handwriting ordinal recognition, are for real-world situations. SVM is one of the most well-known method for categorizing images and patterns. It's intended to divide a collection of training images into two
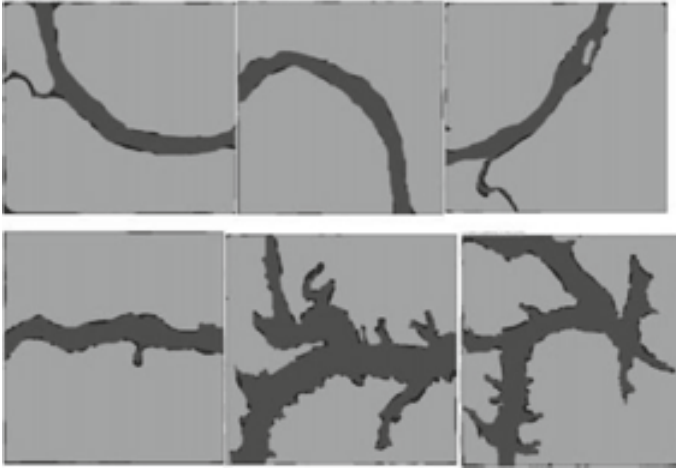
Fig. 6 Sample mask images

distinct classes, $(x_1, y_1), (x_2, y_2),\ldots, (x_n, y_n)$, where x_i is in a d-dimensional feature space in Rd and y_i is in a class label in $-1, +1$ with $i = 1\ldots n$ [1].

Random Forest is a classifier that uses many decision trees on different subsets of the input dataset and averages the results to increase the dataset's predicted accuracy.

Text categorization, image classification, handwritten digital recognition, voice recognition, tone recognition, object detection, and data classification are a few examples of real-world applications.

The Fig. 7 depicts the segmentation using SVM and Random Forest. It begins with analysing training images. Gabor filter and canny edge detection for feature extraction. View Labelled Images (Masks). Prepare your data for SVM. Incorporate our training data into a model. Keep the model for further usage. Using the training model to foretell the test images.

- **Load dataset**

Import the necessary packages first. The software is then loaded with the dataset, which includes photos of rivers. TheCv2.imread() is used to read the images from the specified path.

- **Create a Data Frame**

A Data Frame was created to store the pixels from the original image. This is the first Feature of the programme.

- **Extraction of Features**

Gabor Filter is used to extract other features, and edge detectors like Sobel Edge Detection, Prewitt Edge Detection, Canny Edge Detection, Roberts Edge Detection,
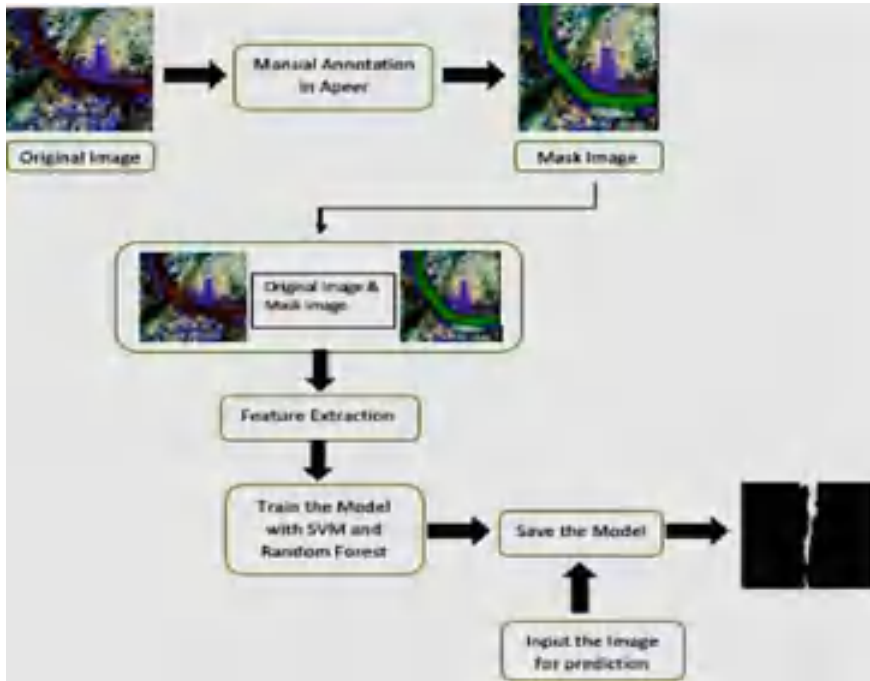
Fig. 7 River segmentation using SVM and random forest

and Roberts Edge Detection are utilized. The Data Frame now includes all of the extra functionality.

- **Add Mask Images Dataset**

The river mask photos comprising the dataset are loaded into the application. To read the images the Cv2.imread() function is used that were shot along the designated path. The Data Frame is then updated to include these image pixels.

- **Testing and Training**

The definitions of the independent variables follow. To ensure accuracy, split the data into train and test groups after the model has been trained. The dataset is divided into different ratios, such as 60:40, 70:30, and 80:20 which are used for training and testing sets as the input.

- **SVM Model and Random Forest**

Random Forest and support vector machine (SVM) are used to train the rivers data.

- **Save the Model**

In case it is required again, the model is saved. The model is first tested on test images before being deployed on unknown images. In order to use it in the future to identify additional pictures, save the acquired model as a pickle string to a disc.

5 Results and Discussion

The dataset includes 310 satellite images taken from National remote sensing centre (NRSC). The performance measures considered are–accuracy and kappa coefficients. Based on the input or training data, accuracy is the metric used to discover which model performs best at finding links and patterns between variables in a dataset. The degree of agreement between frequencies of two sets of data gathered on two separate occasions is shown by the Kappa coefficient.

5.1 Support Vector Machine

Outputs: Fig. 8 is the Input satellite image and Fig. 9 is the river network predicted image by SVM.

Figure 10 is the Input satellite image and Fig. 11 is the river network predicted image by SVM.

5.2 The Random Forest

Figure 12 is the Input satellite image and Fig. 13 is the river network predicted image by Random Forest

Fig. 8 .

Fig. 9 .

Fig. 10 .
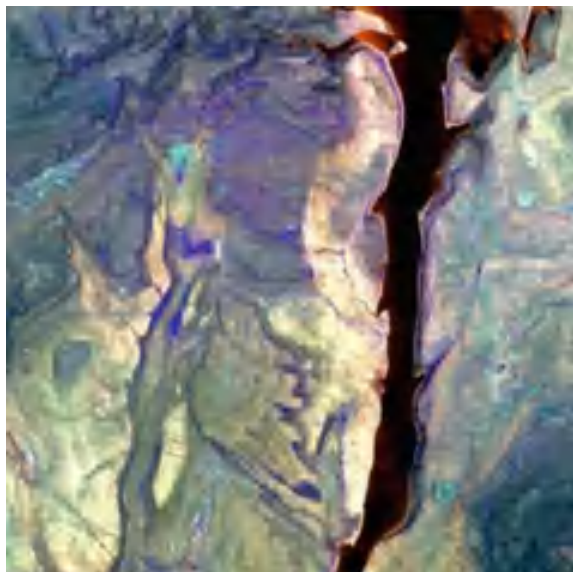
Figure 14 is the Input satellite image and Fig. 15 is the river network predicted image by Random Forest.

By comparing Random Forest algorithm and support vector machine algorithm. Random Forest gives a better accuracy. With a combination of numerical and category information, Random Forest performs well. Thus by applying all the relevant

Fig. 11 .

Fig. 12 .

feature scaling, extraction algorithms the Random Forest and SVM, Random Forest predicting the river network with high accuracy (Tables 1, 2 and 3).

By comparing Random Forest algorithm and support vector machine algorithm. Random Forest gives a better accuracy. With a combination of numerical and category information, Random Forest performs well. Thus by applying all the relevant

Fig. 13 .
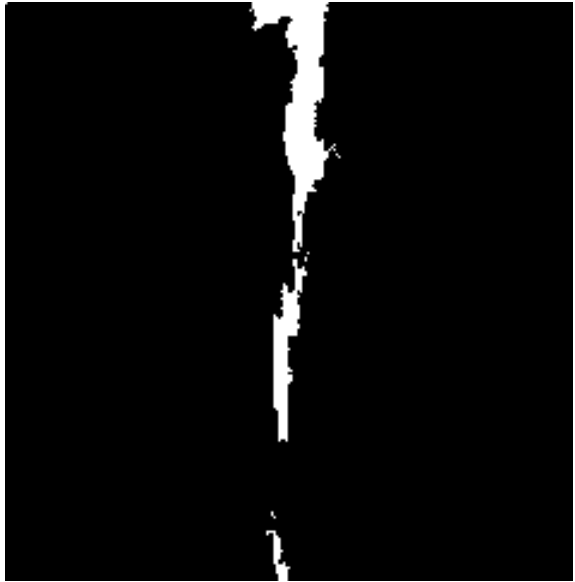
Fig. 14 .
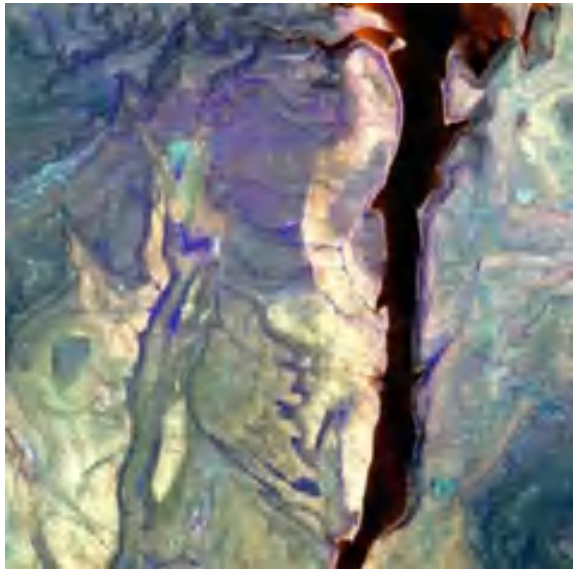
feature scaling, extraction algorithms the Random Forest and SVM, Random Forest predicting the river network with high accuracy (Fig. 16).

Fig. 15 .

Table 1 Accuracy of support vector machine (SVM)

Number of estimators	Train split ratio (%)	Test split ratio (%)	Accuracy (%)
100	60	40	80.01
150	60	40	80.09
200	60	40	82.00
100	70	30	81.89
150	70	30	82.12
200	70	30	82.96

Table 2 Accuracy of random forest

Number of estimators	Train split ratio (%)	Test split ratio (%)	Accuracy (%)
100	60	40	81.36
150	60	40	81.42
200	60	40	82.01
100	70	30	82.39
150	70	30	83.11
200	70	30	83.64

Table 3 Comparison of support vector machine and Random Forest for 70–30% train and test split ratio

Number of estimates	Support vector machine accuracy (%)	Random forest Accuracy (%)
100	81.89	82.39
150	82.12	83.11
200	82.96	83.64

Fig. 16 Comparison of support vector machine and random forest for 70–30% train and test split ratio
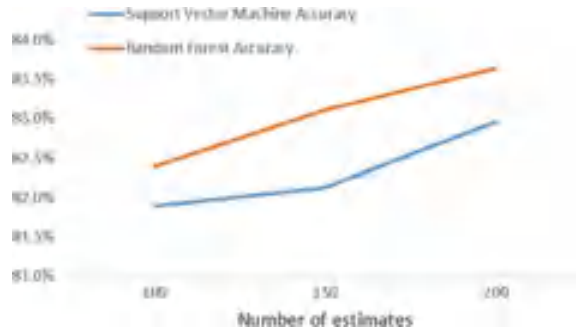
6   Conclusions

The proposed project is used for detection of river networks from satellite imagery using support vector machine and Random Forest classifiers. Using preprocessing techniques such as PCA, Gabor filter and GLCM, and feature extraction filters such as Canny Edge, Sobel, Prewitt Operator, and Roberts Edge Detector. The accuracy obtained for the support vector machine classifier is 82.96% and for Random Forest classifier is 83.64%. Random Forest is observed to give better results when compared to support vector machine. The model can be extended with deep learning techniques in order to reduce the noise. The further enhancement of the project can be done by finding out the width of the river that can be calculated with the knowledge of the geographical coordinates. The proposed model can be used for the detection of rivers and its tributaries.

References

1. Giannini MB, Parente C (2014) An object based approach for coastline extraction from Quickbird multispectral images. Int J Eng Technol (IJET) 6:2698–2704. ISSN: 0975-4024
2. Cai Y, Sun G, Liu B (2005) Mapping of water body in Poyang Lake from partial spectral unmixing of MODIS data. In: IEEE international geoscience and remote sensing symposium (IGARSS'05), vol 32, pp 4539–4549. Seoul, Korea
3. Gao F, Ma F, Wang J, Sun J, Yang E, Zhou H (2018) Visual saliency modeling for river detection in high-resolution SAR imagery. IEEE Access 6:1000–1014. https://doi.org/10.1109/ACCESS.2017.2777444

4. Jain SK, Singh RD, Jain MK, Lohani AK (2005) Delineation of flood-prone areas using remote sensing techniques. Water Resource Manage 19:333–347

5. Harrison JA, Maranger RJ, Alexander RB, Giblin AE, Jacinthe P-A, Mayorga E, Seitzinger SP, Sobota DJ, Wollheim WM (2009) The regional and global significance of nitrogen removal in lakes and reservoirs. Biogeochemistry 93(1–2):143–157

6. Tranvik LJ, Downing JA, Cotner JB (2009) Lakes and reservoirs as regulators of carbon cycling and climate. Limnol Oceanogr 54(6):2289–2314

7. Paravolidaki V, Moirogiorgou K, Ragia L, Zervakis M, Synolakis C (2016) Coastline extraction from aerial images based on edge detection. ISPRS Ann Photogrammetry, Remote Sens Spat Inf Sci III-8:153–158. https://doi.org/10.5194/isprsannals-III-8-153-2016

8. Ryu J, Won J, Min KD (2002) Waterline extraction from Landsat TM data in a tidal flat: a case study in Gomso Bay, Korea. Remote Sens Environ 83:442–456

9. Sharma R, Tateishi R, Hara K, Nguyen L (2015) Developing superfine water index (SWI) for global water cover mapping using MODIS data. Remote Sens 7:13807–13841

10. Huang X, Xie C, Fang X, Zhang L (2015) Combining pixel- and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery. IEEE J Sel Top Appl Earth Observations Remote Sens 8(5):2097–2110. https://doi.org/10.1109/JSTARS.2015.2420713

11. Zhang Y, Liu X, Zhang Y, Ling X, Huang X (2019) Automatic and unsupervised water body extraction based on spectral-spatial features using GF-1 satellite imagery. IEEE Geosci Remote Sens Lett 16(6):927–931. https://doi.org/10.1109/LGRS.2018.2886422

12. Zhou Y, Luo J, Shen Z, Hu X, Yang H (2014) Multiscale water body extraction in urban environments from satellite images. IEEE J Sel Top Appl Earth Observations Remote Sens 7(10):4301–4312. https://doi.org/10.1109/JSTARS.2014.2360436

13. Yang Y, Liu Y, Zhou M, Zhang S, Zhan W, Sun C, Duan Y (2015) Landsat 8 OLI image based terrestrial water extraction from heterogeneous backgrounds using a reflectance homogenization approach. Remote Sens Environ 171:14–32

14. Zhang Z, Zhang X, Jiang X, Xin Q (2019) Automated surface water extraction combining Sentinel-2 imagery and OpenStreetMap using presence and background learning (PBL) algorithm. IEEE J Sel Top Appl Earth Observations Remote Sens 12(10):3784–3798. https://doi.org/10.1109/JSTARS.2019.2936406

Implementing Digitalization in Square Blockchain Technology Using HDS Methodologies

S. L. Aruna Rao, P. S. Latha Kalyampudi, D. Suguna Kumari, A. Aruna Jyothi, and B. Srinivasulu

Abstract A square chain is a digitized, decentralized, open record of all cryptographic cash trades. Persistently creating as "completed" squares are recorded and added to it in consecutive solicitation, it licenses exhibit individuals to screen propelled cash trades without central recordkeeping. Each center point gets a copy of the square chain, which is downloaded normally. At first made as the accounting methodology for the virtual cash Bitcoin, square chains—which use what's known as appropriated record advancement. At this moment, the development is chiefly used to affirm trades, inside cutting edge financial gauges anyway it is possible to digitize, code, and expansion in every way that really matters any record into the square chain. Doing so makes a perpetual record that can't be changed; in addition, the record's validness can be affirmed by the entire system using the square chain as opposed to a lone united force. The proposed system is trades that are irreversible and assented to by all people in the framework. Django is used to make the structures of the undertaking and python is used to make the classes Block, Chain, and Transactions. The correspondence between the center points that are made is practiced by Python-Data Base Communication. The yield of this undertaking makes chains that do a couple of trades and the squares are made. The centers made here are decentralized that is the passageway isn't through only a solitary center point. All the centers are related and the affiliation is through taking care of the past hash an impetus in the current center nearby its hash. The security is given with the objective that the centers can't be changed and the data set aside can't be changed. This endeavor gives security and the lesser worth-based charges.

Keywords Bitcoin · SHA-256 · Hashing · Digital signature · Squares

S. L. Aruna Rao · P. S. Latha Kalyampudi (✉) · A. Aruna Jyothi · B. Srinivasulu
Department of Information Technology, BVRIT HYDERABAD College
of Engineering for Women, Nizampet, Hyderabad, India
e-mail: latha.k@bvrithyderabad.edu.in

B. Srinivasulu
e-mail: srinivas.nivas47@gmail.com

D. Suguna Kumari
Department of Computer Science Engineering, GRIET, Hyderabad, India

1   Introduction

Blockchain is a cryptographically ensured about, time-ventured, open and scattered database of each bitcoin trade that has ever occurred on the framework. "Passed on" here infers that the information in the blockchain is conveyed to and recorded by every center point in the framework. There is no one central database. Any customer can insinuate this summary of trades and check unequivocally what number of bitcoins have ever had a spot with a specific area whenever. Thusly the structure is clear, twofold spending is thwarted, and there is no prerequisite for a trusted in central force. The bitcoin blockchain is reinforced by the undertakings of various excavators customers who put handling ability to the task of dealing with a numerical issue that will compensate them with bitcoins, while all the while making the hashes that ensured the squares of data in the chain. The hashes are cryptographically made strings of data that make the information in the blockchain in every practical sense hard to play with. The cost of intercession assembles trade costs, obliging the base realistic trade size and expelling the open doors for small nice trades, and there is an increasingly broad cost in the loss of ability to make non-reversible portions for non-reversible organizations. With the possibility of reversal, the necessity for trust spreads. Dealers must be cautious about their customers, irritating them for additional information than they would somehow require. A particular degree of deception is recognized as unavoidable. These costs and portion vulnerabilities can be kept up a key good ways from up close and personal by using physical cash, yet no framework exists to make portions over exchanges direct without a trusted in party.

Exactly when two customers partake in a bitcoin trade, information about this trade gets impart from their wallets to one another customer (center point) in the framework. This information is painstakingly stamped and time-ventured, so any individual who sees it understands who sent the money, who got it, how much money it was, and when. At the point when the center points have looked at the information and transaction (that is, watched that everything is real), they each update their copy of the blockchain on their PCs to consolidate this new data.

The data gets squeezed into a square close by data of various trades that are going on around a comparable second. A square takes after a 1 megabyte pile of successively mentioned information about trades. These squares will get related together orchestrated by creation to shape the blockchain.

The Fig. 1 explains, at the point when an additional swap over or an alternate to a current swap over comes in to a blockchain, by and huge a larger part of the hubs inside a blockchain usage must execute calculations to assess and check the historical backdrop of the individual blockchain hinder that is proposed. In the event that a lion's share of the hubs go to an accord that the history and mark is substantial, the new square of exchanges is acknowledged into the record and another square is added to the chain of exchanges. On the off chance that a lion's share doesn't surrender to the option or adjustment of the record section, it is denied and not added to the chain.
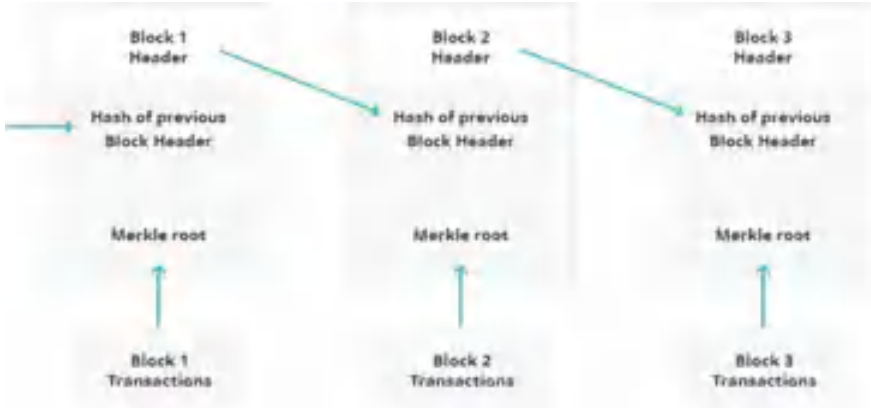
Fig. 1 Block diagram

Each blockchain square involves:

- certain data
- the hash of the square
- the hash from the past square

From Fig. 2, the data set aside inside each square depends upon the sort of blockchain. For instance, in the Bitcoin blockchain structure, the square keeps up data about the recipient, sender, and the proportion of coins.

A hash takes after a one of a kind imprint (long record containing certain digits and letters). Each square hash is made with the help of a cryptographic hash figuring (SHA 256). Subsequently, this helps with recognizing each square in a blockchain structure with no issue. The second a square is made, it normally adds a hash, while

Fig. 2 Functional diagram of blockchain

any movements made in a square impact the distinction in a hash also. Simply communicated, hashes help to recognize any changes in squares.

A square chain is an open record of all bitcoin trades that have ever been executed. A square is the current part of a square chain which records a couple or the whole of the continuous trades, and once completed, goes into the square chain as unchanging database. Each time a square gets completed, another square is delivered. Squares are connected or linked with each other chains or squares immediately, successive invoke with each square carry a hash of the past square. To use standard banking as a relationship, the square chain takes after a full history of banking trades.

Bitcoin trades are entered successively in a square chain just the manner in which bank trades are. In the meantime, squares, take after solitary bank enunciations. The full copy of the square chain has records of each bitcoin trade anytime executed. It would along these lines have the option to give understanding about real factors like how much worth had a spot with a particular area whenever beforehand.

2 Literature Survey

During starter research we found an article looking at the current status of blockchain research. In it Yli-Huumo et al. present a systematic review of 41 companion kept an eye on papers disseminated up until 2015. From the way wherein they talk, it seems like they could simply find a total of 41 buddy assessed articles now. One of the most interesting things that they raise at the soonest reference purpose of the article is that 80 percent of the articles they found were on the use of blockchain for Bitcoin—computerized money. After a wide preamble to blockchain, they give a survey of the methodology they used for their productive arranging study which is extremely similar to what we are achieving for this current assessment. Their four assessments tend to address: subjects tended to in back and forth movement research, applications delivered for blockchain, recurring pattern research openings, and future orientation for blockchain [1]. They began by outlining out the databases they used to search for their composition, by then depicted their screening method. They by then expelled watchwords and data from abstracts. Despite focuses and conveyance date, they furthermore contemplated the source industry or the academic world— and the geographic zone. Besides, they considered the circulation type: gathering, workshop, journal, book part, etc. Finally, they perceived three unmistakable paper types: blockchain report, blockchain improvement, and blockchain application. A lot of this way of thinking was depicted in expansive information stream charts and tables. In their review, they found around five basic subjects in the blockchain composing: security; most likely the years 2015 and 1016 have seen the ascent of business and educational works being dispersed about blockchain and suppy chain. In the Fig. 4 is presented the yearly duty of writing in the overview. The year 2016 is appropriate with most of the conveyances separated contrasting with it. The prior year, 2015 presents relatively little dissemination.

There is to be seen that the current year 2017, with only 2 months gone into it, speaks to unclear number of conveyances from the sum of the year 2015. What affiliations, and to what affiliations have a spot the makers of the papers, is another information of interest. Who is beginning to stand out in the development application?. In the Fig. 5 is outlined the information about the affiliations. Business Companies are the first in the repeat count anyway amazingly close are Universities also, followed in the third circumstance by Research Centers. Collecting Research Centers and Universities as Academic Organizations, by then the Academic Organization takes the principle position.

The number given of business arranged papers addresses those appearing in business magazines, mechanical gatherings, white papers delineating a business game plan and premonition working papers from Advisory associations. Considering the starting period of progress for the plans of supply chain reliant on blockchain. Unfortunately the users and technology suppliers are so unique in the affirmation of thought of the responses for the particular key applications. While automated money and other financial assets applications are going standard into open use, the supply chain and other logistic applications can be becoming snappier than predicted.

Drawback of Existing Method

The work is more on the essential comprehension of square chain yet when the situation is considered for digital currencies like bitcoin there's significantly more than this to the bitcoin arrange. The center comprehension of square chain including chain of squares and approving honesty is increasingly imperative to be considered in building the blocks.

Proposed Method

The objective of this task is to disclose and to make more clear how is a square chain organized at the very center. There are three divisions in usage: The Message() class, the Block() class, and the Chain(). A message is the essential information compartment. It is fixed when added to a square and has 2 hashes that distinguish it: the payload hash and the square hash. Each message is connected to the past message by means of hash pointers. The approve message technique will guarantee the uprightness of each message, however won't check if the hash pointers are right. This is left to the approve technique in the Block() class. A square can contain $1,…,n$ messages that are connected successively in a steady progression. At the point when a square is added to the chain, it's fixed and approved to guarantee that the messages are accurately requested and the hash pointers coordinate. When the square is fixed and hashed, it is approved by checking the normal versus the real. A chain can contain $1,…,m$ hinders that are connected consecutively consistently. The chain trustworthiness can be approved whenever calling the approve technique, which will call each square's approve strategy and will raise an invalid blockchain special case.

A supervisor () work is furnished to collaborate with the square chain by means of the Terminal/Console. The essential activities are: Add Message to Block, Add Block to Chain, Show Block, Show Chain and Exit.

3   Methodology

3.1   *Hashing and Digital Signature (HDS)*

A typical use for hashes today is to unique mark records, otherwise called check zones. This implies a hash is utilized to check that a document has not been messed with or adjusted in any capacity not proposed by the creator. On the off chance that WikiLeaks, for instance, distributes a lot of records alongside their MD5 hashes, whoever downloads those documents can check that they are really from WikiLeaks by computing the MD5 hash of the downloaded records, and on the off chance that the hash doesn't coordinate what was distributed by WikiLeaks, at that point you realize that the record has been adjusted somehow or another (Fig. 3).

Hashes are utilized in blockchains to speak to the present status of the world. The information is the entire the condition of blockchain, which consists all the swapping
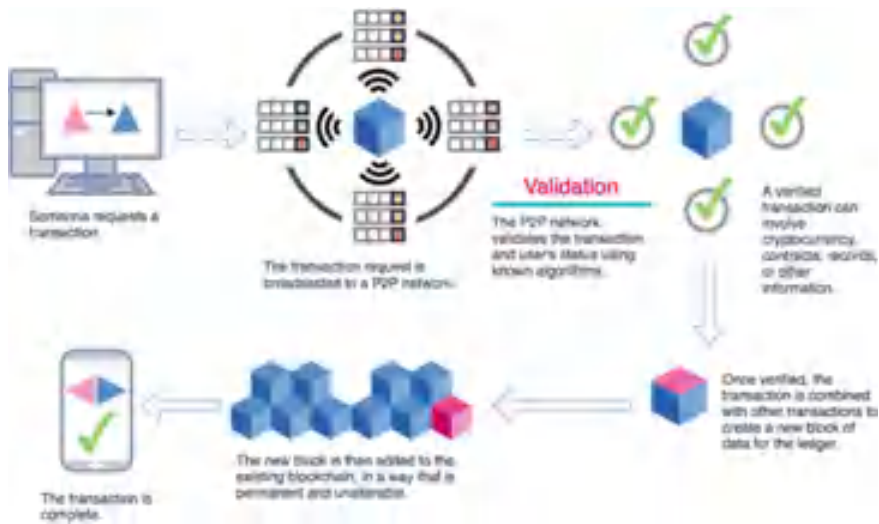
Fig. 3   Blockchain authentication process
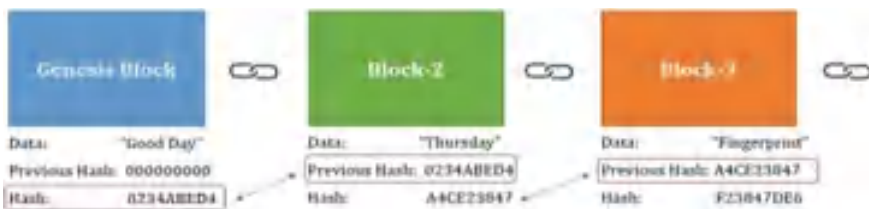
Fig. 4   Hash work use in blockchain
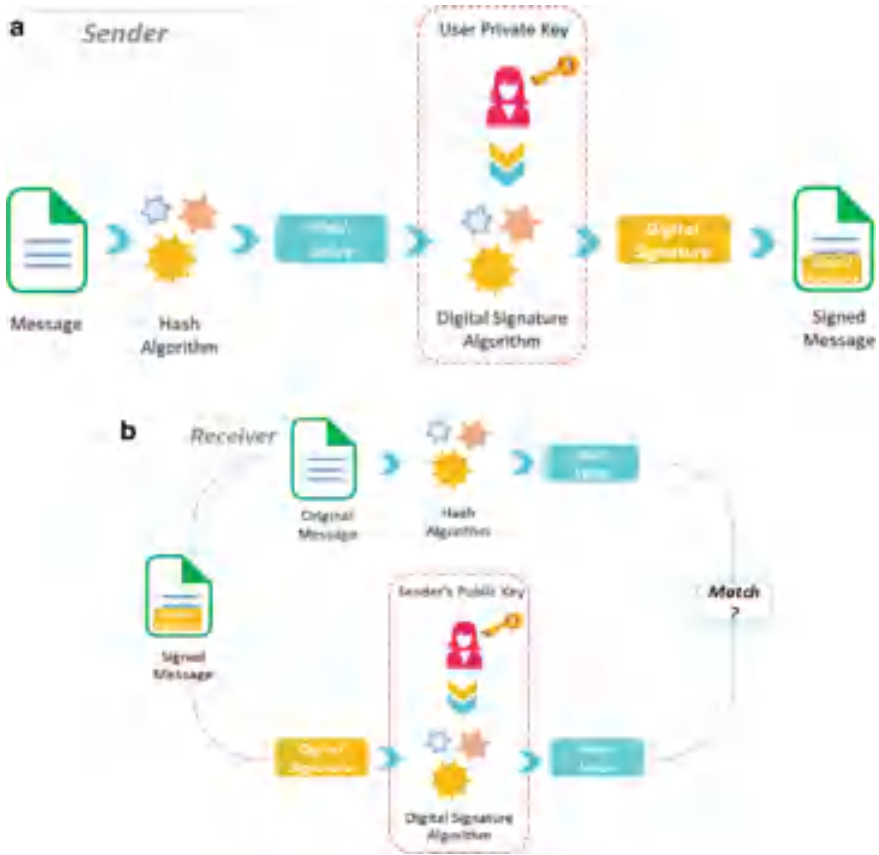
Fig. 5 a Digital signature and hash function work use in blockchain at sender side, **b** digital signature and hash function work use in blockchain receiver side

that have detected until now and the subsequent yield hash speaks to the now status of the blockchain. The hashing is used to concurrent between all collected that the world state would one say one is in the same. Figure 4 shows the hashing function procedure.

Advanced marks, similar to genuine marks, are an approach to demonstrate that someone is who they state they are, then again, actually we use cryptography or math, which is more secure than manually written marks that can be effortlessly fashioned. Computerized marks, in Fig. 5 are utilized today everywhere throughout the Internet. At whatever point you visit a site over ACTPS, you are utilizing SSL, which utilizes advanced marks to set up trust among you and the server. This implies when you visit Facebook.com, your program can check the advanced mark that accompanied the page to confirm that it to be sure begun from Facebook and not some programmer.

In uneven encryption frameworks, clients create something many refer to as a key pair, which is an open key and a private key utilizing some known calculation.

The open key and private key are related with one another through some numerical relationship. The open key is intended to be disseminated freely to fill in as a location to get messages from different clients, similar to an IP address or personal residence.

The private key is intended to be left well enough alone and is utilized to carefully sign messages sent to different clients. The mark is remembered for the message with the goal that the beneficiary can confirm utilizing the sender's open key. Along these lines, the beneficiary can be certain that solitary the sender could have sent this message. Producing a key pair is comparable to making a record on the blockchain, yet without having to really enroll anyplace. Entirely cool likewise, every exchange that is executed on the blockchain is carefully marked by the sender utilizing their private key. This mark guarantees that solitary the proprietor of the record can move cash out of the record.

3.2   SHA-256 Crypto Algorithm

SHA-256 is a cryptographic hash work that takes a contribution of an irregular size and creates a yield of a fixed size. Hash capacities are ground-breaking since they are "single direction." The SHA-256 hash work is used inside the Bitcoin arrange in two principle ways: Mining and Formation of Bitcoin addresses.

Mining is a procedure by which new coins are brought into the current coursing flexibly of the Bitcoin convention, just as a strategy used to make sure about the Bitcoin arrange. Inside a square is what is known as a square header; to develop a square there are 6 boundaries that must be filled in by the excavator (Fig. 6).

As can been seen from the picture above, all together for an excavator to deliver the past square hash boundary, the square header of the past square should be gotten through the SHA-256 calculation twice, this otherwise called twofold SHA-256. That is:

Fig. 6 SHA-256 crypto algorithm process

$$rvos\,loc\,as = -256(-256\,loc\,(adr))$$

The SHA-256 calculation is likewise used to create the Merkle root, which is then in this way embedded into the square header. A thorough clarification of this part of the Bitcoin convention can be found here: Merkle tree and Merkle root explained. Upon fruitful development of a square, the excavator would now be able to start the mining procedure, wherein another utilization instance of the SHA-256 calculation will introduce itself. In this occurrence, one boundary of the square header, the nonce, is a variable that is changed over and over, and after hashing of the square header utilizing the SHA-256 capacity, if the hash is beneath the objective, the digger is viewed as effective.

Production of Bitcoin addresses: So as to deliver a Bitcoin address, a private key, which is an arbitrarily chosen number, is duplicated utilizing an elliptic bend to create an open key. This open key is then gotten through both the SHA-256 and RIPEMD160 hashing calculations.

$$= IEDI160(-256)\quad()$$

where $K =$ the public key and $A =$ Bitcoin addresses.

As we examined, SHA is an abbreviation for Secure Hash Algorithm, so while SHA2 is the replacement to SHA1, it's a totally unique calculation, or somewhat set of calculations, not a minor departure from the first. SHA1 was created by the US government and is nearer in nature to MD5. It makes message digests, 160-piece (20-byte) hash esteems that are spoken to by 40-digit long hexadecimal strings. SHA2 was additionally evolved by the US government, explicitly the NSA, and is really a group of calculations, six diverse hash works that produce digest/hash benefits of changing lengths: 224, 256, 384, or 512 (Figs. 7, 8, 9).

The most well-known is 256, however classification alluding to the SHA2 family can be confounding: The assortment of SHA-2 hashes can prompt a touch of disarray, as sites and creators express them in an unexpected way. On the off chance that you see "SHA-2" , "SHA-256" or "SHA-256 piece," those names are alluding to something very similar. In the event that you see "SHA-224" , "SHA-384" or "SHA-512" those are alluding to the other piece lengths of SHA-2.

4 Implementation

SHA-256 (secure hash algorithm, FIPS 182–2) is a cryptographic hash function with digest length of 256 bits. Here to implement SHA-256 Algorithm we can follow the following steps: Padding bits, Padding Length, Installing the buffers, and Comparison functions. In padding bits consists combination of original bits and some additional bits. The total length is 64 bits. To implement SHA 256 in python we are using hashlib library or module, in this module we can use.encode() and.hexdigest() methods.

Counter (CTR) mode encryption

(a) Encryption

Counter (CTR) mode decryption
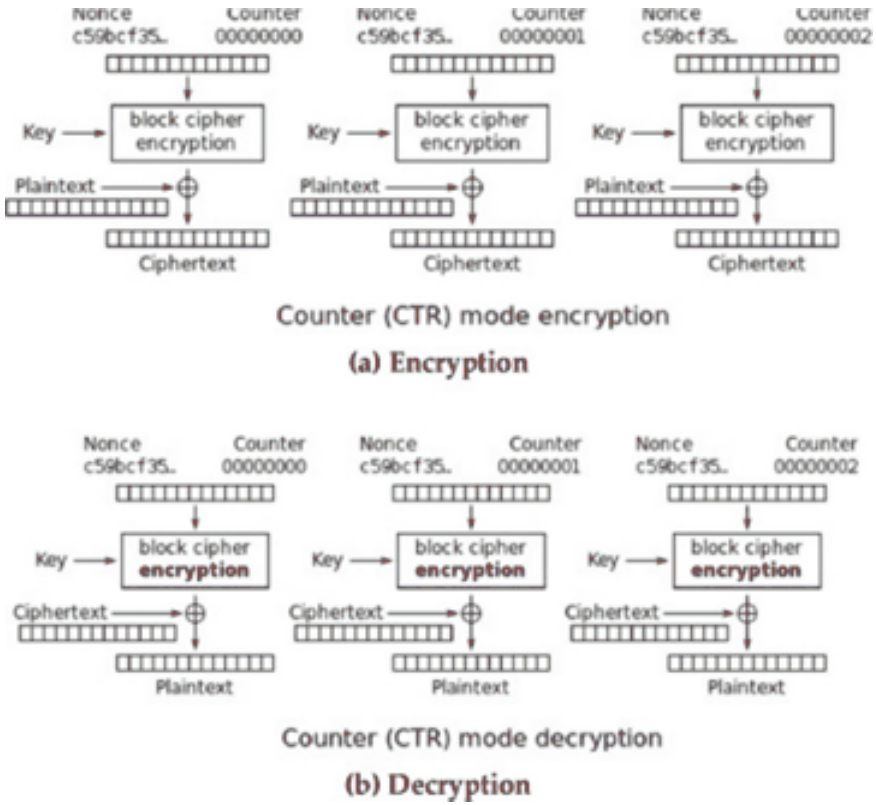
(b) Decryption

Fig. 7 General encryption and decryption in crypto algorithm procedure

Fig. 8 SHA crypto digital
signature generation and
verification
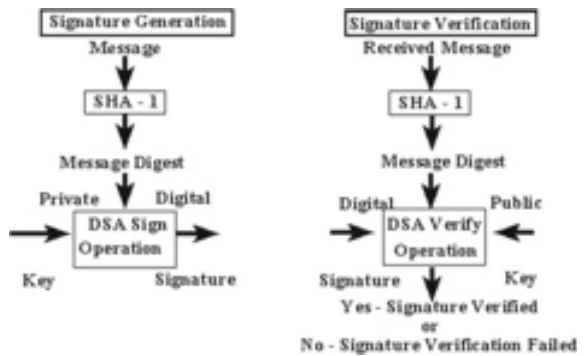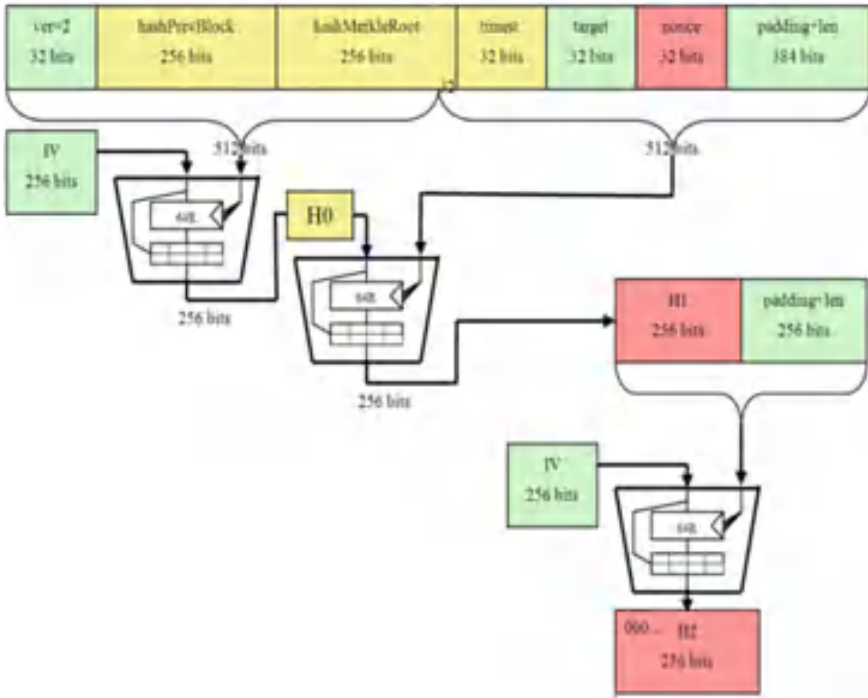
Fig. 9 Bitcoin mining schematic diagram

encode() method converts the string into bytes to be acceptable by hash function and hexdigest() method Returns the encoded data in hexadecimal format.

```
import hashlib.
mx = hashlib. sha512()
mx.update(b"Msg")
var = "Msg"
mx. update(var. encode("utf8"))
hashedMessage = mx. digest()
print(hashedMessage)
```

5 Results

This segment represents the usage of the created crypto-hash decentralized record for dealing with the exchange of oil-based goods dispersion. The crypto-framework was plan for overseeing exchanges information that happen in oil and gas circulation station and the information got data from the big hauler armada while in travel to the goal is distantly moved to the protected database and refreshed right away.

Homepage GUI, it is a login interface for exchange enlistment. Users enrollment page, this is the enlistment page that contained general data concerning outsider in the exchange. First layer login interface, this is a login structure for the transporter. Admin segment for client affirmation, this contains subtleties data about new exchange and confirmation. Admin segment for overseeing transporter geolocation, this put away a far off following data about the exchange for administrator observing in a continuous. GUI for new exchange subtleties, this page put away all the exchange records subtleties incorporate crypto-hash code.

Login page consists of username and password to login to the page that consists of attributes to create the blocks, chains, and transactions. If the user is unauthorized or he forgot his password, the login page would show the error message as shown in the figure. After the user has successfully logged in, the page which is shown above is displayed. This pages contains the creation of blocks, chains, and transactions. For the authentication and authorization, groups and users are the attributes. The overall page allows the user to create the blockchain with security and authentication. The chain page here allows the user to create chains which are the senders or the receivers who are the main entities for the transactions. This is the chains added page that shows the number of chains added and the names of the chains. This is the block page that allows the users to create the blocks that create the transactions between the chains and creating the communication between the previous chain by adding the previous hash value to the present transaction. This is the blocks added page. Here the user can access the created block. For example, Block 4 on b3 conveys that 4 is the index of the block added and b3 is the receiver of that transaction. This is the transaction page. The user creates the transaction between any sender and receiver. The transaction here is provided with security and the transaction is done very fast here. This is the transaction page, this page show the user the transactions that has been created. The every transaction along with sender, receiver, and the money transferred. This is the group page. The user here can select the permissions to the new users. This allows the admin for the safety of the page. This provides the admin the security and the authentication. The crypto-hash database framework was created and execution assessment was done in the Mozilla web designer device. The framework execution measurements was assessed base on the speed of the web application created which incorporates document type, size, time, and the exchange speed as showed in Figure. The document transmission rate is a component of the kind of record being moved in the framework, which creates the outcome given of the prepared and void store (Fig. 10).

6   Conclusion

We have projected a foundation for electronic exchanges outside contingent upon trust. We started with the standard structure of coins created utilizing advanced marks that gives complete control of ownership, yet is imperfect outside an approach to thwart binal spending. To solve this, we projected a delivered system applying
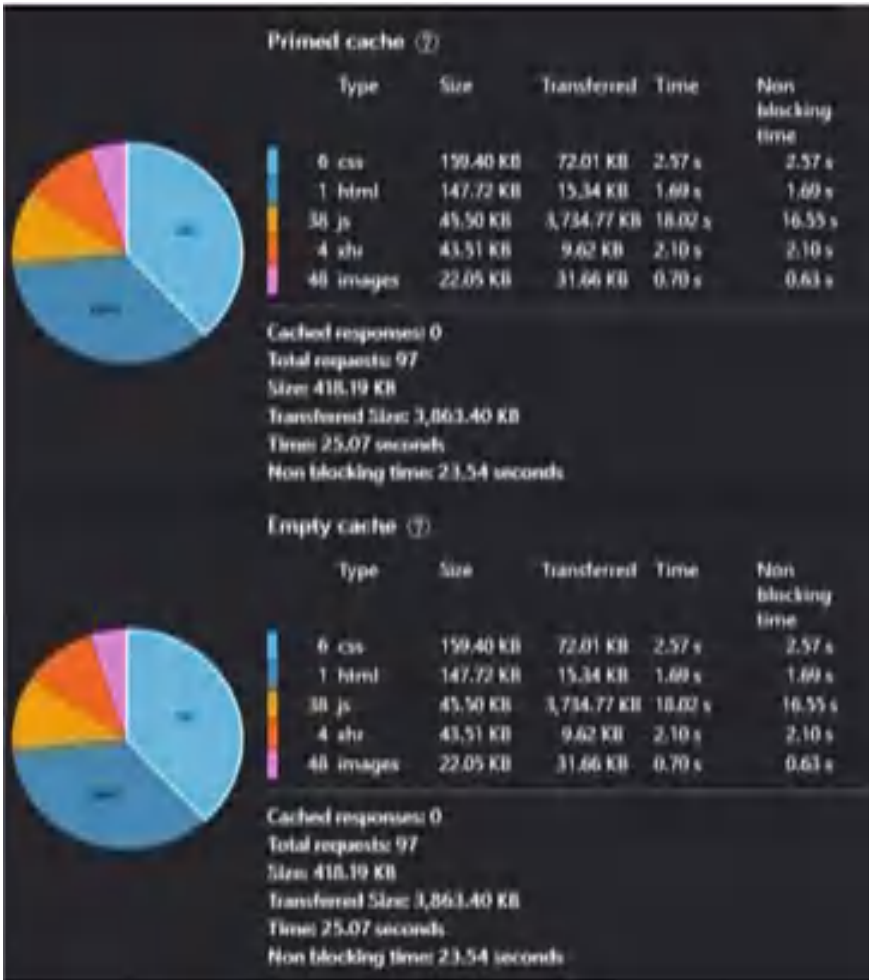
Fig. 10 Performance evaluation

validation of work to record an open past of exchanges that rapidly produces expected computationally groundless for an assailant to change if fair centers control a most of Computer power. Bureaucracy is energetic in allure unstructured honesty. Centers undertake the same time accompanying little arrangement. They forbiddance should do, because ideas are not directed to a distinguishing spot and just bear be sent on a best endeavor premise.

References

1. Kshetri N (2017) Can blockchain strengthen the internet of things? IT Professional 19(4):68–72. Available: http://ieeexplore.ieee.org/document/8012302/
2. Miraz MH (2017) Blockchain: technology fundamentals of the trust machine. Mach Lawyering, Chin Univ of Hong Kong. Available: https://doi.org/10.13140/RG.2.2.22541.64480/2
3. Tapscott D, Tapscott A, Revolution B (2016) How the technology behind bitcoin is changing money, business, and the world, 1st edn. Penguin Publishing Group, New York, USA
4. Ali M, Miraz MH (2013) Cloud computing applications. In: Proceedings of the international conference on cloud computing and eGovernance. ICCCEG 2013, Dubai, pp 1–8. Available: http://www.edlib.asdf.res.in/2013/iccceg/paper001.pdf
5. Ali M, Miraz MH (2014) Recent advances in cloud computing applications and services. Int J Cloud Comput (IJCC) 1(1):1–12. Available: http://asdfjournals.com/ijcc/ijcc-issues/ijcc-v1i1y2014/ijcc-001html-v1i1y2014/
6. Liang X et al (2017) ProvChain: a blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In: Proceedings of the 17th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGrid '17), Madrid, Spain, pp 468–477. Available: https://dl.acm.org/citation.cfm?id=3101176&CFID=994896989&CFTOKEN=44228545
7. Miraz MH, Ali M, Excell P, Picking R (2015) A review on internet of things (IoT), Internet of everything (IoE) and internet of nano things (IoNT). In: The proceedings of the fifth international IEEE conference on internet technologies and applications (ITA 15). Wrexham, UK, pp 219–224. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7317398
8. Miraz MH, Ali M, Excell PS, Picking R (2018) Internet of nano-things, things and everything: future growth trends. Future Internet
9. Miraz MH, Ali M (2018) Blockchain enabled enhanced IoT ecosystem security. In: proceedings of the first international conference on emerging technologies in computing 2018 (iCETiC '18), London, UK
10. Underwood S (2016) Blockchain beyond bitcoin. Communications of the ACM 59(11):15–17. Available: https://doi.org/10.1145/2994581
11. Gartner (2017) Top trends in the Gartner hype cycle for emerging technologies. Gartner, Inc., Gartner Hype Cycle. Available: http://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/
12. Ajao LA, Adedokun EA, Nwishieyi CP, Adegboye MA, Agajo J, Kolo JG (2018) An anti-theft oil pipeline vandalism detection. Int J Eng Sci Appl 2:41–46
13. Vestergaard C (2018) Better than a floppy the potential of distributed ledger technology for nuclear safeguards information management, innovative approaches to peace and security from the stanley foundation. Anal Brief 1:1–8
14. Lavanya BM (2018) Blockchain technology beyond bitcoin: An overview. Int J Comput Sci Mob Appl 6:76–80
15. Aminu SA, Olawore OP (2014) Empirical investigation of challenges of distribution of premium motor spirit (PMS) in federal capital territory (Fct), abuja and environs Nigeria. Int J Manag Sci Humanit 2:17–59

Comic Character Recognition (CCR): Extraction of Speech Balloon Context and Character of Interest in Comics

R. Baskaran and **M. S. Karthika Devi**

Abstract Document analysis is an active field of research which can attain a complete understanding of semantics of a given document. Further, digitalization of physical forms of various data has become vital in today's data extraction. Comic digitalization is becoming widespread as it is one of the easily understandable graphic content and is attention seeking from early education readers to middle aged groups. Extracting the frames, panels, and speech balloons from digital comic is crucial for techniques that felicitate comic reading. However, Automatic Panel Extraction for Digital Comic is challenging, largely because of its layout design, visual symbols, speech balloons attached to almost all the panels, throughout the page. In this proposed work, it is proposed to automatically extract panels from digital pages using contour analysis and watershed canny operator. The first method identifies the difference between frames associated with the panels, whereas the second method identifies the difference in the color between the panel and the gutters. Speech balloons are segmented by methods of K-means clustering and contour analysis. K-means clustering is used to identify the closest related components, and contour analysis is used to differentiate the speech balloons from the other components in the comic panel. Text Area Recognition is a subtle approach that is implemented by Optical Character Recognition (OCR). And finally, comic character extraction has three steps of annotating the object, training the model, and detection of the comic character. The annotation is performed by bounding box algorithm, followed by training of the custom model using a pre-trained YOLOV3 algorithm. Once the custom model is trained, it is provided with comic strips as input to detect the dominant characters with their probability of confidence, which is the correctness of comic character. The implementation considerably serves better in performance when compared to the previous traditional methods of comic component analysis and extraction. Evaluation on the same has also been performed for the trained as well as input dataset.

R. Baskaran · M. S. Karthika Devi (✉)
Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India
e-mail: mskarthikadevi2014@annauniv.edu

R. Baskaran
e-mail: baaski@annauniv.edu

1 Introduction

Comic is the most attention gaining graphic and text content that is popular in the European countries, Japan, Korea and is wide spreading in various other places. They provide deep idea of simple social as well as other issues that need to be made aware in a faster rate. The digital forms of these comics are more easy and simpler to handle. While context extraction and text separation are one of the active works in document analysis, one of the key challenges in comics processing is the element extraction such as panels, balloons, text or processing the comic characters, which are much more complex that the other elements tail detection plays a major role in identifying the protagonist of the respective context. Another interesting element is the explosion balloon, whose irregularity in shape makes it tedious for extraction.

Comics have become a source of portraying news and any formal message through a graphical gesture, through which the reader or the pedestrian reading it from new boards, gets a better clarity of it. The main work behind this transformation is digitalization.

The transformation of physical to digital content has become a widespread process of interest and research today. Digitalization of comics has a critical role of understanding and extracting the components of the strip. These components include panels, speech balloons, text, and comic character detection. In the real time, this type of document analysis is on the run for research, and various advanced algorithms are implemented to produce accurate and true results.

2 Literature Review

Comic strips differ from traditional document analysis and are to have a different flow of analysis and interpretation when compared to that of a normal image processing document. The components of the document such as panels, speech balloons, text, and comic characters are studied and algorithm based on how the components can be extracted are analyzed and implemented in the proposed model.

2.1 Detection of Comic Components

The components are applied through different algorithms and have different methodologies that are studies from works of various authors. The extractions here include various components of a comic strip. A comic strip as a whole is hard to analyze;

therefore, we go for further breakdown of the comic components. Panels are the boundaries in a comic beyond which we find the content of interest. Speech balloons are considered a key element in comics. They play a major link between graphic and textual elements and part of the graphical style of the albums or series. Followed by the main role players are the comic characters, as the entire storyline is based on these comical characters. Considering speech balloons, they can have various shapes, contours and are surrounded by a closed black line. Speech balloons have various cases of completely closed, open, or half closed. There are also studies and work on special speech balloons that demonstrated over the expressions and feeling expressed by a protagonist in a comic panel and difference in their semantics.

Rigaud et al. [1] proposed a method to recognize the panels and text based on the connected components by adding some other features such as the topological and spatial relations. He proposed an adaptive binarization process based on minimum connected component thresholding followed by a text or graphic separation based on contrast ratio and text line grouping. Supervised learning and unsupervised learning have equal share of methods in the part of identifying the various components of a comic. While there are algorithms where components are trained on various features of labeled data, we have components by training the model on unlabeled data too. Li et al. [2] proposed unsupervised speech text localization for comics based on the training of a Bayesian classifier on aligned connected components and then detecting the rest of the text using the classifier for text or non-text separation. Devi et al. [3] incorporated deep learning approaches to analyze the key components contained in comic books. Also explored automatic extraction of comic elements.

Several techniques have been developed to automatically extract panels [4]. Most of them are based on white line cutting with Hough transform [5], recursive X-Y cut [6], or density gradient. These methods do not consider empty area and border-free panels [4]. These issues have been corrected by connected component labeling approaches, but these approaches are sensitive to regions that sometimes connect several panels, which potentially increases the detection error rate and it had been dealt by Rigaud et al. [1], Arai and Tolle [7].

The recent interests are on new methods that are restricted to European comics with different background colors [8]. They are based on watershed [9], line segmentation using canny operator and polygon detection [5], and region of interest detection such as corners and line segments [10]. Complex manga layouts are processed using recursive binary splitting strategy to partition the panel block into disjoint panel regions to find the optimal panel position. Recent methods incorporate three types of visual patterns extracted from the comic image at different levels, and a tree conditional random field framework is used to label each visual pattern by modeling its contextual dependencies [11].

Neural networks models are well developed and have significant achievements in most computer vision tasks such as image classification, object detection, object segmentation, and text recognition. These achievements open new avenues to many domains which require the image analysis tasks. Training models for the custom identification object of interest and to bring out profitable information and correctness (probability of right detection) are the area of interest in the present studies.

2.2 Text Area Recognition

Text recognition is challenging based on what kind of dataset that is to be worked upon. This is tedious as the forms they take are different and vary from one comic to another. They vary in style, the font and even languages for instance.

Li et al. [2] proposed unsupervised speech text localization for comics based on the training of a Bayesian classifier on aligned connected components and then detecting the rest of the text using the classifier for text or non-text separation. Rigaud et al. [1] also proposed a method to recognize the panels and text based on the connected components by adding some other features such as the topological and spatial relations. He proposed an adaptive binarization process based on minimum connected component thresholding followed by a text or graphic separation based on contrast ratio and text line grouping.

The proposed approach requires pre-trained OCR system able to recognize text lines. Tesseract is the most popular OCR systems presently available. It comes with pre-trained data for several languages [12]. Tesseract is considered one of the most accurate free open-source OCR engines. It is open-source software that can be easily integrated into research experiments. Karthika Devi et al. [13] used quote content source identifier to automatically generate comic strips with appropriate characters and speech balloon to convey the moral behind Thirukkural.

2.3 Comic Characters Recognition

Rigaud et al. [1] proposed a deep learning CNN algorithm to focus on various comic characters of the given comic. In the work, he proposed to identify the protagonist based of the existing dataset for the proposed dataset. For the experiment analysis, the proposed work creates a dataset on its own. The English comics are collected from the free public domain collection of comic books. The manual annotations are made by bounding boxes of panels and speech balloons. Rigaud et al. [14] also proposed a knowledge-driven approach of understanding image comics where he did identify various character with their semantic relations in an unsupervised way. The idea here was an iterative approach of how the system is recursively. Nguyen et al. [15] proposed a deep learning-based character detection by training a new dataset based on the available dataset. By this, there is better chances to compare the previous work with the betterment of the future work. Nhu-Van Nguyen et al. also proposed about how comic strips are classified into sub images corresponding to their panels, which allows them in advanced search of various keywords based on comic characters, actions, and various region of interest.

3 Proposed System

3.1 Overall Design

The proposed architecture of the system is divided into three stages that include low level processing, Inference Engine and Validation of results.

The input to the system is a set of comic strip of interest, which is processed through a series of stages. At each stage, every components are analyzed and extracted. The flow of identification of these components begins with the detection of panels in a given comic strip. This is performed majorly through contour detection algorithm. Followed by the extraction of speech balloons along with their tails detection. This is achieved by CNN and k-means clustering. Text extraction is performed by Optical Character Recognition (OCR).

The final stage of processing is to recognize the character if interest. This would include annotating a large dataset to form the character of interest. This is done through labeling the comic strips with the character of interest. Then, training the annotated dataset using CNN and further selection of the character based on the labeling of the comic are performed. Thus, the components of the comic strips are extracted.

The architecture of the proposed system and the various algorithms used to implement them are stated below in Fig. 1.

Object Detection

Panel Extraction

A comic strip can be precisely defined as a collection of panels. Panels form the basic cover to the other comic components that include speech balloons, narration boxes, and comic character. The panels are usually divided by dark frames, with subjective space in between known as the gutters. Usually, panels are rectangular in shape, but may vary to authors. Extraction of panel makes it simpler, as now it is also a comic strip with less complexity and small region of interest.

> **Algorithm**
> **Input**: Comic Strip
> **Output**: Comic Panel
> 1. Select the comic strip for which panel extraction is to be performed.
> 2. Apply contour analysis and canny operator algorithm
> 3. Find the min max edges beyond which contour doesn't exist.
> 4. Return the edges as the borders of the identified panels.

Speech Balloon and Tail Extraction

Speech balloons are the subject containing component of a comic strip. Beyond what a graphical image can convey, the speech balloons are an associativity between the comic character and the subject of conveyance to the reader. The speech balloons
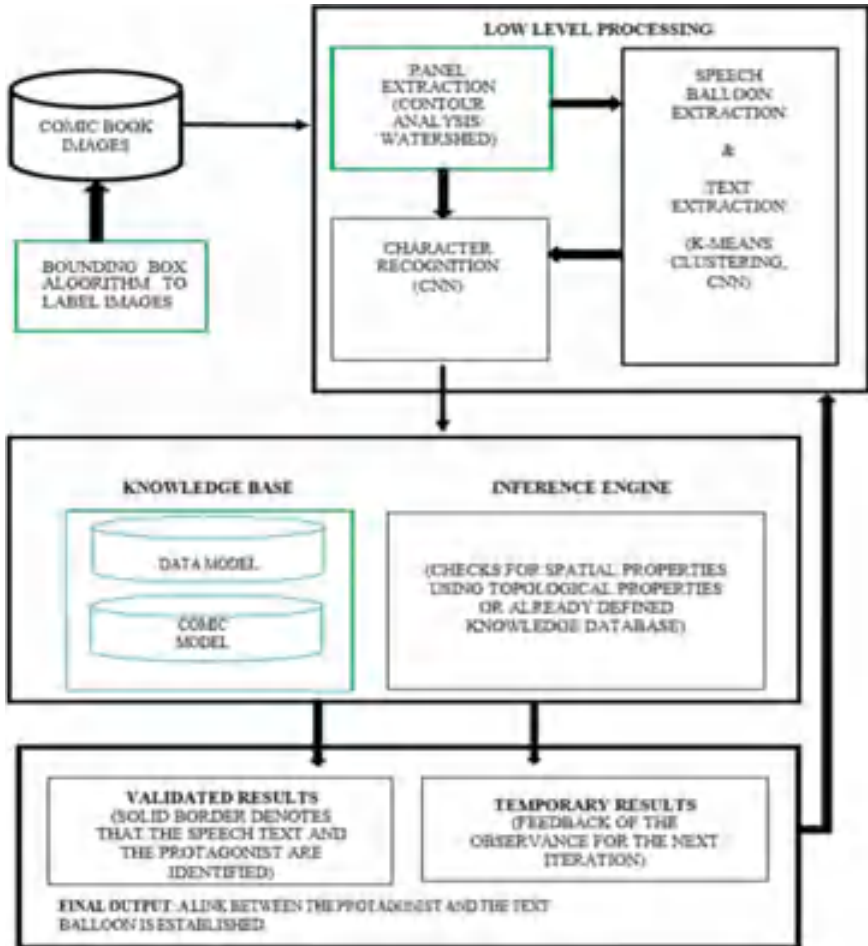
Fig. 1 Proposed architecture for identifying components of a comic

may differ in shape, size, and model based on the author, type of comic, and the situation of the storyline. There are also different speech balloons to demonstrate the various moods that can be portrayed such as anger, shiver, cold, suspense, etc. Therefore, it is very important to know the use of such minimal information regarding the components of various components.

Tail is the end of a speech balloon that denotes the protagonist associated with the speech balloon considered. Protagonist here refers to the comical character who is the character of interest and conveys the subject in the particular panel considered. The tails turn out to bring the relationship between the speech balloon and the protagonist.

In our implementation, we look forward to segmenting the formal speech balloons along with their tails from a given panel as input.

> ***Algorithm***
> ***Input****: Comic Panel*
> ***Output****: Speech Balloons with tail*
> *1. Select the comic panels for which speech balloons need to be extracted.*
> *2. Consider different cases such as no speech balloon, complete speech balloon, open speech balloons, and overlapping speech balloons.*
> *3. Segment the speech comic strip through which the speech balloons are alone identified.*
> *4. Compute this for different panels of different cases.*

Text Extraction Text Area Recognition

Text area refers to the text content in a panel, or precisely the speech balloon. Text area may be the narration boxes too. The text area helps us to inversely find out the speech balloon too. Text recognition is also used in process such as summarization and key words extraction, which is not our region of interest.

> ***Algorithm***
> ***Input****: Segmented images*
> ***Output****: Text identified regions*
> *1. Select the comic panel with the segmented format from where text need to be extracted.*
> *2. Apply Optical Character Recognition algorithm to highlight the text areas in the panels.*

Comic Character Recognition

Comic character recognition is a vital role as they form the major difference from any other normal images. Comic extraction is implemented by the process of training the images with annotated images with dominant characters. By creating a model, now we can detect the characters in any given comic strip of the same characters.

Step 1: Annotating the Dataset

Annotating the dataset is the process of labeling the dataset. When training a model, first it must be made understood of what exactly the comic is built upon. We choose a particular dataset (a set of comic images), and we determine the dominant comical character in that particular comic strip and label them. The output of annotating the comic strip/ panel is a set of xml files. These xml files have the depth, width, and position the particular object that is annotated.

> ***Algorithm***
> ***Input:*** *Comic strips*
> ***Output:*** *XML files of the annotated comic characters*

1. The annotating tool is chosen, and the directories of source and destinations are selected. With this, the various comic strips and images are loaded in the tool.
2. Then choose the rectangle tool and apply bounding box around a comic character.
3. Right click on the bounded area and provide the appropriate label for the particular comic character bounded.
4. XML files are created for each bound box created and the objects bounded can also be visible in the right pane.

Step 2: Training the Model

Training the model is based on YOLO model, which is pre-trained with which we train our model of custom dataset. Training the model is the process where the system, or the algorithm that is used for further character of interest detection is made to understand what are the features upon which it needs to identify a comic character, or differentiate the particular character from the others. When considering only few characters of interest, the model now classifies the different sets of objects and trains further on the similar features of the classified objects.

Algorithm
Input: *XML files with respective Comic strips. Model*
Output: *Trained model*
1. Train the model using YOLO model for the annotated XML files.
2. Provide the iteration up to which the training has to done and the epoch values.

Step 3: Detection of Comic Characters

*Co*mic character detection is the resultant step of the final extraction process. Now that we have a new model from the annotated images and the pre-trained model, this is further used to detect the comic characters of interest in any given comic strip. The output is a bounding box of the detected comic character with the identified name and the probability that the model justifies the comic character.

Algorithm
Input: *Trained Model, Testing data*
Output: *Detected comic characters*
1. Provide the trained model with the testing data, which includes xml files and images.
2. Detection of comic characters is done through solid bound boxes and also in the form of edges returned by the algorithm.

4 Experiment and Result Discussion

4.1 Dataset Description

The dataset is a collection of comic strips collected from the internet (www.google.com). The comic strips are provided as raw images for panel extraction, speech balloon, and tail detection, as well as text extractions. The role is vital and different in case of comic character recognition.

4.2 Experimental Results

The panel extraction is carried out by canny operator and contour analysis where the algorithm scans for edges, minimum and maximum of a rectangular frame and returns the points. The output of the algorithm when run as a crop image results perfectly in the extraction of the panel in the strip. And also the speech balloons along with the tails are segmented and extracted from a given input image is shown in Fig. 2. From the extracted panels, texts are extracted using Optical Character Recognition (OCR), and the same is shown in Fig. 3. The XML files that were annotated are served as input, along with the YOLOV3 pre-trained model for training the new model as per the comic strip dataset. The number of iterations that the model must be trained is provided accordingly as one of the input information. The trained model provided along with the data, the comic characters of dominance are identified and detected as the final output by bounding box is shown in Fig. 4.

Fig. 2 Speech balloon extraction

Fig. 3 Text area recognition

Fig. 4 Comic character detected strips

4.3 Evaluation Metrics

Confusion matrix for binary classification

The evaluation of the comic strips against the comic character recognized at the end of all the modules is performed for dividing the condition binary, on four major cases. As per our dataset of Tom and Jerry comics, the four cases are defined as

True Positive

True Positive is a test case when the object is correctly identified. The True Positive Rate is given by Eq. 1.

$$TPR = TP/P = TP/(TP + FN) = 1 - FNR \tag{1}$$

When the actual character in the strip is either tom or jerry and the detected output is the same with respect to the character.

True Negative

True Negative is a test case when the object is correctly rejected. The True Negative Rate is given by Eq. 2.

$$TPR = TP/P = TP/(TP + FN) = 1 - FNR \qquad (2)$$

When the comic character is neither tom nor jerry and is not identified as one of them with respect to the character.

False Positive

False Positive is a test case when the object is correctly rejected. The False Positive Rate is given by Eq. 3.

$$FPR = FP/N = FP/(FP + TN) = 1 - TNR \qquad (3)$$

When the comic character is neither tom nor jerry but is identified as either of them.

False Negative

False Negative is a test case when the object is wrongly rejected. The False Negative Rate is given by the Eq. 4

$$FPR = FP/N = FP/(FP + TN) = 1 - TNR \qquad (4)$$

When the comic character is tom nor jerry but the model does not identify the object as tom or jerry.

To find the accuracy of cases, we go with the equations:

(i) **Accuracy of True positive**

= (Number of tom or jerry identified correctly)/Total number of tom or jerry in strip.

(ii) **Accuracy of True negative**

= (Number of other characters identified correctly)/(Total number of other characters in strip)

(iii) **Accuracy of False positive**

= (Number of other characters identified wrong)/(Total number of other characters in strip)

(iv) **Accuracy of False positive**

= (Number of tom or jerry identified wrong)/Total number of tom or jerry in strip.

Table 1 shows the accuracy percentage of the results obtained.

Precision Recall

Precision is the fraction of relevant comic characters detected among total rightly identified comic characters, while recall (also known as sensitivity) is the fraction of the total amount of relevant comic characters to those identified to be dominant characters. Both precision and recall are therefore based on an understanding and measure of relevance of the comic characters.

The precision and recall can be calculated from the binary classification data already found. The mathematical representation for precision and recall is given by Eqs. 5 and 6, respectively.

$$\text{Precision} = \text{True Positive}/(\text{True Positive} + \text{False Positive}) \tag{5}$$

$$\text{Recall} = \text{True Positive}/(\text{True Positive} + \text{False Negative}) \tag{6}$$

Table 2 gives the computed results of the precision and recall from the computed data of Table 1. From the dataset we have trained and the model that has been developed, we have got the accuracy results at various stages of the process. By various stages, it means that by increasing the number of training samples (comic strips). Table 1 shows the accuracy percentage of the various samples. The chart demonstrated in Fig. 5 gives a graphical representation and visual comparison of the same. In Fig. 6, it shows the total accuracy of the entire module of character recognition.

Table 1 Evaluation–accuracy percentage

Dataset images	True positive (%)	True negative (%)	False positive (%)	False negative (%)	Total accuracy (%)
20 images	55	27.6	63	11	51
40 images	69	77	13	23	79
70 images	89	92	3	6	95
120 images	94	94.2	3	5	97.5
150 images	94.2	94.2	2.5	5	97.8

Table 2 Evaluation using precision and recall

Comic character detection								
No. of. images	Trained model-1			Trained model-2			Trained model-3		
	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	F1
20	0.0833	0.6043	0.0732	0.01992	0.6415	0.0303	0.0169	0.6333	0.01504
40	0.0750	0.8415	0.0688	0.02154	0.7867	0.06914	0.0376	0.76143	0.04720
70	0.0936	0.9674	0.1061	0.0551	0.8287	0.10992	0.0416	0.8532	0.07143
120	0.0945	0.9690	0.1176	0.0691	0.97250	0.12427	0.0499	0.9670	0.06742
150	0.1049	0.9741	0.1259	0.1052	0.9760	0.20912	0.0612	0.9781	0.13351
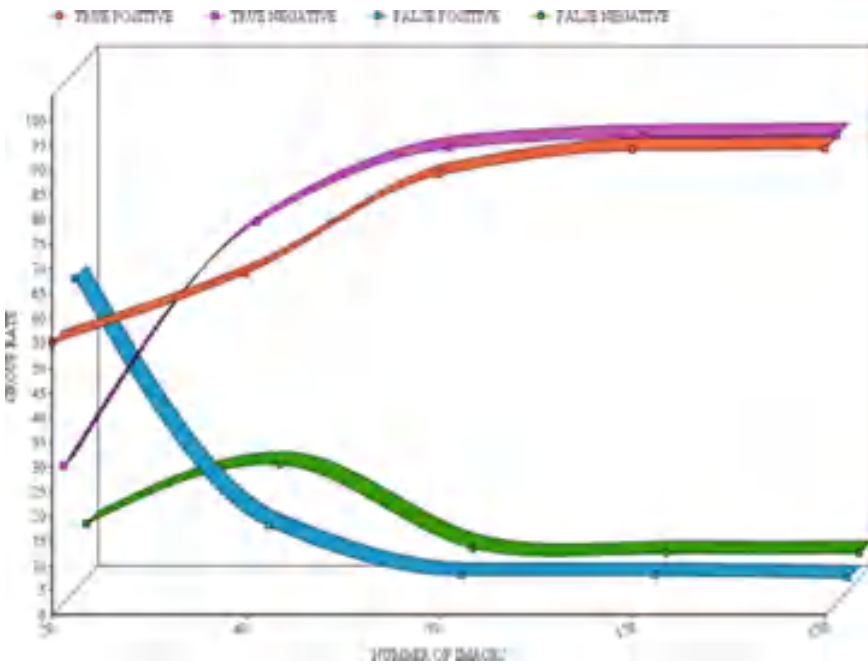
Fig. 5 Evaluation accuracy graph
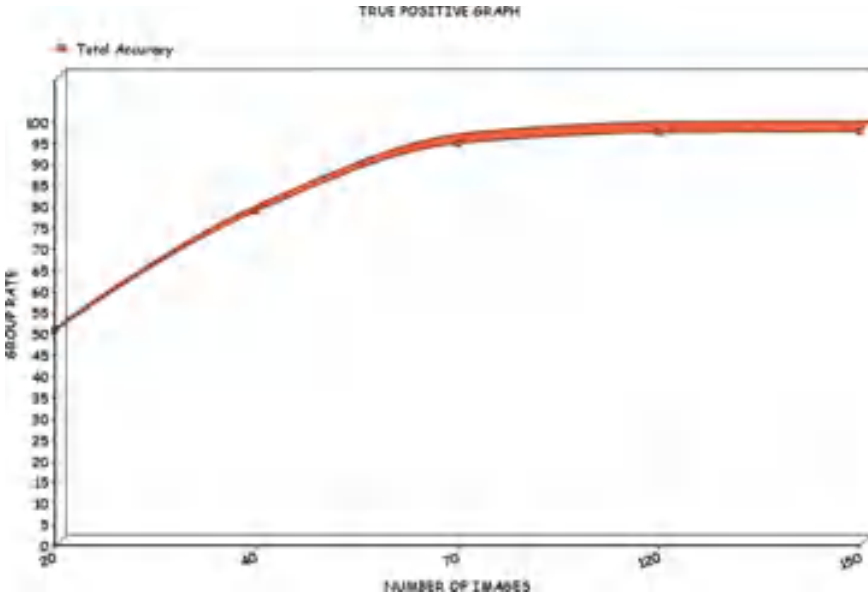
Fig. 6 Total accuracy graph

5 Conclusion

Understanding and analysis the components of comic strips requires a model which quite knows how to differentiate the various components in the comic. This proposed model is used to extract the components such as panels, different cases of speech balloons, text extraction, and comic character recognition for the trained characters of interest. As the proposed work takes more complex dataset with large number of characters, it is a tedious hard to collect large volume of ground truth data and train the comic characters for such data. This is an enormous task which is time-consuming and requires the resources to perform so. The components are now extracted, leading to bring forth the relationship and meaning of these components. Each component in a comic strip can be denoted through various styles that represent different meaning. Thus, the future work can be a solid extension of analyzing different semantics of the components.

References

1. Rigaud C, Tsopze N, Burie JC, Ogier JC (2011) Robust frame and text extraction from comic books. In: Proceedings of international workshop on graphics recognition, pp 129–138
2. Li L, Wang Y, Tang Z, Lu X, Gao L (2013) Unsupervised speech text localization in comic images. In: Proceedings of the IEEE international conference on document analysis and recognition, pp 1190–1194
3. Devi MK, Fathima S, Baskaran R (2020) CBCS-comic book cover synopsis: generating synopsis of a comic book with unsupervised abstractive dialogue. Procedia Comput Sci 172:701–708
4. Takahashi A, Oie T, Hirano K, Higuchi M, Kawasaki S, Koike A, Murukami H (2011) Fast frame decomposition and sorting by contour tracing for mobile phone comic images. Int J Syst Appl Eng Develop 5:216–223
5. Li L, Wang L, Tang Z, Gao L (2014) Automatic comic page segmentation based on polygon detection. Multimedia Tools Appl 69(1):171–197
6. Han E, Kim K, Yang H, Jung K (2007) Frame segmentation used MLP based X-Y recursive for mobile cartoon content. In: Proceedings of the international conference on human-computer interaction, pp 872–881
7. Arai K, Tolle H (2011) Method for real time text extraction of digital manga comic. Int J Image Process, pp 669–676
8. Augereau O, Iwata M, Kise K (2018) A survey of comics research in computer science. arXiv preprint arXiv: 1804.05490
9. Ponsard C, Ramdoyal R, Dziamski D (2012) An OCR-enabled digital comic books viewer. In: proceedings of international conference on computers for handicapped persons, pp 471–478
10. Pang X, Cao Y, Lau WR, Antoni B (2014) A robust panel extraction method for manga. In: Proceedings of the ACM international conference on multimedia, pp 1125–1128
11. Wang Y, Zhou Y, Tang Z (2015) Comic frame extraction via line segments combination. In: Proceedings of IEEE international conference on document analysis and recognition, pp 856–860
12. Smith R (2007) An Overview of the tesseract OCR engine. In: Proceedings of IEEE international conference on document analysis and recognition, vol 2, pp 629–633
13. Karthika Devi MS, Umaa Maheswwari G, Ramachandran B (2022) Dialogue extraction and translation from stories on thirukural using verb cue quote content source identifiers. In: Senjyu T, Mahalle PN, Perumal T, Joshi A (eds) ICT with intelligent applications. Smart innovation, systems and technologies, vol 248
14. Rigaud C, Guerin C, Karatzas D, Burie JC, Ogier JM (2015) Knowledge-driven understanding of images in comic books. Int J Doc Anal Recogn (3):199–221
15. Nguyen NV, Rigaud C, Burie JC (2018) Digital comics image indexing based on deep learning. Proc J Imaging (7)

COVID-19, Normal, and Pneumonia Classification Based on Deep Features Using Transfer Learning

Bipin Bihari Jayasingh and **Talapaneni Jyothi**

Abstract The coronavirus is not the only virus that can cause pneumonia, but pneumonia caused by COVID-19 is more likely to be severe than other types of pneumonia. Pneumonia is a dangerous consequence that occurs when the virus enters the lung tissue of the lower respiratory tract. This may occur when the infection is absorbed. The images of the internal organs of the chest are obtained during an X-ray examination. The main objective of this study is to classify the three classes: COVID-19, normal, and pneumonia. The data set used in this study includes 6432 radiographs. Using transfer learning, image classification for deep features analyses the input image and generates results based on categories. Since deep features are the most important part of medical image categorization, a model that converts the raw image into a format that in-depth features can understand is required. In this study, several deep features are studied by using pre-trained CNN models with transfer learning such as InceptionResNetV2, InceptionV3, and NasNetMobile. Accuracy, precision, recall, sensitivity, specificity, and AUC are the few metrics used to check the model's efficiency. The Xception performs better at classifying COVID-19 with 98.26% accuracy. The InceptionResNetV2 model achieved the highest overall accuracy of 92.80% for pneumonia and normal classes. The model concludes that it correctly categorized the diseases in 92.80% of pneumonia and normal classes. The proposed technique is useful in clinical practice and helps physicians identify diseases from chest radiographs. This enables physicians to help patients promptly.

Keywords COVID-19 · Deep features · Image classification · Medical images · Normal · Pneumonia · Transfer learning and performance validations

B. B. Jayasingh (✉) · T. Jyothi
IT Department, CVR College of Engineering, Hyderabad, India
e-mail: bipinbjayasingh@cvr.ac.in

T. Jyothi
e-mail: 20B81DB007@cvr.ac.in

1 Introduction

COVID-19 is a disease caused by the cov-2 virus. Patients with the virus will not recover without special therapy. Some will need medical assistance and may become critically ill. Serious diseases are more likely to affect people over the age of 65 and those with underlying medical conditions like cancer, diabetes, chronic respiratory disease, or cardiovascular disease. Anyone at any age can die from COVID-19.

A virus can cause pneumonia. Pneumonia affects tiny air sacs called alveoli. A dry cough, chest pain, and difficulty breathing are symptoms of pneumonia. They performed diagnostic procedures such as chest radiographs, blood tests, and sputum cultures to identify the disease and confirm its existence or absence. Risk factors for pneumonia include asthma, diabetes, and an impaired immune system. Researchers have developed vaccines for several types of pneumonia. If the patient's condition was too critical, oxygen therapy, which enhances breathing, would be needed.

Computer vision techniques classify images based on their visual content, but Jayasingh et al. [1] have visualized using big data analytics. An image classification system may look for pictures. The researchers developed vaccines against a variety of types of pneumonia. Pixels receive land cover classifications thanks to image categorization. The image classification steps include the objective, and the characteristics of the image data should determine the classification classes. Colour, textures, multi-spectral or multitemporal characteristics, and other characteristics should distinguish classes. The training set of data is used to select classification competencies like supervised and unsupervised learning. Find a decision rule using training data, a decision rule for a later classification will be selected: the decision rule groups all pixels into a class. Segmentation by field and pixel by pixel is available. Analysis of results helps determine the accuracy and reliability of categorized results. A pre-trained feature is a layer's recurrent response to an input relevant to the final output of the model in a hierarchical model. Based on the first frame, they answered, one characteristic is considered "deeper" than another. They realize that learning is transferred when new knowledge is used in new situations. The transfer of learning is the enhancement of learning in a new task through the transfer of knowledge from a related task already learned. Transfer learning is an optimization that allows quick development when modelling the second task.

This paper is further comprised: Sect. 2 deals with related work, Sect. 3 elaborates upon the method used for the following proposed workflow, Sect. 4 shows the proposed work results, Sect. 5 deals with the conclusion for better analysis, and classification of pneumonia, normal, and COVID-19 diseases via CXR scans.

2   Related Work

Umar Ibrahim et al. [2] developed a methodology for diagnosing viral thoracic infections and COVID-19. First, scientists built a machine that can identify viral pneumonia from radiographs. Two fully interconnected layers and transmission learning are utilized to educate the computer's AI engine. Rahim Zadeh et al. [3] developed an updated in-depth NBC to detect COVID-19. ResNet50V2, Xception, and a mixture of the two generated the pattern. Nada M. Elshennawy et al. developed a pneumonia detection through chest X-ray scans. Randomly splitting sample photos established training and validation categories. CNN, RNN, LSTM, ResNet152V2, MobileNetV2, and LSTM-CNN are recommended. Mporas et al. [5] identified COVID-19 from chest X-rays. 2905 X-ray pictures were classed as "normal", "COVID-19", and "viral pneumonia" 2002 pictures data set includes 300 * 400 chest X-rays. CNN pre-trains DenseNet, ResNet, and Xception. Sharma et al. [6] used X-rays to diagnose COVID-19. Kaggle initially provided healthy and unwell chest X-rays. The images shows the negative and positive COVID-19 results with an 1024-pixel pictures introduced the models using CNN pre-tained models like VGG16 and VGG19 etc. Shah et al. [7] used convolutional neural networks and Kaggle photos to diagnose pneumonia. Each diagnosis is based on publicly available radiographic chest data. X-rays are grayscale. CNN is recommended. Sai Krishna et al. [8] examined deep learning techniques for pneumonia. Chest X-rays may generally detect the problem. A diagnosis can be subjective for diseases not seen on chest X-rays or complicated with other diseases. Clinicians require computer-aided diagnostics.

Bhardwaj et al. [9] initiated a novel on efficient deep learning approach for detecting COVID-19 using X-ray imaging modality and further compared them with different versions of the images in the data set and other techniques to add more data. An approach to the detection of COVID-19 disease was provided by Cengiz and Çınar [10] as a system for the concatenation effect of the profound characteristics in the classification problem. Proposed pipeline for COVID-19 detection three data sets are submitted to the proposed pipeline: normal, COVID, and pneumonia by Chilakalapudi et al. [11]. They applied various algorithms to each data set to create a new set of characteristics and supplemented the methodology by applying various classification algorithms to the elements. Pneumonia Detection Using Convolutional Neural Networks by Activist, Militante et al. [12], was developed with 26,684 images in the RSNA data set. They separated the patients suffering from pneumonia and those having the highest resolution of 1024×1024 pixels into two groups.

Heidari et al. [13] used chest X-rays and preprocessing techniques had improved COVID-19 prediction using NBC. First, train, validate, and test with X-rays. DAC 3-class classification is 94.5% accurate. CAD has 98.4% sensitivity and 98.0% specificity for COVID-19. EFiky [14], deep-COVID-19 uses deep learning to detect COVID-19 in radiographic pictures. Six commercial convolution neural networks were evaluated for analysing COVID-19 X-ray pictures. Experiment findings show the effectiveness and reliability of VGG16, VGG19, and MobileNets. Gupta et al. [15] used cnns to detect pneumonia. David Raju et al. [16] classified abnormal and

normal chest X-rays using CNN and SVM classifiers. The best CNN model is chosen via analysis. Using supervised classifier algorithms and pre-formed CNN models to analyse chest radiographs can assist diagnose pneumonia, according to statistics.

3   Methodology

The data set was collected from kaggle.com [17] under open access. The data set comprises CXR scan images and the three classes COVID-19, normal, and pneumonia with different input shapes like 224 * 224 pixels, and the colour format of the CXR scan images is grayscale. It has 6432 total images, 4273 of which are pneumonia chest X-rays, 1583 of which are normal chest X-rays, and 576 of which are cleaned COVID-19 images organized into the test and training directories. They split the data set into two records (training and testing), each with three sub-folders (COVID-19, pneumonia, normal). The test data represents 20% of the total 6432 X-rays in the data set. In data preprocessing, preparing raw image data into formatted data is the first step in the deep learning process, also there are data mining approaches discussed by Jayasingh [18]. Reading the image, resizing the image, eliminating the noise, and converting the images to a denoising format are the steps involved in image data preprocessing. Although the images in the proposed data set have a higher resolution, resizing them to smaller dimensions could simplify the template training process. They set the image sensor input on the network input at 224 * 224 * 3. The parameters used during the implementation are the batch size as 32, 20 epochs, optimizer as Adam, the activation function for the last layer is softmax, the activation function for fully connected layers is a ReLu activation function, the learning rate is 0.0001, and the dropout layer is 0.6. The suggested method addresses the proposed workflow as shown in Fig. 1.

3.1   Deep Learning Image Classifiers

Using chest X-rays, deep learning image classifiers can identify diseases related to COVID-19, normal diseases, and pneumonia. The process of our method is represented in Fig. 2 and includes three different pre-trained CNN models of the recognition system as suggested by Ponnampalli et al. [19], including the InceptionResNetV2, InceptionV3, and NasNetMobile.

3.2   Transfer Learning and ImageNet Weights

To support academics, students and other researchers in image and vision studies, they organized the photos into a hierarchical framework in the ImageNet database. It
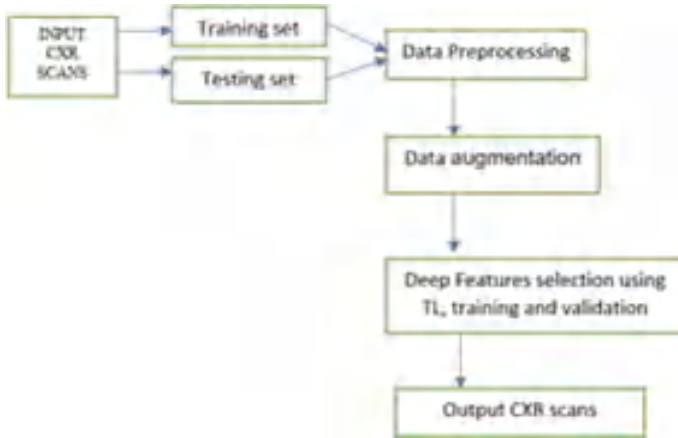
Fig. 1 Proposed workflow
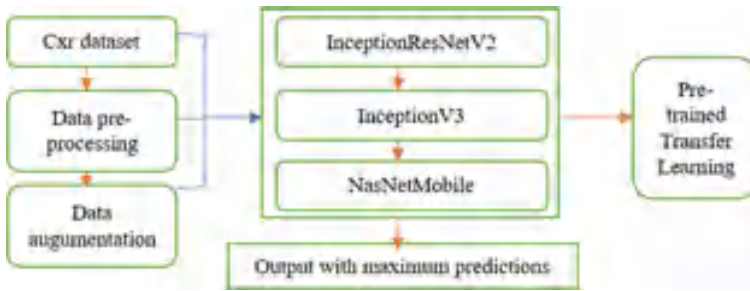
Fig. 2 How modelling is performed

commissioned researchers from around the world to find solutions that would result in the lowest error rates for the ILSVRC.

InceptionResNetV2: In the classification system used by CNN, convolutional layers get input images to process. An input image's receptive field, a small area connected to a convolutional layer, stretches through its whole depth. For an RGB image, the depth is three. Feature mappings organize convolutional layer units. Same-feature map units share the filter bank. The final normal dual module flattens and analyses the output using neurons from the number of classes in a fully linked layer. Softmax transforms completely bound layer output into class probabilities as shown in Fig. 3.

The classification and CNN training are shown. Convolutional filters evaluate the input image by applying a small receptive field and detecting all possible image positions and depths. After a global average pooling layer and one fully connected layer process output feature maps, convolutional layers or pooling filters evaluate them at each surface position. A green functionality map shows the functionality's
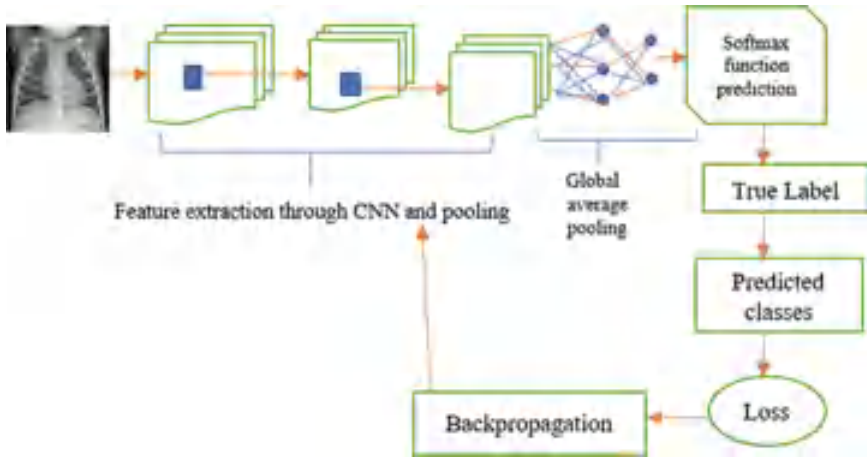
Fig. 3 InceptionResNetV2 architecture

output. ReLu activation of convolutional neurons is followed by a connection to a softmax function, which turns the output into picture category probabilities. During training, the label and classes are used to calculate loss through the object function, cross-entropy. Adam updates the convolutional filter weights and fully connected layer, then retrogrades the network losses.

InceptionV3: On the ImageNet data set, the InceptionV3 image recognition model can achieve the highest level of accuracy. InceptionV3: From the ImageNet data set, the InceptionV3 image recognition model can reach the highest level of accuracy. This model consists of 48 layers that it is already trained on the image network data set over several classes as shown in Fig. 4.
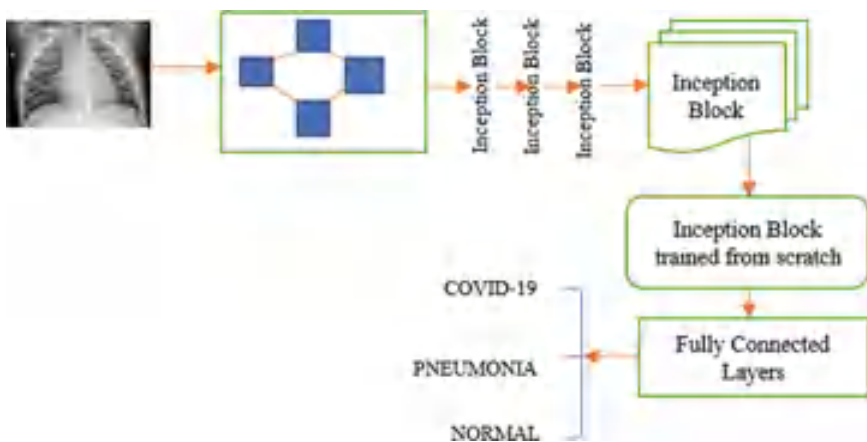
Fig. 4 InceptionV3 architecture
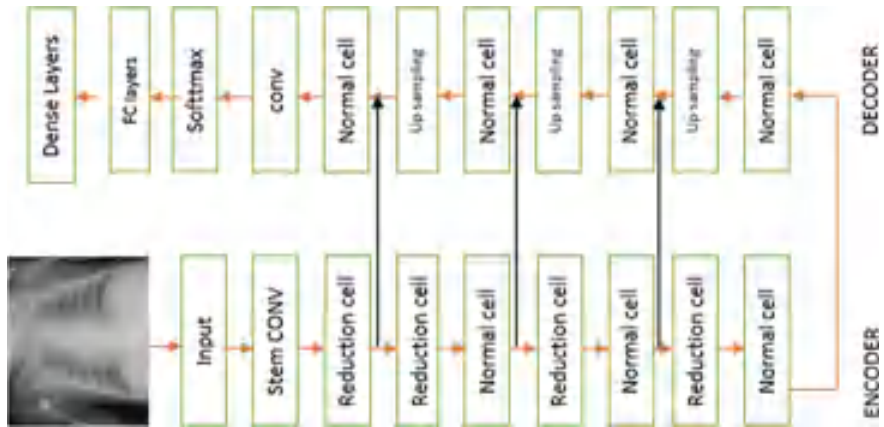
Fig. 5 Layers of NasNetMobile

NasNetMobile: A Neural Search Architecture Network called NasNet is a machine learning model. Core ideas differ from well-known models such as Google Net, and they will soon lead to a great improvement in artificial intelligence. The neural architecture research led to the discovery of NasNet, a network of convolutional neurons. A normal cell refers to a convolutional cell that generates a map of entities of the same size. A normal cell refers to a convolutional cell that generates a map of entities of the same size. The "reduction cell" is a type of convolutional cell in which the height and width of the entity map are halved and then returned as shown in Fig. 5.

4   Results

The deep learning techniques proposed in this work, consisting of the following hardware and software, are required: Processor: Nvidia GPU; 11th Gen Intel (R) Core (TM) i5-1135G7 @ 2.40 GHz, 2.42 GHz; RAM: 8.00 GB; Storage: 476 GB. Windows OS, Python 3.10.5, TensorFlow 2.8.2, Open CV Python 4.6.0, and other required plugins. The proposed models were trained over 20 iterations with the Adam optimizer. The velocity and acceleration rate are both set to 0.0001. The "Loss" output of the process is used to correct tags that may already be present. How many of the images validated in Fig. 6 were identified are shown by the curves for best-performed models are InceptionResNdetV2, InceptionV3, and NasNetMobile.

The effectiveness of a classification model is evaluated using a confusion matrix using the numerical values of the test data for which the actual values are known. It showed the confusion matrix for the detection data set in Fig. 7. The matrix displays improperly and correctly classified images. The x-axis of the array displays the expected or detected labels, while the y-axis displays the actual labels. In this case,
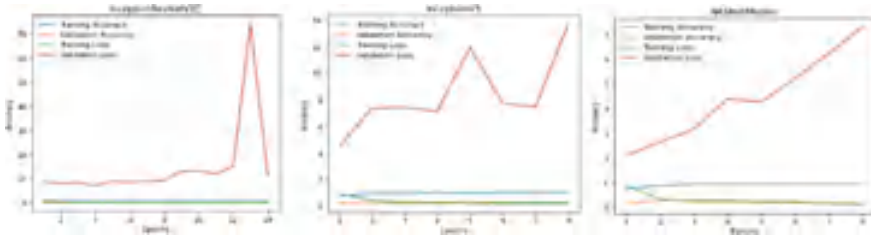
Fig. 6 Training and validation losses as well as training and validation accuracy curves
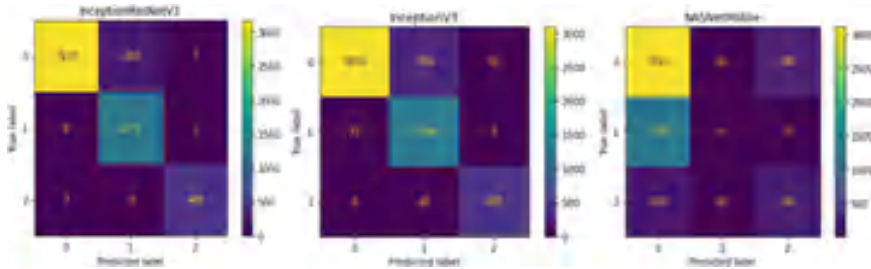
Fig. 7 Outcomes of the confusion matrix

define the "positive results" as indicating the presence of the illness and "negative results" as indicating its absence. In addition, the assessed classifiers for accuracy, precision, recall, and F1-score are shown in Fig. 7 of the top three best models.

The accuracy, precision, recall, F1-score, and area under the curve of all in deep learning techniques presented for COVID-19, normal, and pneumonia were evaluated using CXR scans. Table 1 demonstrates the performance validation of multiple models. Among all classification methods, Xception was the best accurate at 98.26%. An accuracy of ranking of 99.56% was reached by InceptionResNetV2. The MobileNetV3Small registered the highest percentage of reminders (98.26%). For general classification purposes, InceptionResNetV2 outperformed all other methods with an F1-score of 98.46%. Overall, ResNet50V2, ResNet101V2, DenseNet169, DenseNet201, InceptionResNetV2, and InceptionV3 AUC for classification are the highest with 97%.

Table 2 shows the performance validation statistics for multiple models, including accuracy, recall, and F1-scores, and the AUC was the highest of any of the categorization approaches. ResNet50V2 achieved a perfect accuracy score of 99.81%. InceptionResNetV2 has the highest level of accurate recall classifications (92.80%). InceptionResNetV2 achieved a record F1-score of 96.14%, making it the most effective approach to overall categorization. Globally, the best AUC for classification is achieved by Xception, InceptionResNetV2, and InceptionV3 achieved with a score of 100%.

Table 1 Overall accuracy, precision, recall, F1-score, and AUC for the COVID-19 class

COVID-19

S. No	Models	Accuracy	Precision	Recall	F1score	AUC
1	VGG16	75.21	69.48	75.22	72.23	95.00
2	VGG19	68.91	87.81	68.91	77.22	98.00
3	ResNet50V2	94.78	99.32	94.78	97.00	100
4	ResNet101V2	97.82	98.25	97.83	98.04	100
5	ResNet152V2	66.52	98.71	66.52	79.48	97.00
6	DenseNet121	84.13	100	84.13	91.38	97.00
7	DenseNet169	86.95	99.01	86.96	92.59	100
8	DenseNet201	68.04	99.05	68.04	80.67	100
9	Exception	98.26	89.15	98.26	93.49	99.00
10	MobileNet	93.69	96.85	93.70	95.25	99.00
11	MobileNetV2	86.95	3.74	3.04	3.36	44.00
12	InceptionResNetV2	97.39	99.56	97.39	98.46	100
13	InceptionV3	94.56	97.75	94.57	96.13	100
14	NasNetMobile	52.17	43.40	52.17	47.38	86.00

Table 2 AUC values, F1-score, recall, accuracy, and precision for the normal class

Normal

S. No	Models	Accuracy	Precision	Recall	F1score	AUC
1	VGG16	60.64	96.11	60.65	74.37	90
2	VGG19	78.23	97.06	78.23	86.64	95
3	ResNet50V2	75.74	**99.81**	75.75	86.13	99
4	ResNet101V2	92.04	99.06	92.04	95.42	99
5	ResNet152V2	62.20	96.94	62.20	75.78	92
6	DenseNet121	73.81	99.64	73.28	84.81	98
7	DenseNet169	83.93	99.00	83.94	90.85	98
8	DenseNet201	84.96	97.25	84.96	90.69	96
9	Xception	84.66	99.76	84.67	91.60	**100**
10	MobileNet	84.31	99.35	84.32	91.22	98
11	MobileNetV2	91.69	68.43	91.69	78.38	65
12	InceptionResNetV2	**92.80**	99.72	**92.80**	**96.14**	**100**
13	InceptionV3	90.40	99.20	90.40	94.60	**100**
14	NasNetMobile	91.10	68.15	91.11	77.98	67

Table 3 Precision, recall, accuracy, F1-score, and AUC for pneumonia class

Pneumonia						
S. No	Models	Accuracy	Precision	Recall	F1score	AUC
1	VGG16	60.64	96.11	60.65	74.37	90
2	VGG19	78.23	97.06	78.23	86.64	95
3	ResNet50V2	75.74	99.81	75.75	86.13	99
4	ResNet101V2	92.04	99.06	92.04	95.42	99
5	ResNet152V2	62.20	96.94	62.20	75.78	92
6	DenseNet121	73.81	99.64	73.28	84.81	98
7	DenseNet169	83.93	99.00	83.94	90.85	98
8	DenseNet201	84.96	97.25	84.96	90.69	96
9	Xception	84.66	**99.76**	84.67	91.60	**100**
10	MobileNet	84.31	99.35	84.32	91.22	98
11	MobileNetV2	91.69	68.43	91.69	78.38	65
12	InceptionResNetV2	**92.80**	99.72	**92.80**	**96.14**	**100**
13	InceptionV3	90.40	99.20	90.40	94.60	**100**
14	NasNetMobile	91.10	68.15	91.11	77.98	67

InceptionResNetV2 received a record F1 rating of 96.14%, making it the most efficient approach for overall categorization. The several classification strategies of InceptionResNetV2 achieve the greatest accuracy at 92.80%. The overall precision of Xception categorization was 99.76%. InceptionResNetV2 networks were classified with a recall score of 92.80%. InceptionResNetV2 has proven to be the most effective method for broad-based classification with an F1-score of 96.14%. The Xception, InceptionResNetV2, and InceptionV3 100% AUC for classification are the best of any model (Table 3).

5   Conclusion

As part of this research, the proposed deep characteristics are used from transfer learning models to categorize cases of COVID-19, normal cases, and pneumonia cases using digitized client experience images to treat classification techniques. They recommend the use of real-time chest X-rays for the most precise diagnosis of pneumonia, normal chest, and COVID-19. The various transfer learning models and graph-analysis results were compared with one another. More image data is required for better pre-trained CNN models to generalize correctly. Data has more useful applications after being preprocessed. The InceptionResNetV2 performs better at classifying COVID-19 with 99.56% precision and the F1-score with 98.46% than the other models. The InceptionResNetV2 performs better at classifying normal with 92.80% accuracy than the other models, and InceptionResNetV2 performs better at

classifying pneumonia with an accuracy of 92.80% when compared to the VGG, ResNet, DenseNet, Xception, and MobileNet, Inception, and NasNet families with various versions of the COVID-19 image data set. For computer-assisted diagnostics, doctors and researchers can use the model to treat COVID-19, normal, and pneumonia.

References

1. Jayasingh BB, Patra MR, Mahesh DB (2016) Security issues and challenges of big data analytics and visualization. In: 2016 2nd international conference on contemporary computing and informatics (IC3I) 5(6):216–221
2. Umar Ibrahim A, Ozsoz M, Sale S, Al-Turjman F, Habeeb Kolapo S (2021) Convolutional neural network for diagnosis of viral pneumonia and COVID-19 alike & diseases. Expert Syst 13(7)
3. Rahim Zadeh M, Attar A (2020) A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Inf Med Unlocked 19:5
4. Elshennawy NM, Ibrahim DM (2020) Deep-pneumonia framework using deep learning models based on chest X-ray images. Diagnostics 10(9):649
5. Mporas I, Naronglerdrit P (2020) COVID-19 identification from chest X-rays. Int Conf Biomed Innov Appl (BIA) 12(4)
6. Sharma S, Tiwari S (2021) COVID-19 diagnosis using X-ray images and deep learning. Int Conf Artif Intell Smart Syst 23(6)
7. Shah S, Mehta H, Sona wane P (2020) Pneumonia detection using Convolutional neural networks. Third Int Conf Smart Syst Inventive Technol (ICSSIT) 43(12)
8. Sai Krishna D, Rao MM, Dhanush BS, Harshvardhan S, Prudhvi B, Rana P, Mittal U (2021) Pneumonia detection using deep learning algorithms. 2nd Int Conf Intell Eng Manag (ICIEM) 12(4)
9. Bhardwaj P, Kaur A (2021) A novel and efficient deep learning approach for COVID-19 detection using an X-ray imaging modality. Int J Imaging Syst Technol 31(4):1775–1791
10. Cengil E, Çınar A (2021) The effect of deep feature concatenation on the classification problem: an approach to COVID-19 disease detection. Int J Imaging Syst Technol 32(6):26–40
11. Chilakalapudi HP, Venkatesan R, Kamatham Y (2021) Parameter-based performance evaluation of deep learning models for classification of CoViD and pneumonia CT images. In: 2021 IEEE 4th international conference on computing, power and communication technologies (GUCON)
12. Militante SV, Dionisio NV, Sibbaluca BG (2020) Pneumonia and COVID-19 detection using Convolutional neural networks. Third Int Conf Vocat Educ Elect Eng (ICVEE) 32(5)
13. Heidari M, Mirniaharikandehei S, Khuzani AZ, danala G, Qiu Y, Zheng B (2020) Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with pre-processing algorithms. Int J Med Inf 144(34):104–284
14. EFiky A (2021) HDeep COVID-19: deep learning for COVID-19 detection from X-ray images. Int J Innov Technol Exploring Eng 11(7):1–6
15. Gupta P (2021) Pneumonia detection using convolutional neural networks 7(7):77–80
16. David Raju K, Jayasingh BB (2019) Influence of syntactic, semantic and stylistic features for sentiment identification of messages using SVM classifier. Int J Sci Technol Res 8(10):2551–2557. ISSN 2277-8616
17. https://www.kaggle.com/datasets/prashant268/chest xray-covid19-pneumonia

18. Jayasingh BB (2016) A data mining approach to inquiry-based inductive learning practice in engineering education. In: 2016 IEEE 6th international conference on advanced computing, pp 845–850
19. Ponnampalli S, Venkata Suryanarayana Birudukota N, Kamal A (2022) COVID-19: vaccines and therapeutics. Bioorg Med Chem Lett 75(5):128987

Attention-Based Approach for English to Hindi Translation

H. S. Gururaja, M. Seetha, Niranjan Hegde, and Ankit Das

Abstract Artificial Intelligence (AI) is cleverly evolving with time and in large volumes of computational and processing strength and excessive call for the final quarter, greater than a billion dollars. Through language translation, people from different parts of the world can communicate, work together, and develop relationships. Machine Translation is undergoing a major transformation thanks to the use of neural networks in machine learning (MT). Looking at the identical aspect, a Machine Translation for English to Hindi has been proposed using the likes of Neural Machine Translation techniques along with attention mechanisms. Neural Machine Translation (NMT) is a modern approach which gives extremely good enhancements in evaluation of traditional system translation techniques. Neural Machine Translation has been capable of reaping massive development over ancient techniques: Rule primarily based model and Statistical Machine Translation. Aiming on the trouble of managing a lengthy distance dependency, attention mechanism is incorporated into the interpretation model, as a result the preprocessing module, encoder-decoder framework, and attention module of the system also are adopted.

Keywords NMT · Attention mechanism · RNN · LSTM · GRU · Transformer

1   Introduction

Machine Translation (MT) is one of the most essential domains within the place of Natural Language Processing. It converts scripts from a single language to scripts written in another language. Basically, MT goes for performing a mechanical substitution of phrases in a single language for phrases in every other language, however that doesn't necessarily produce an excellent translation due to the fact that the

H. S. Gururaja (✉) · N. Hegde · A. Das
Department of ISE, B.M.S. College of Engineering, Bengaluru, India
e-mail: gururajhs.ise@bmsce.ac.in

M. Seetha
Department of CSE, G. Narayanamma Institute of Technology and Science, Hyderabad, India

popularity of the complete terms and their closest opposite words in the sentences in the target language is needed. Every phrase in a language now no longer has equal phrases inside the desired language, and a few phrases can have multiple meanings. Solving this trouble with the corpus statistical and neural techniques is an increasingly-growing area which ends in decent translations, handling various versions in linguistic typology, translation of idioms, and isolation of anomalies.

NMT is a utility of deep learning wherein huge datasets of translated sentences educate the model so that it can do translation from one language to another language. It is an intensive departure from older translation methods. NMT employs the method of continuous representations as opposed to discrete symbolic representations in SMT. Alternatively, in NMT, the entire translation process is versioned by a single neural network, thereby liberating the requirement for function engineering. Whereas the SMT models make use of tuned components for the purpose of learning whereas the architecture of NMT models is end-to-end. Besides its simplicity, NMT has accomplished contemporary overall performance on several language pairs. It has turned out to be an important era in today's world as many businesses are making use of the MT systems. The strength of NMT is its ability to directly and consistently analyze the mapping of input text content to relevant output text content.

The encoder-decoder structure is a method of using the recurrent neural networks for a series-to-series prediction problem. In order to clear up this hassle of the encoder-decoder version encoding the input series to one constant size of vector from which to decode every output time step, attention mechanism is proposed as a solution. This problem is thought to be more elaborate whilst interpreting lengthy sequences. Language translation fosters global collaboration, fact-sharing, and relationship-building by bringing people from different countries together. Translation is essential for spreading new facts, knowledge and ideas around the world. Clearly, achieving strong communication across cultures is critical. Translation is something which could extrude history. Communication between languages and cultures must be clear and effective in our multilingual and multicultural society. Companies, schools, scientific missions all enjoy the incorporation of translation.

The most frequently spoken language in the world is English, which has enormous global significance. As a result, translating information from English to native languages and vice versa becomes necessary. With the improvements in the deep learning domain, Machine Translation is becoming feasible without human involvement. Hindi to English translation has been attempted to obtain on this assignment the use of Neural Machine Translation [3, 5].

2 Related Work

There have been various attempts to build a robust translational model for text translation between many different languages. English being the highest spoken language and Hindi being the second highest spoken language is not an exception. In 2019, two models NMT-1 and NMT-2 was proposed and was published with paper 'Neural

Machine Translation: English to Hindi (2019)' [1, 2] by Sahinur Rahman Laskar, Partha Pakray, Abinash Dutta, and Sivaji Bandyopadhyay. This paper compared both the models which were based on LSTM and Transformers Architecture. Comparison between two exceptional NMT structures, specifically NMT-1 which is based at the Long Short-Term Memory (LSTM) version [6] and NMT-2 which relies upon the transformer version are discussed with the context of English to Hindi translation. Using the Bilingual Evaluation Understudy (Bleu) metric, system effects are assessed. The common Bleu ratings of NMT1 structures had been 35.89 and NMT-2 structures changed into 34.42. The effects display higher overall performance than current NMT structures.

In 'English-Hindi Neural Machine Translation-LSTM Seq2Seq and ConvS2S (2020)' [7] the authors have analyzed and done comparison of NMT fashions on various parameters for the English-Hindi language pair: a sequence-to-sequence learning structure where each encoder and decoder use (1) LSTMs and (2) Convolutional Neural Networks (CNNs) and embedded in each method attention mechanism to achieve. The evaluation offers insights which show the models and parameters best suited for this task. There was another paper Neural Machine Translation with GRU-Gated Attention Model (2019) [9] where the authors recommend a singular GRU-gated attention model (GAtt) for NMT. The secret is that people will be able to extend the variance of context vectors with the aid of using refining source representations in step with the partial translation generated while using the decoder. The authors stated that the LSTM layer desires to condense all essential statistics in the context vector which may not be needed. Bahdanau presented an extension of the use of an attention mechanism that looks for positions in which suitable information are centered in order to improve the accuracy of the LSTM Seq2Seq structure. The decoder anticipates the target word using context vectors that resemble these centered positions. It preserves a fixed-length context vector resembling the preceding sentence instead of immediately collecting lengthy input sentences into a single vector and choosing a subset of them for attention. The technique stores the attention weights associated with the context vector of the preceding phrase in an attention vector (x).

The translation network with transformers described in 'Re-Transformer: A Self-Attention-Based Model for Machine Translation (2021)' [8] depicts the effects of self-attention on NMT tasks. A multi-head self-attention visualization heat map was also proposed by the authors of this research, which largely demonstrates usage of attention weights of the source text after processing of the self-attention mechanism. Their analysis of the multi-head self-attention, in which the interpretation version will pay attention to the sequence and the manner in which it allows to convert the supply sentence to the target sentence was largely improved, thanks to the heat map. The above works have successfully shown that attention mechanisms increase the efficiency of translation.

Phrase-Based Machine Translation of Digaru-English (2020), by Kri and Sambyo [4] used a phrase-based translation between Digaru, an underdeveloped language spoken by the Tawra Mishmi Tribe of Arunachal Pradesh, and English. The statistical version offers an advantage over the Rule-Based Translation system because it no longer depends on language functions, such as grammar and language-specific translation styles, and it also permits us to design several language pairs with little modification. The Corpus Instruction of the Digaru-English Parallel Corpus includes Tokenization, which could be very useful for Statistical Machine Translation (SMT), true casing, which restores case facts to poorly or uncased text and avoids inappropriate distinction between lowercase and uppercase phrases, and cleaned phrases, which eliminates lengthy and empty sentences so as to improve readability. By creating an alignment model between the input and output languages, translation models are built. The output of the language model and the output of the translation model in addition to the supply sentence is the starting point for the decoder.

In the same direction, the model proposed in this paper implements the attention mechanism with a Gated Recurrent unit for English to Hindi language translation using Seq2Seq architecture.

3 Proposed Method

The steps involved in the system include data collection, data preprocessing, NMT Attention Net model training with attention, and model testing same have been discussed in the below subsections. Figure 1 shows the proposed method's system architecture.

3.1 Data Collection

Dataset Collection is the most essential step in the lifecycle of a machine learning model. The quantity and quality of the accumulated statistics makes a decision on the overall performance of the output. More the quality of the data, the more accurate the results will be. For this purpose, a Hindi-English dataset was collected from Keras official blog [2]. A total of 2915 English sentences and their matching Hindi translations make up the dataset.

3.2 Data Preprocessing

Data preprocessing in machine learning is a critical step which improves the quality of data with a view to sell the extraction of significant insights from the statistics. Quality data holds more weightage than the quantity of data. It is the approach
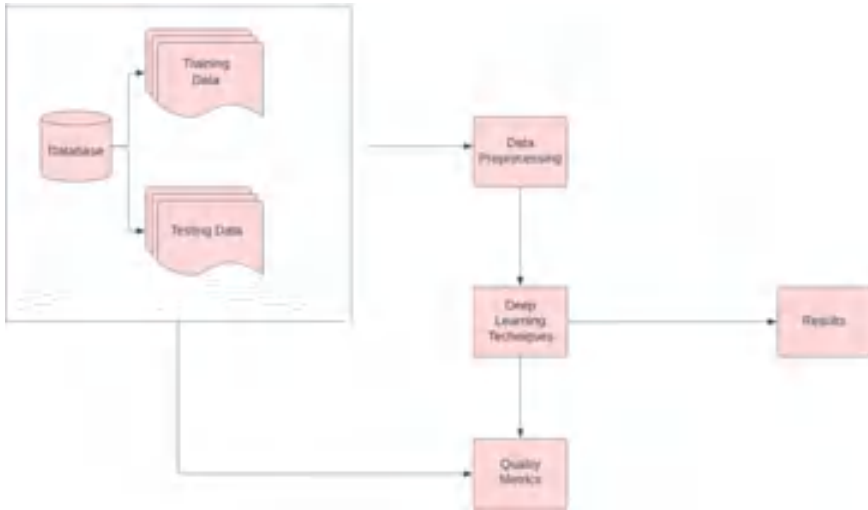
Fig. 1 System architecture

of preparing (cleansing and organizing) the raw statistics with a view to make it appropriate for constructing and training ML models. In simple words, it's a data mining approach that transforms uncooked facts into an understandable and readable format.

The preprocessing techniques employed in the research paper are shown in Fig. 2. The raw text of both English and Hindi is cleaned to remove stop words and are tokenized. Then the embeddings of the tokenized words are generated which are provided as input to the encoder-decoder network.
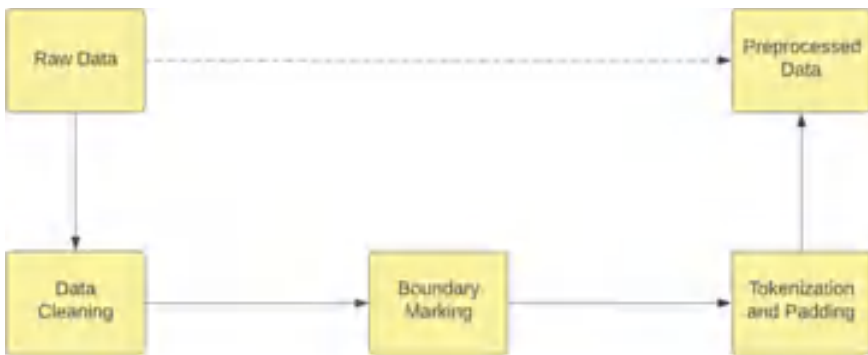
Fig. 2 Preprocessing techniques

3.3 Algorithm Selection

The human mind is integrated into a system through deep learning, a subset of machine learning. It is stimulated by how a human mind functions, therefore it is a fixed of neural network algorithms that try to mimic how a human mind functions and learn from experiences. Neural networks are a collection of algorithms that are made to recognize styles. They are loosely based on the human brain.

The output of the previous step is used as the input for the current step in a recurrent neural network (RNN), which is a type of neural network. While all inputs and outputs in a typical neural network are independent of one another, in instances when the next phrase of a sentence must be anticipated, the preceding phrases and hence the preceding words may be necessary.

Convolutional Neural Networks (ConvNet/CNN) are a collection of deep learning algorithms that can take an input image and provide meaning (learnable weights and biases) to several features of the image. They have the best chance of telling apart two things that are similar. Compared to individual techniques, ConvNet requires a lot less preprocessing. With enough training, ConvNets can investigate their own filters and attributes, in contrast to the initial approach where filters were created manually. They are also known as deep convolutional networks or shift-invariant networks. Feature maps are translation-equivariant responses produced by space invariant artificial neural networks, which are primarily based on the shared-weight structure of the convolution kernels or filters that slide along entry functions.

3.4 Encoder-Decoder Model for RNNs

The preferred NMT technique is the encoder-decoder model for recurrent neural networks (RNN). It competes and, in some circumstances, outperforms classical statistical translation techniques. Encoder reads and encodes input textual content right into a fixed-size or a constant vector. A decoder then outputs a translation within the target language from the encoded vector.

The Sequence-to-Sequence (Seq2Seq) version, as shown in Fig. 3, includes subnetworks, which comprise of both the encoder as well as the decoder. The encoder, on the left hand, receives sequences from the incoming/source language as inputs and produces a condensed illustration of it, summarizing all of its information. The output then generated becomes an input for the decoder. The decoder can have other inputs as well.

The decoder generates a detailed element of its output sentence, primarily based on the source sentences entered at every step. The decoder, very similarly, updates its own state for the following time steps. The size of the source series can be different from that of the output series. Getting the encoder to offer the maximum whole and significant translation of its sentences in single output detail to the decoder is the most important factor of this model. The longer the entered text, the extra tough it
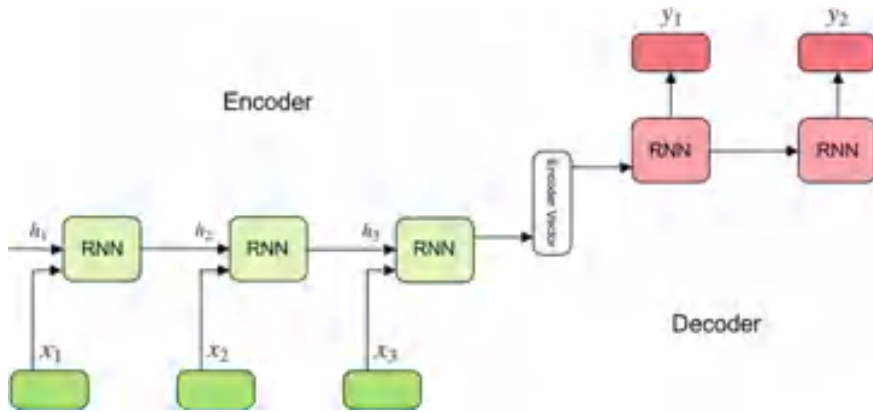
Fig. 3 Simple encoder-decoder architecture

will be to compress it right into a standalone constant size vector. It is cumbersome for the encoder version to memorize prolonged sentences and convert it properly into a fixed-period vector. Moreover, the decoder performs more effectively from the final encoder's hidden kingdom. Hence, it's very much difficult for the decoder to summarize longer entered sequences at once.

3.5 NMT Attention Net Model

The suggested model, NMT Attention Net, is a kind of recurrent neural network that uses Gated Recurrent Unit (GRU) as a basis model for encoder and decoder networks. In addition, the encoder and decoder model also have an attention mechanism. In their study, Bahdanau et al. performed RNNs for the encoder and decoder, as well as attention architectures and sequencing architectures.

According to Bahdanau et al., the attention mechanism is composed of step-by-step computations of alignment scores, weights, and context vectors.

  i. Alignment scores: The encoded hidden states, h_i, are taken in the alignment model and the preceding decoder output to evaluate a score that gives an idea how properly the factors of the source sequence synchronize with the current output.
 ii. Weights: The weights are computed by making use of a SoftMax operation to the formerly computed alignment scores.
iii. Context Vector: At every time step, the decoder receives a unique context vector. The hidden states of all encoders are weighted together to calculate it.

In this mechanism, unlike Seq2Seq without attention where the last hidden state output is used, here all the hidden state outputs of forward and backward propagation of both encoder and decoder are used. Figure 4 depicts the attention architecture used in the proposed model.
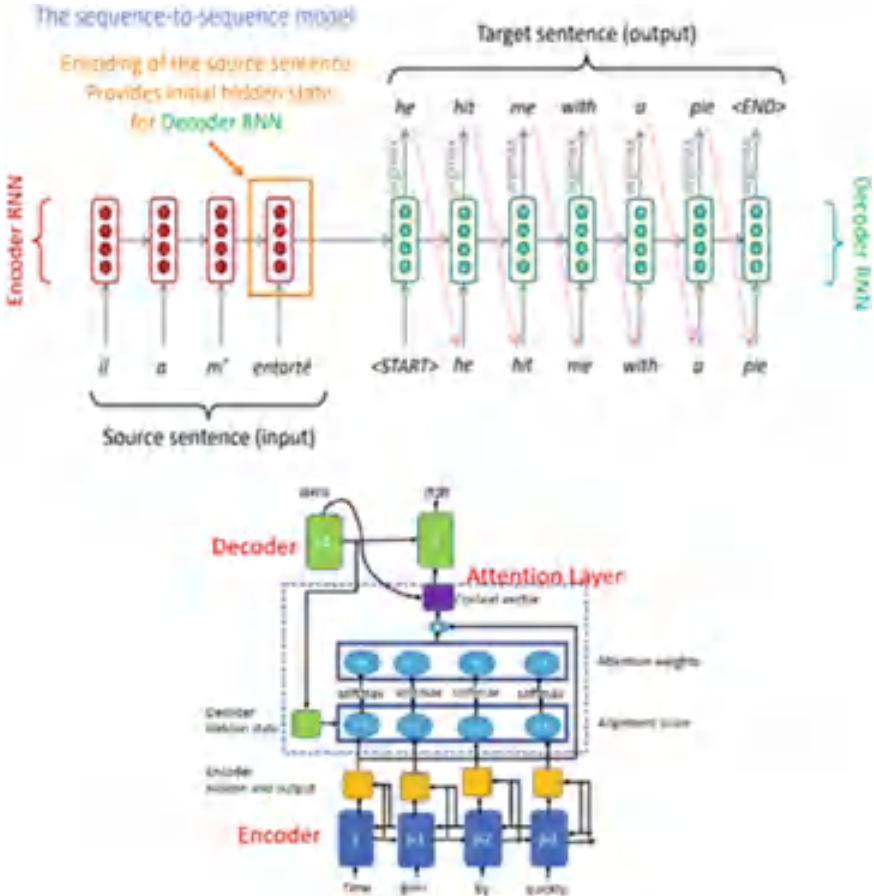

Fig. 4 Attention mechanism with Seq2Seq model

The most crucial components of the input sequence might be used by the decoder in a flexible way thanks to a weighted combination of all encoded input vectors. Highest weights are allotted to the most relevant vectors. It predicts the subsequent word via means of focusing on some applicable elements of the series, rather than searching on the whole series. It acts as an interference among the encoder and decoder which extracts beneficial records from the encoder and transmits it back to the decoder.

This mechanism was deployed to deal with the issue that occurs with the usage of a constant-length encoding vector, in which the decoder would have limited access to the source sentence. The primitive methods would become trickier for lengthy and/or complicated sequences, in which the dimensionality could be pressurized to be similar to shorter or less complicated sequences.

4   Results and Discussion

The NMT Attention Net model has been tested with 250 instances of data which included both the test set data and the data collected from external sources in the ratio 9:1. To compute the Bleu score [10], first we take the target sentence and find out the precision scores for n-grams (generally taken upto 4-g). Then we combine the precision scores to find the geometric average precision score using:

$$\text{Geometric Average Precision}(N) = \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

$$= \prod_{n=1}^{N} p_n^{w_n}$$

$$= (p_1)^{\frac{1}{4}} \cdot (p_2)^{\frac{1}{4}} \cdot (p_3)^{\frac{1}{4}} \cdot (p_1)^{\frac{1}{4}}$$

In the next step, we calculate the Brevity Penalty using:

$$\text{Brevity Penalty} = \begin{cases} 1, & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c <= r \end{cases}$$

In the last step, we multiply the geometric average precision score and Brevity Penalty to compute the Bleu score.

$$\text{Bleu}(N) = \text{Brevity Penalty} \cdot \text{Geometric Average Precision Scores}(N)$$

Table 1 Bleu score comparison of proposed model with NMT-1

Model	Bleu score
NMT-1 (2019) [2]	0.3589
NMT attention net (Eng-Hin) [Proposed model]	0.302
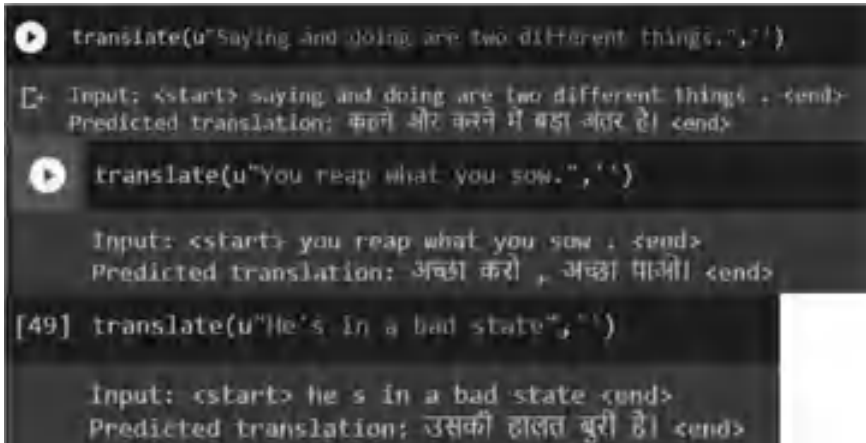
Fig. 5 Sample outputs of NMT attention net proposed model

It is observed that the proposed model performed better for long sentence inputs than very short sentences. For 250 instances of data, Bleu score was found to be 0.302 which was slightly better than the earlier models for text translation of English to Hindi considering the training set sample. Table 1 shows the performance of the proposed model and the previous models.

The model results are shown in Fig. 5 and their corresponding Google Translation outputs are shown in Fig. 6. It is observed that the results from the model are almost identical to output of Google Translator in terms of context of the sentence.

Fig. 6 Outputs of Google Translator for same inputs

5 Conclusion

The attention mechanism has revolutionized the manner in which we create NLP systems and is presently a general tool in most of the current NLP fashions. In effect, this lets the model remember all of the phrases within the input and give attention to particular phrases while formulating an answer.

An NMT Attention Net model has been proposed which is an experiment in which Attention mechanism has been applied to standard Seq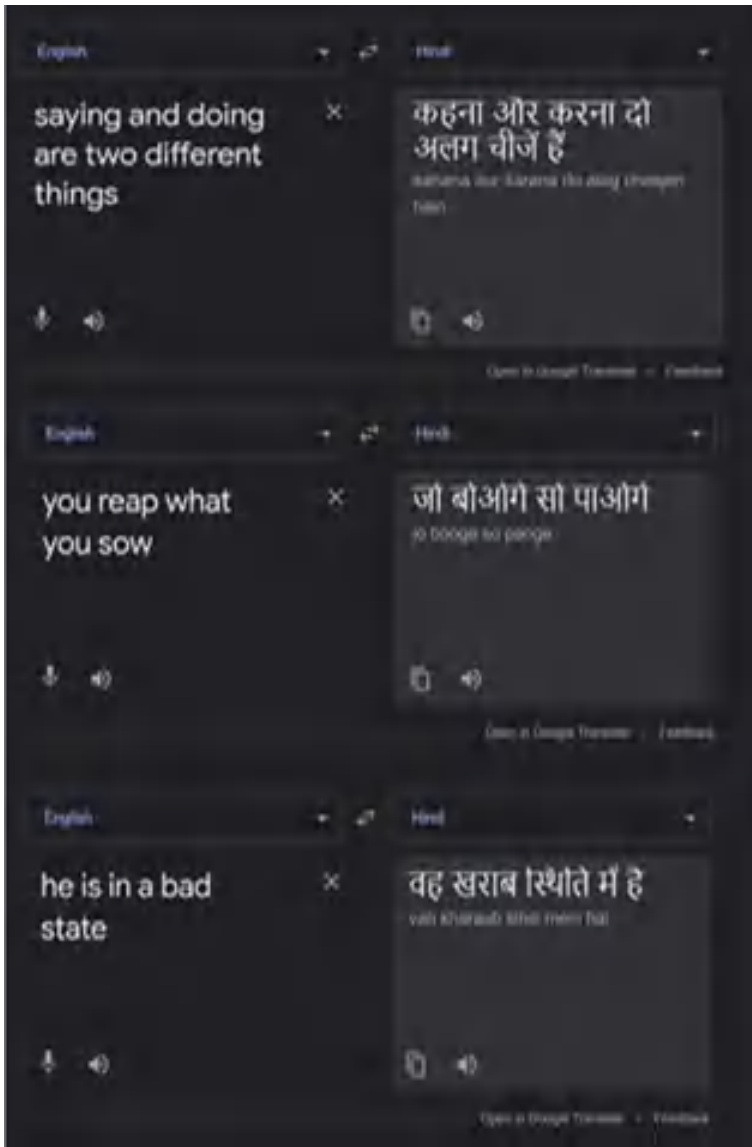2Seq models for translating English sentences into Hindi. Considering the low training set sample which was around 2700, the model did fairly well. The results also convey the fact that neural networks require a large number of quality training data. The results open the door for more experimentation on the model. The biggest advantage of this model is that it performs really well for even complex sentences whose translations require a diligent reordering.

Further work can be done to improve Bleu scores. The model can be implemented using data of various languages. Hence a multilingual product can be built with the aforementioned model as the basis.

References

1. Singh SP, Darbari H, Kumar A, Jain S, Lohan A (2019) Overview of neural machine translation for English-Hindi. In: International conference on issues and challenges in intelligent computing techniques (ICICT), vol 1. IEEE, pp 1–4
2. Basmatkar P, Holani H, Kaushal S (2019) Survey on neural machine translation for multilingual translation system. In: 3rd International conference on computing methodologies and communication (ICCMC), IEEE, pp 443–448
3. Wang H, Wu H, He Z, Huang L, Church KW (2021) Progress in machine translation. Engineering
4. Rushanti K, Sambyo K (2020) Phrase-based machine translation of Digaru-English. In: Electronic systems and intelligent computing, Springer, Singapore, pp 983–992
5. Tan Z, Wang S, Yang Z, Chen G, Huang X, Sun M, Liu Y (2020) Neural machine translation: a review of methods, resources, and tools. AI Open 1:5–21
6. Cui Y, Wang S, Li J (2015) LSTM neural reordering feature for statistical machine translation. arXiv preprint arXiv:1512.00177
7. Gaurav T, Sharma A, Sahotra A, Kapoor R (2020) English-Hindi neural machine translation-LSTM Seq2Seq and ConvS2S. In: 2020 International conference on communication and signal processing (ICCSP), IEEE, pp 871–875
8. Liu H-I, Chen W-L (2021) Re-transformer: a self-attention based model for machine translation. Proc Comput Sci 189:3–10
9. Zhang B, Xiong D, Xie J, Jinsong S (2020) Neural machine translation with GRU-gated attention model. IEEE Trans Neural Netw Learn Syst 31(11):4688–4698
10. Malik P, Baghel AS (2016) An improvement in BLEU metric for English-Hindi machine translation evaluation. In: International conference on computing, communication and automation (ICCCA), pp 331–336

A Dependable and Secure Communication Infrastructure for Sensor Networks

A. Vijaya Krishna, I. Ravi Prakash Reddy, A. Anny Leema, and M. Dinesh

Abstract The military, health care, home automation, remote monitoring, and many industries make use of Internet of Things (IoT), a newer era of WSNs devices. Device security has risen to the top of the priority list as a result of their widespread use and recent significant distributed denial of service (DDoS) attacks involving huge of these devices. Because these wireless devices have limited power resources, the security solution must also be energy efficient. Because both users and attackers have remote management or access capabilities can attack on IoT devices. We created a secured communication system for these LoWPAN devices in this study. We created and implemented a one-of-a-kind security solution for detecting and preventing RPL attacks in IoT. We examined how long the batteries in IoT devices lasted and how much energy they consumed before and after we implemented our suggested repair. Using received signal strength indicator (RSSI) tunneling, the proposed security technique detects and corrects routing protocol layer (RPL) issues.

Keywords IoT · WSNs · Attacks

A. Vijaya Krishna
Department of Computer Science and Technology, GNITS, Hyderabad, India
e-mail: Vijayakrishnaakula2022@gnits.ac.in

I. Ravi Prakash Reddy
Department of Information Technology, GNITS, Hyderabad, India
e-mail: irpreddy@gnits.ac.in

A. Anny Leema
SCOPE, VIT, Vellore, India
e-mail: annyleema.a@vit.ac.in

M. Dinesh (✉)
College of Computing and Informatics, Saudi Electronics University, Riyadh, Kingdom of Saudi Arabia
e-mail: dineshmsaudiuniversity@gmail.com

1   Introduction

The importance of IoT has grown recently. To capture sound, temperature, and pressure events in the Internet of Things, numerous remote sensor devices have been placed. The dispersed sensor nodes work on an as-needed basis and send whatever data they find to the sink node [1]. Environmental monitoring is one application of the Internet of Things [2–7]. Examples include tools for industrial control, cyber-physical systems, and public safety equipment. In a range of applications, including environmental monitoring, health monitoring, and smart home models, IoT use has increased substantially [8–11]. Energy optimization and security is an exigent endeavor.

The following is a summary of the key contributions of the research:

- By considering sensor node threshold values, a heterogeneous communication architecture that uses less energy can be created.
- Choosing the appropriate parent member node when using the RPL protocol to cut out pointless branches and save time.
- Improving network performance by thwarting DoS attacks with pruning.

The remaining sections of the paper are laid out as follows. In the second part, we cover related work on topics including secure routing and energy efficiency. Details of the energy model are presented in Sect. 3. The fourth section describes the suggested trimming model for denial of service attacks. The fifth section concerns experimental analysis. The sixth section closes the research paper.

2   Literature Review

Over the past ten years, many academics have worked to develop a real-time routing system for energy optimization and security risks in data transfer. The cluster leader is chosen at random by the LEACH approach, developed by the authors of [12]. Every node has equal opportunity to become a cluster leader once every 1/p epoch. After multiple loops, a random numeral between 0 and 1 is provoke [13]. A node with an energy level below the threshold is considered a cluster head node [14]. Next cluster leader is chosen among non-leader nodes. After a cluster has been formed, the ADV message is sent out to all of the nodes in the cluster by the leader node. The nodes use their allocated TDMA slot to talk to the cluster master. Fundamentally, large-scale sensing environments are incompatible with the LEACH technique. Advanced nodes have more energy than standard nodes in a heterogeneous network [15]. The stable election process (SEP), whose results are based on weighted probability, is the technique for choosing cluster heads. Its performance and stability are better than the LEACH method. The hybrid model zones stable election algorithm was introduced by the authors in [16]. (Z-SEP). All three zones are covered by Z-SEP at random. Zones 1 and 2 are evenly and asymmetrically spread with highly advanced nodes. A report on these gadgets' security was just issued. Seventy percent of these gadgets are

vulnerable, according to one research [6]. Because IoT devices have limited capacity, distributed denial of service (DDOS) attacks have the potential to bring the entire network down. DDOS attacks on IoT devices have also been reported in a few cases. In the hypothetical scenario, Dyn, a DNS service, received queries from millions of IoT devices running dangerous operating systems, rendering Netflix, CNN, and Twitter unavailable [7]. Classification of Routing Algorithms.

2.1 Types of Attacks

(i) Wormhole Attack

The attacker constructs a phonepey communication link between nodes during a wormhole assault. Compared to other pathways or routes in the network, this road or route seems to be shorter. All pursing nodes are compelled to use the contaminated channel as a result of the interruption in the node routing mechanism. By employing this technique, the attacker is able to create a tunnel utilizing one or more malicious node. Any node can gather and spread the sent packet during this attack.

(ii) Hello Flooding

Flooding is commonly referred to as a denial of service (DoS) assault. A continuous hello flooding attack creates a lot of traffic by delivering packets or messages over and over again, causing network nodes to send destination-oriented directed acyclic information object (DIO) messages and forcing them to reset their trickle timers. Many RPL strikes might be used to carry out this onslaught.

(iii) Rank Attack

The relative positions of network nodes in a directed acyclic graph with the end destination circled are used to determine their order in the network's overall ranking. The attacker invites uncompromised nodes to join the destination-oriented directed acyclic network by broadcasting the lower rank to those nodes in a rank assault. This technique can be used to carry out a variety of distributed denial of service attacks, including wormhole and sinkhole attacks.

3 Energy Model for IoT

The proposed mechanism is broken down into three parts. They are monitoring RSSI levels, gathering surroundings information, and validating RSSI data for energy optimization. The goal of this project was to create a routing system that saves energy while increasing node longevity. To minimize energy loss, we investigated network heterogeneity at three distinct levels based on baseline node energy. Normal nodes have the least amount of energy, followed by intermediate nodes and then advanced nodes (Fig. 1).
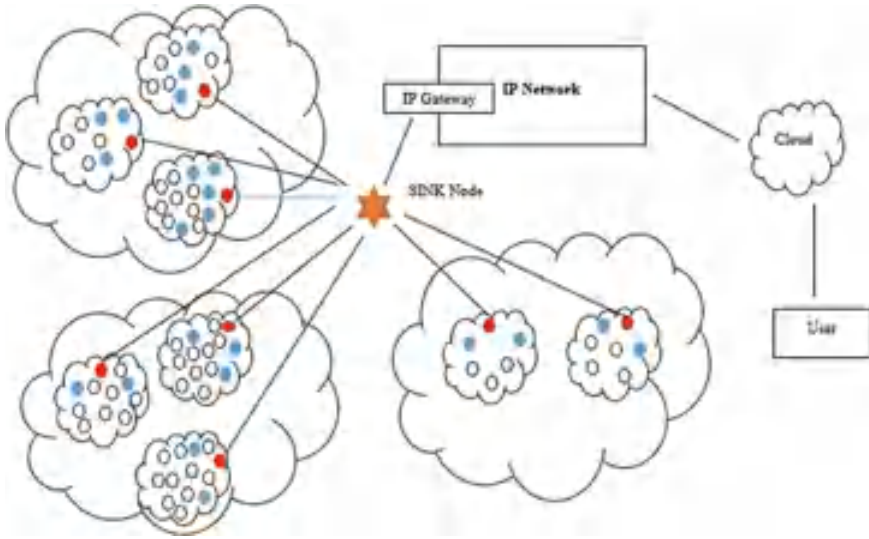
Fig. 1 Architecture of the proposed model

Let's pretend the initial energy of the basic nodes is 0, the advanced nodes are 0(1 + X), and the intermediate nodes are 0(1 + Y). Normal, intermediate, and advanced node energies are all reflected in Eqs. 1, 2, and 3.

$$\vartheta_{Nn} = k\vartheta_0(1 - x - y) \tag{1}$$

$$\vartheta_{In} = ky\vartheta_0(1 + Y) \tag{2}$$

$$\vartheta_{An} = kx\vartheta_0(1 + X) \tag{3}$$

Advanced network nodes can be recognized by their energy level, denoted by the letter X. The energy of the network's intermediate nodes, represented by the letter y, is calculated using the formula $Y = X/2$, where k is the size of the node in question. Total network energy consumption is shown as a percentage.

$$\begin{aligned} \vartheta_{tot} &= k\vartheta_0(1 - x - y) + ky\vartheta_0(1 + Y) + kx\vartheta_0(1 + X) \\ &= k\vartheta_0(1 + xX + yY) \end{aligned} \tag{4}$$

For the purpose of choosing the cluster leadership, we use the SEP and LEACH. Based on each type of node's potential to become the cluster head, the threshold value is established. The following nodes should be picked, assuming $S1$, $S2$, and $S3$, are the sets of each class of node that have not yet been selected as the cluster head:

For N_n:

$$P_{Nn} = \frac{P}{1 + xX + yY} \tag{5}$$

$$T_{K_{Nn}} = \begin{cases} \frac{P_{Nn}}{1 - P_{Nn}^{(u \bmod 1/P_{Nn})}}; & \text{if } k_{Nn} \in S_1 \\ 0; & \text{otherwise} \end{cases} \tag{6}$$

For I_n:

$$P_{In} = \frac{P(1 + Y)}{1 + xX + yY} \tag{7}$$

$$T_{K_{In}} = \begin{cases} \frac{P_{In}}{1 - P_{In}^{(u \bmod 1/P_{In})}}; & \text{if } k_{In} \in S_2 \\ 0; & \text{otherwise} \end{cases} \tag{8}$$

For A_n:

$$P_{An} = \frac{P(1 + X)}{1 + xX + yY} \tag{9}$$

$$T_{K_{An}} = \begin{cases} \frac{P_{An}}{1 - P_{An}^{(u \bmod 1/P_{An})}}; & \text{if } k_{An} \in S_3 \\ 0; & \text{otherwise} \end{cases} \tag{10}$$

The average probability of picking the cluster heads can be found by combining the results of Eqs. (5, 7, and 9), based on this we can calculate create a CH and clusters for energy optimization.

4 Trimming Model for Security Attacks

When denial of service attacks like Wormhole and Hello flooding assault are initiated, it causes ripple effects throughout the network. Due to the attacks capacity to disrupt networks and drain device resources, these modifications might have an effect on network performance indicators. If such an attack occurs, against a network of IoT devices, the devices would stop working, resulting in disastrous consequences. As a result, the recommended solution was designed with IoT device architecture in mind (Fig. 2).

In our trimming model, we uses RPL protocol for avoiding attacks. This mechanism helps to avoid the network congestion, delay, and attacks while routing the packets and leads to improve the QoS in routing. The intruder is not allowed into the network easily without the authentication from the CH; all the nodes are joined only the network with proper authentication from the CH. After joining into the network,
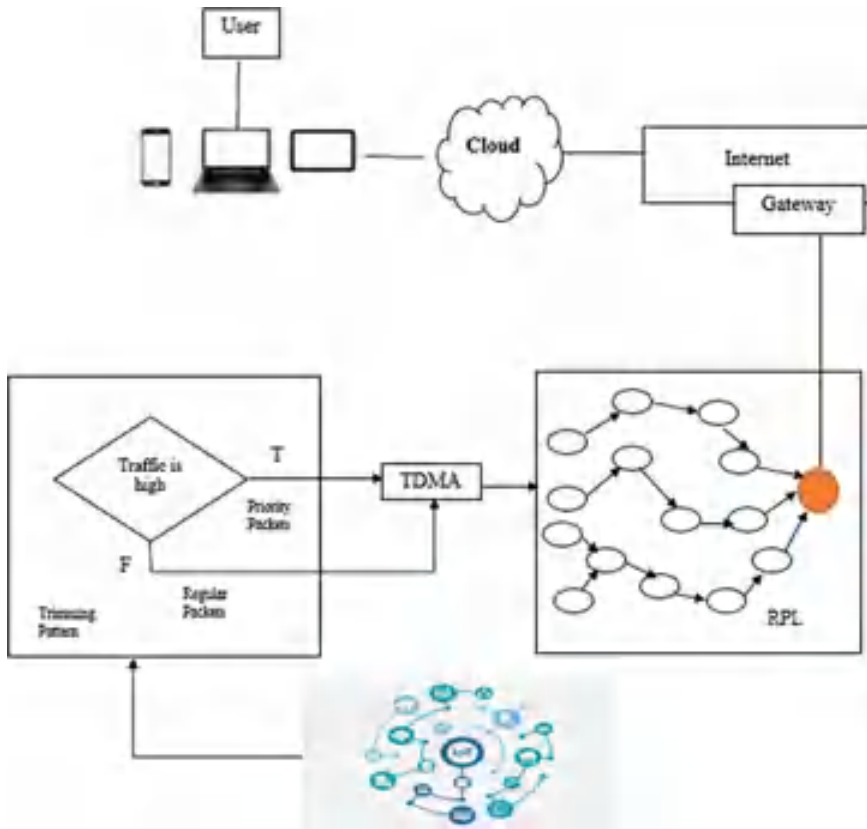
Fig. 2 Architecture of the trimming mechanism for denial of service attacks

the CH can issue the TDMA for their cluster nodes to send their packet. In case if any node accidentally joins in the network without any authentication and falsely propagates the wrong routes and hello messages continuously that can be detected easily with the help of our proposed model. By using the trimming model, we can monitor and shape the traffic while routing. When there is network congestion or some emergency situations, the trimming mechanism can assign time slots to all nodes in their network. Based on this all nodes can give willingness to forward their packets in the network. Then the trimming model observes any suspicious node attacks, it can alert all the nodes and take alternative actions.

5 Experimental Analysis

The evaluation of the proposed Energy Trimming model (ETM) method was performed using the QualNet simulator. The performance of the ETM method is evaluated with parameters like network lifetime and throughput and also it is compared with the SEP and Z-SEP.

5.1 Network Lifetime

See Fig. 3.

5.2 Throughput

See Fig. 4.

Fig. 3 Network lifetime

Fig. 4 Network throughput

6   Conclusion

In this research, we developed an energy-aware paradigm for effective Internet of Things communication based on a threshold factor. In homogeneous networks, conventional algorithms such as are commonly utilized. Regarding QoS parameters, these protocols are inefficient. The IoT network in the real world consists of numerous sensor nodes that are both energy-constrained and vulne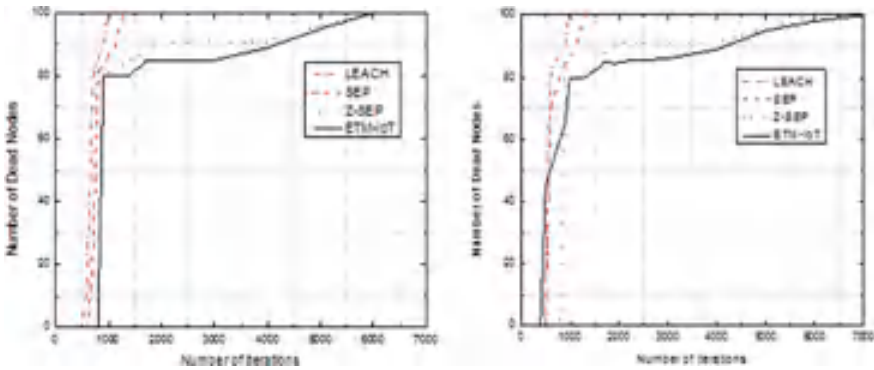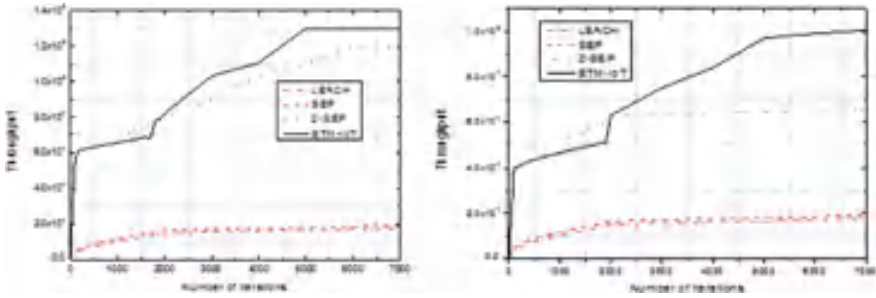rable to security assaults. Before entrusting them with vital jobs, we must ensure their security. In an effort to lessen the impact of energy consumption and network security breaches, we deploy a variety of operating circumstances.

References

1. Kumar M, Verma S, Lata K (2015) Secure data aggregation in wireless sensor networks using homomorphic encryption. Intl J Elec 102(4):690–702
2. Shahzad MK, Ho Cho T (2017) An energy-aware routing and filtering node (ERF) selection in CCEF to extend network lifetime in WSN. IETE J Res 1–13. https://doi.org/10.1080/03772063.2016.1241721
3. Wang T-Y, Cheng Q (2008) Collaborative event region and boundary-region detections in wireless sensor networks. IEEE Trans Sig Process 56(6):2547–2561
4. Usman MR, Shin SY (2016) Channel allocation schemes for permanent user channel assignment in wireless cellular networks. IETE J Res 62(2):189–197
5. Wang TY, Yang MH, Wu JY (2014) A sliding window approach for dynamic event-region detection in sensor networks. In: 2014 International conference on information science, electronics and electrical engineering, vol 3. IEEE, pp 2025–2028
6. Yin J, Hao Hu D, Yang Q (2009) Spatio-temporal event detection using dynamic conditional random fields. In IJCAI 9:1321–1327
7. Cheng H, Su Z, Loret J, Chen G (2014) Service oriented node scheduling schemefor wireless sensor networks using Markov random field model. Sensors 14(11):20940–20962
8. Noura A, Mehedi Hassan M, Ykhlef M, Fortino G (2019) An efficient event matching system for semantic smart data in the internet of things (IoT) environment. Future Gener Comput Syst 95:163–174

9. Michele M, Aoudia FA, Gautier M, Berder O, Benini L (2017) WULoRa: an energy efficient IoT end-node for energy harvesting and heterogeneous communication. In: Proceedings design, automation and test in Europe conference and exhibition (DATE), pp 1528–1533
10. Saritha V, Nagaraju D (2020) Energy-aware dynamic task offloading and collective task execution in mobile cloud computing. Int J Commun Syst 33(13):27–39
11. Xianjiao Z, Xu G, Zheng X, Xiang Y, Zhou W (2018) E-AUA: an efficient anonymous user authentication protocol for mobile IoT. IEEE Internet Things J 6(2):1506–1519
12. Gao T, Zhang L, Gai Y, Shan X (2007) Load-Balanced cluster-based cooperative MIMO transmission for wireless sensor networks. In: Proceedings 4th international symposium on wireless communication systems, pp 602–606
13. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wirel Commun 1(4):660–670
14. Saravanakumar R, Susila SG, Raja J (2010) An energy efficient cluster based node scheduling protocol for wireless sensor networks. In: Procedings 10th IEEE International conference on proceedings solid-state and integrated circuit technology (ICSICT), pp 2053–2057
15. Smaragdakis G, Matta I, Bestavros A (2004) SEP: a stable election protocol for clustered heterogeneous wireless sensor networks. Boston University Computer Science Department Technical Report
16. Faisal S, Javaid N, Javaid A, Khan MA, Bouk SH, Khan ZA (2013) Z-SEP: zonal-stable election protocol for wireless sensor networks. J Basic Appl Sci Res 3(5):132–139
17. Al-Janabi S, Samaher A-J, Al-Shourbaji I, Shojafar M, Shamshirband S (2017) Survey of main challenges (security and privacy) in wireless body area network
18. Anastasi G, Falchi A, Passarella A, Conti M, Gregori E (2004) Performance measurements of motes sensor networks. ACM MSWiM, pp 174–181

Image Captioning for Assisting the Visually Impaired

Sukhabogi Sandhya, Mittapalli Manaswini, and Teegala Aarthika

Abstract It is an arduous task for the visually impaired to comprehend their surrounding environment. Understanding the road signs while outdoors, knowing what is in their immediate vicinity or even reading and understanding notice boards might be a problem for them. This paper proposes an application mechanism that can assist the visually impaired people in better comprehending their surroundings. It is a process of generating a suitable textual description of the captured image by using the concept of neural networks and NLP. Along with an OCR module for reading text on notice boards and documents, the description that is generated is read out as an audio output to assist the blind and visually handicapped. A camera and a system with a GPU integrated are required. This makes it simpler for those who are blind or visually impaired to be aware of and knowledgeable about what is present in their immediate surroundings. The description's audio output can be produced in a variety of languages, including Telugu, Hindi, and English, so it can be customized to the demands of the local vernacular.

Keywords Image captioning · CNN · NLP · Text-to-speech API · OCR · Sequence processor · Visually impaired · Assisting tool

1 Introduction

For many years, vision impairment has been one of the most important global challenges. 246 million of the 285 million visually impaired people worldwide in 2010 have low vision levels, while 39 million are entirely blind, according to a WHO report titled "Global Data on Visual Impairments 2010" [1]. With a continually growing population, the number of people who are visually impaired has significantly increased.

S. Sandhya (✉) · M. Manaswini · T. Aarthika
Department of Computer Science, G. Narayanamma Institute of Technology and Science, Hyderabad, India
e-mail: s.sandhya@gnits.ac.in

Globally, there is a severe problem with visual impairment. The Lancet Global Health 2020 report [1] states that it not only has an impact on the individual's life but is also one of the causes of the world financial crisis. It also restricts a person's options for employment, which leads to unemployment and a lack of financial independence. Therefore, persons who have poor vision or no vision need support.

1.1 Methods Available for Visually Impaired People

Experts in the subject have suggested a variety of aids and methods to assist those who are blind. The most crucial prerequisite for outdoor safety is proper sidewalk use. People who are blind or visually challenged may utilize walking canes, guide dogs, GPS-enabled devices, and other aids.

Although they can detect impediments, guide dogs might be tough to communicate with by their human companions. Walking canes are undoubtedly useful for spotting abnormalities at ground level, but they are ineffective for spotting overhanging objects like cables, open windows, or tree branches. While GPS-enabled smartphones can provide directions in an outside setting, they cannot detect obstacles.

1.2 Literature Survey

Researchers have recently been concentrating on the problem of converting visual content into plain language description. Making captions for pictures is a job that applies to both computer vision and NLP. The biggest difficulty in achieving this goal is expressing in natural language how things relate to one another in the image (like English).

Historically, models have created text descriptions for photos using established templates. Additionally, search-based strategies have recently been developed to address the issue. The lack of diversity in this method prevents it from producing lexically rich text descriptions, though. With neural network's enhanced efficiency, this flaw has been eliminated. Numerous cutting-edge algorithms make use of neural networks to anticipate the following lexical unit in the output sentence while producing textual descriptions from input photos.

The successful use of sequence-to-sequence training with neural networks for machine translation served as the basis for several of the proposed approaches, which are based on RNNs. The generation of image captions also uses object search-based methods on the acquired image, although most models do not take the semantics of the image into account. Image caption creation is similar to transforming an image into a sentential format by sequential training, making it ideally suited to the encoder and decoder system.

1.3 Objectives

1. Live Capturing Image of the surrounding environment to generate a sentential description (Caption) of the captured data and an audio output of the generated caption.
2. Detecting textual data present in the environment like notice boards, banners and documents and reading it aloud (in English, Hindi, or Telugu)—This application recognizes the text present in front of it using the OCR module and generates an audio output for the visually impaired people to hear and understand in the language of their choice.

1.4 Hardware and Software Requirements

Suitable System Requirements include a robust CPU and GPU with at least 8 GB of memory, 8 GB of RAM, and an active internet connection are needed to enable Keras to download the inceptionv3/vgg16 model weights in order to train the model. Required libraries for Python (with their versions) used while testing this application are Python—3.6.7, Numpy—1.16.4, TensorFlow—1.13.1, Keras—2.2.4, nltk—3.2.5, PIL—4.3.0, Matplotlib—3.0.3, tqdm—4.28.1, Flickr8k Dataset (contains 8091 images with each containing five captions associated with each), gTTS API. The development has been done in Jupyter Notebook.

2 Proposed System

2.1 Dataflow and Architecture

The architecture consists of

1. Image Feature Extractor: It is a VGG 16 model that is pre-trained on ImageNet dataset. The outer dense layer from the model architecture is removed, and the penultimate feature extractor layer of VGG 16 model is used to preprocess our image dataset. The features extracted from original images are converted to RNN compatible feature vector dimensions. Thus, it's also known as encoder.
2. Sequence Processor: RNN, which was given the feature vectors by CNN, is the application's next stage and decodes them. Based on the projected word order, the description is constructed. The word embedding layer handles the text input, and then an LSTM RNN layer processes the caption text. LSTM helps the information to persist. It is a special type of RNN that works efficiently with any kind of sequential data fed to it. This model translates the features extracted by the image feature extractor to a natural sentence.

Fig. 1 Image and caption data pre-processing

3. Decoder: The final step is prediction, which comes after tokenization. Here, a dense layer processes the fixed-length feature extractor and sequence processor vectors in order to produce the final prediction.
4. Audio Output: Google text-to-speech commonly known as the gTTS API has been used to generate audio output from the generated caption.
5. OCR module: Pytesseract and Python Imaging Library (PIL) libraries have been utilized to read out the text present in an image in multiple languages that can be switched from the dropdown present in the user interface of the application (Figs. 1 and 2).

2.2 Implementation—Training and Testing

The steps involved in order to prepare the data is mentioned in Fig. 1. The caption dataset contains punctuations, singular words, and numerical values that need to be cleaned before it is fed to the model because an uncleaned dataset will not create good captions for the images. The following steps need to be performed:

1. Convert the captions into lowercase
2. Remove all the punctuations from the tokens
3. Tokenize the captions into different tokens
4. A vocabulary of words is created
5. Add "start_index" and "end_index" as tokens for the model to identify the start and the end of the caption
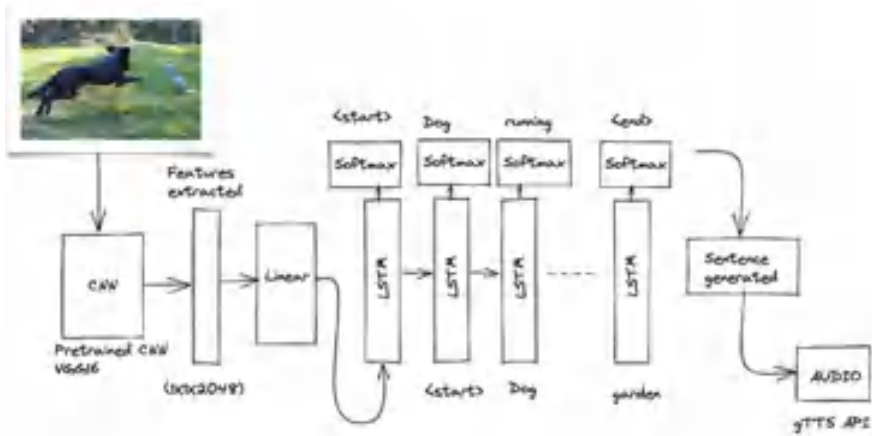6. Adding start and end sequence tokens for each captions

Fig. 2 Model architecture

7. Tokenizing the captions for further processing as the model can't take texts as an input, they need to be converted into vectors

The vocabulary will then be tokenized. This mapping can be learned using the Tokenizer class offered by Keras from the loaded captions data. Given the text-loaded image captions, it is suitable for a tokenizer. Each word in the vocabulary must be mapped with a different index value. The tokenizer.pkl pickle file will be used to construct tokens from vocabulary using a function provided by the Keras package. Since each caption for an image is a different length, start and finish sequence must be appended to the tokens to make it easier to distinguish them.

To extract the picture characteristics, the VGG16 model and weights are then loaded. The features are extracted from the input image by the pre-trained CNN, or VGG16 model. The feature vector is further converted linearly to have the same dimension as the input dimension of the LSTM network. Target caption text is predefined and is used to train the LSTM model. This is done to help the model comprehend where the labeled sequence begins and ends. Our image collection is preprocessed using the feature extractor component of VGG after the outer dense layer from the architecture is removed. This is as a result of the fact that we are only using it for feature extraction and not object classification. In the feature extraction process, VGG16 model gives out 2048 features from the input image of size 224*224.

After that, only those photographs containing a caption will be used to develop and train the LSTM model. To create the caption, the CNN's output is fed through an RNN-LSTM. Once the convolution layer has finished processing, the volume of the picture is lowered in the pooling layer. Use of the SoftMax layer allows for the multi-classification of objects utilizing formulas. The encoded result is sent to the LSTM model in the output layer, which is the final layer of the CNN model. Recurrent neural networks, or RNNs, feed continuing steps with the results of past steps. In order to anticipate the sequence based on the previous step, Long Short-Term

Memory (LSTM), an expanded form of RNN, stores both the predicted sequence and all previous steps. It gathers the necessary data from the processing of the inputs, forgets the gate, and also eliminates the unnecessary data.

We selected 2000 photos for testing and 6081 images for training from the whole dataset of JPEG-format photographs. This is a supervised learning task where the model is trained given both input and output. The training data now contains 2048-length feature vectors for each image, along with numerical representations of the captions for each image. It is not practical to store the vast amount of data created for 6081 photos in memory, so a generator approach that produces batches is utilized. The sequence of input and output will be produced by the generator. The model is trained with 100 epochs. Each caption is split into words. For an individual pair of an image and the list of captions, for each caption present in that list, the number of data points is equal to the length of the sentence. The sequential training is carried out in this manner. The model is given one word and a picture as input, and it makes predictions about the subsequent words based on those inputs. The model will then be given these two words and the image as input once more in order to generate the following word, and so on until the last word of the caption is produced. The result is that we will have two features, $\times 1$ (picture) and $\times 2$ (text sequence), as well as one target variable, y (generated word). The model is trained in this manner. Knowing the previous word must be stored in order to predict the subsequent word. This is the primary justification for LSTM use in our model.

The created image description has been automatically read out using the gTTs API as the last step. It is being read out in English right now. The OCR module is the next component of our prototype. It is situated as a button in our online application's user interface, and clicking it enables us to submit a document or a screenshot of a notice board that has been photographed. When an image is submitted, the text (in English, Hindi, or Telugu) is identified and read out using the gTTs API.

2.3 Evaluating the Model Performance

Leveraging the Bilingual Evaluation Understudy Score, models are evaluated (BLEUs). It serves as a measure for evaluating the actual and anticipated captions that have been gathered and assessed. It enumerates the degree to which the predicted text resembles the expected text. A BLEU score of 0.55 for the BLEU-1 measure has been attained, and it gets lower as the number of n-grams rises.

3 Results

The prototype built by us is a software web application which includes a dashboard with the following options as shown in Figs. 3, 4, 5, 6, 7, and 8.
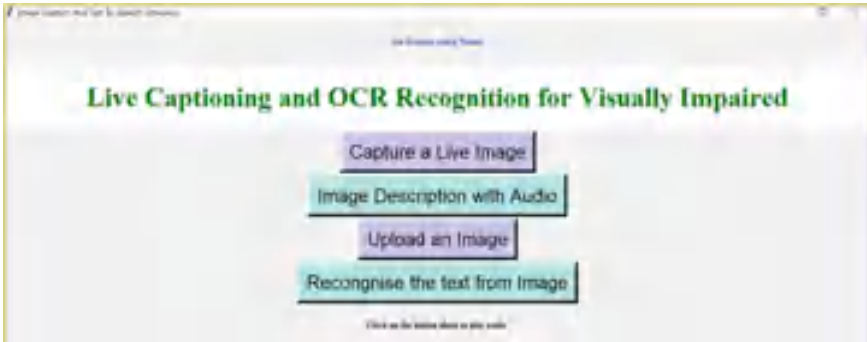
Fig. 3 Dashboard of our prototype web application
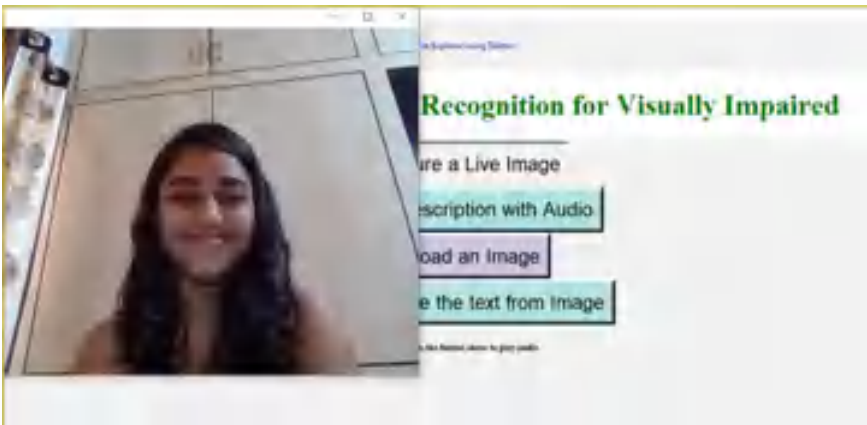
Fig. 4 Image of a girl that has been captured live

The GUI for the web application has been built using Tkinter and Flask for back end. On clicking the first button "Capture a live image," the camera module in-built in the system is opened up to capture the image in the person's vicinity. On clicking the next button, internally the image is sent as an input to the trained model, and the description is generated along with the audio output. An example of the image captioning on the prototype dashboard frontend is shown in Figs. 4 and 5.
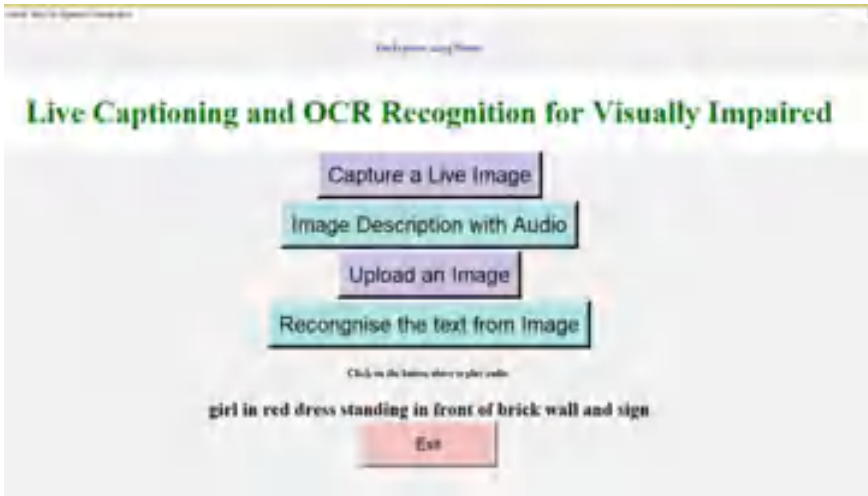
Fig. 5 The caption generated "*Girl in a red dress standing in front of a brick wall and sign*" and the audio being read out in the background

Fig. 6 Example of a sign board in Telugu

Fig. 7 Example of a sign board in Hindi
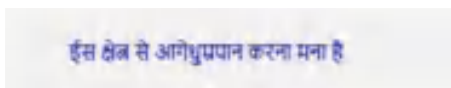
Fig. 8 OCR module identifying the text from sign boards and the output being read out in the background

4 Conclusion and Future Work

The construction of a caption generator for an image and audio output uses a CNN-LSTM model. The usage of an OCR module helps blind people better understand their surroundings in a variety of ways. To demonstrate the application, a software prototype has been designed. This model can be used for a variety of purposes, including use cases in the computer vision and natural language processing (NLP) fields, such as alternative text prediction in social media and primarily in the robotics industry. The image's text description can increase the effectiveness of content-based image retrieval, broadening the spectrum of applications for visual comprehension in industries like medical, security, and the military, among others, with a wide range of potential uses. Additional improvements can be made to the generation of captions using the attention module and several targets that are visible in an image [10]. Additionally, by employing a larger dataset and the idea of attention module, the BLEU scores can be improved.

As a future scope, this software of this prototype needs to be built upon spectacles [11] in order to make its use convenient to visually impaired people. It can also be implemented as an android application for easier and practical usage.

References

1. Vision Loss Expert Group of the Global Burden of Disease Study (2020) Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. Lancet Global Health 2020. https://doi.org/10.1016/S2214-109X(20)30425-3
2. Roentgen AR, Gelderblom GJ, Soede M-I, De Witte LP (2009) The impact of electronic mobility devices for persons who are visually impaired: a systematic review of effects and effectiveness. J Visual Impair Blin 103(11):743–753
3. Global Data on visual impairments 2010, WHO, Feb (2021). Accessed 21 Feb 2021. [Online] Available: https://www.who.int/blindness/GLOBALDATAFINALforweb.pdf.
4. MikroKopter (1 Jan 2015) [Online] Available: http://www.mikrokopter.de/
5. AscTec (15 Jan 2015). [Online] Available: http://www.asctec.de/
6. Aibotix (15 Jan 2015) [Online] Available: http://www.aibotix.de/
7. Microdrones (15 Jan 2015) [Online] Available: http://www.microdrones.de/
8. Hoffmann G, Rajnarayan DG, Waslander SL, Dostal D, Jang JS, Tomlin CJ (2018) The stanford testbed of autonomous rotorcraft for multi agent control (STARMAC). In: Proceedings. 23rd digital avionics system conference, October 2004, pp 12.E.4–1–12.E.4–10; [10] Bai S, An S (2018) A survey on automatic image captioning. Beijing Jiaotong University, No. 3 Shang Yuan Cun, Hai Dian District (Beijing, China, 2018)

9. Pedersoli M, Lucas T, Schmid C, Verbeek J (2017) Areas of attention for image captioning. In: Proceedings of the IEEE conference on international conference on computer vision, Venice, Italy, October, pp 1251–1259
10. Bai J, Lian S, Liu Z, Wang K, Liu D (2017) Smart guiding glasses for visually impaired people in indoor environment. IEEE Trans Consum Electron 63(3):258–266
11. Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc

Implementation of 64-Bit Inexact Speculative Half Unit Biased Floating-Point Adder

Vani Dasu and K. Ragini

Abstract In recent times, arithmetic operations involve very small or large numbers. These operations become easy by using floating-point numbers. Generally, the results of these operations are rounded to the nearest value. But while rounding, an error may occur. Half Unit Biased (HUB) representation is useful in avoiding this error. HUB-based numbers are obtained after the represented numbers are shifted by half a unit in the last place. For HUB-based adders, speed is restricted by how the carry propagates throughout the adder. An idea to study the effect of inexact speculative techniques on the speed of floating-point adders is put forth in this paper. Hence, an inexact speculative HUB floating-point adder is proposed. In this adder, carry is propagated in shorter paths rather than the entire architecture of the adder. Thus, the speed of adding two numbers increases when an intermediate carry is estimated by using only a few of the previous stages. In this Paper, the proposed adder has been implemented on Field Programmable Gate Array(FPGA). The speed of the proposed adder has increased by 81.19% compared to conventional 64-bit floating-point adder.

Keywords Floating-point · HUB · Unbiasing · Rounding · Adder · Inexact speculative

1   Introduction

Arithmetic computations on very small or large numbers involve rounding. Among the different rounding modes, the default mode in many processors is round-to-nearest (RN). But while rounding such numbers, an error may occur. To avoid this error, HUB representation is useful [1]. To get numbers in HUB format, Exactly

V. Dasu (✉) · K. Ragini
Department of Electronics and Communication Engineering, G. Narayanamma Institute of
Technology and Science, Hyderabad, Telangana, India
e-mail: srinagvani@gmail.com

K. Ragini
e-mail: K.ragini@gnits.ac.in

Represented Numbers (ERN) are shifted by half a unit in the last place. This shifting can be also interpreted as appending a hidden least significant digit set to one [2]. It is because, in a binary representation, shifting by half a unit in the last place is equivalent to shifting to the midpoint of two bits which makes the Least Significant Bit (LSB) as "1". HUB format simplifies rounding by truncating the extra bits. As this truncation avoids carry propagation, it simplifies the two's complement together with RN operations.

For large numbers, the carry propagation is the most time taking step in any adder. The carry propagation chain can be minimized using speculative techniques. In this Paper, speculative technique has been implemented in a 64-bit floating-point adder.

The HUB format and the inexact speculative HUB floating-point adder have been discussed in this Paper. The HUB format is presented in Sect. 2. Section 3 specifies the effect of inexact speculative techniques on the HUB floating-point adder. And the last Sect. 4 explains the FPGA implementation results of unbiased 64-bit inexact speculative HUB floating-point adder.

2 HUB Format

HUB format resembles the conventional ERN where the unit digit in the last place (ULP) is shifted by half unit [3]. Hence, numbers represented by this format are called "Half Unit Biased" (HUB)-based numbers. When compared to the conventional ERN, only the significand differs for the HUB format, but the exponent is the same. Also, the implicit least significant bit (ILSB) will be set to one due to the shifting [4, 5].

HUB representation is advantageous as two's complement is calculated only by the bit-wise inversion [5]. Also, the equivalent RN number is obtained by truncating the HUB number. The main advantage is that a HUB number is obtained only by appending "1" as the implicit least significant bit to the original number [6]. This is shown in equation below.

$$M_x = 1.M_{x-1}M_{x-2}\cdots M_{x-f}1 \tag{1}$$

HUB formats are preferred when rounding is performed. Conversion to the HUB format does not require bulky, lengthy calculations [7]. This conversion is performed by removing the Most Significant Bits (MSB) from the original number. These transactions require less additional hardware cost. Also, the error while rounding has an upper bound of 0.5 ULP. Hence, HUB format is preferred [8].

Figure 1 shows rounding in conventional ERNs, whereas Fig. 2 shows rounding in HUB format [1]. It is seen that the middle point between two ERNs in the conventional format is the ERNs in the HUB format [7]. The distance between consecutive ERNs is the same in both representations. Thus, precision is also the same [7].
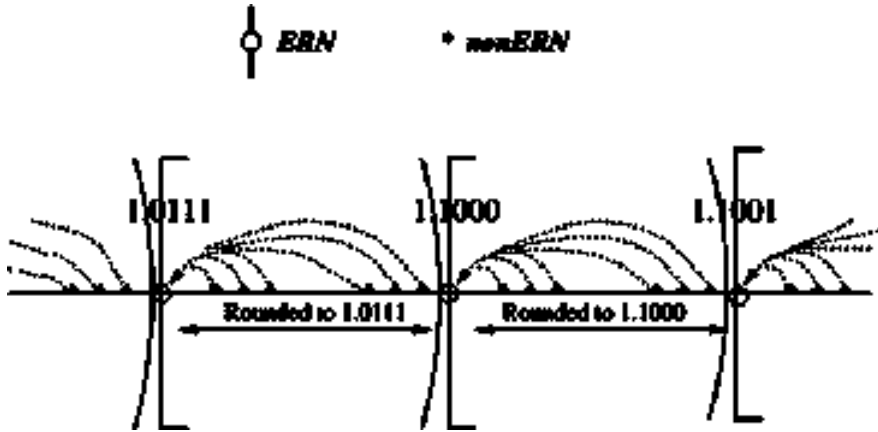
Fig. 1 Rounding in the conventional format
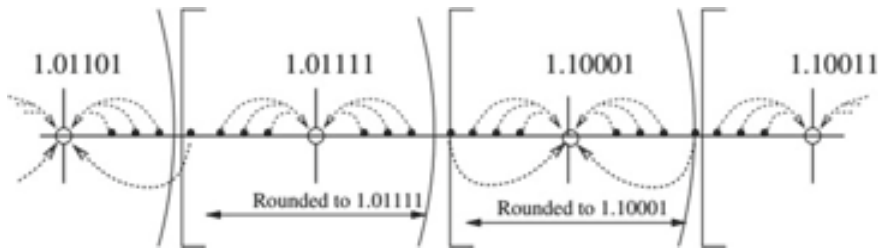
Fig. 2 Rounding in the HUB format

3 Inexact Speculative HUB Floating-Point Adder

In digital adders, the propagation of carry, however, limits the speed of addition. But in inexact speculative adders, the carry propagation chain is divided to remove this limitation [9].

Speculative adders follow the idea that carry propagation chain for addition need not span throughout the adder. Instead, an intermediate carry is estimated using a limited number of previous stages. Thus, the carry propagation chain is split which improves the speed. The carry generated by the initial sub-adder is propagated to the next position only after the second sub-adder calculates the sum of the next set of bits. This generates multiple carries concurrently and the appropriate carry for addition is choosen. Propagation delay can be drastically reduced if a carry-look-ahead adder (CLA) is used as a sub-adder [10]. Figure 3 shows the 16-bit inexact speculative adder [11]. This is extended for 64-bit numbers in this Paper.

An inexact speculative adder includes a speculator (SPEC), sub-adder, and compensator (COMP). One set of these blocks performs addition operation on four bits of the operands. These blocks are then repeated multiple times depending on
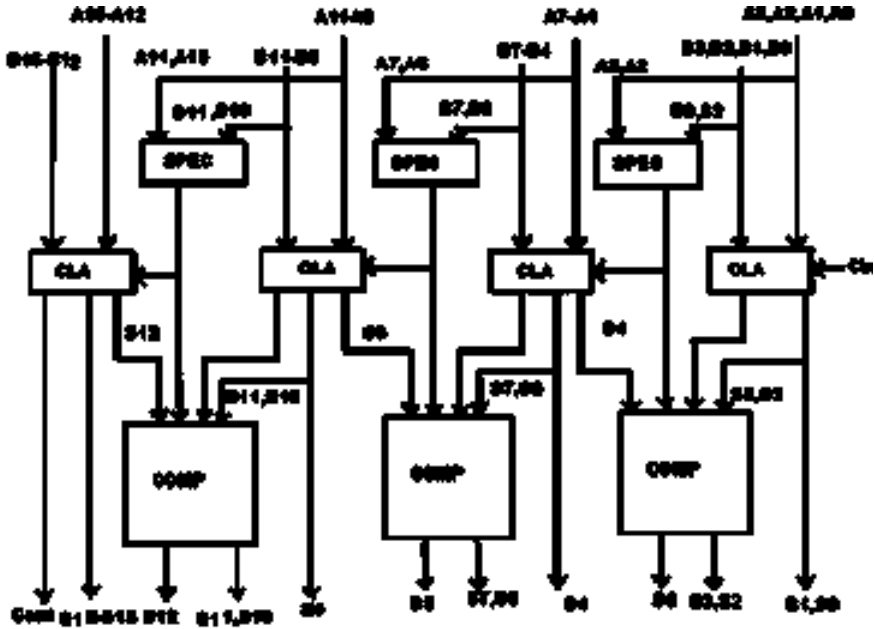
Fig. 3. 16-bit inexact speculative adder

the total number of bits of the operands. As seen in Fig. 3, these blocks are repeated three times for 16-bit operands.

The speculator block is shown in Fig. 4. This speculator block obtains MSB and MSB-1 bits from both the operands. This block generates carry according to the structure seen in Fig. 4 and is denoted by Cso. The Cso bit is the carry for sub-adder block performing addition for the next set of four bits.

The sub-adder block is a 4-bit CLA which takes only 4-bits as input. The compensator block uses an XOR gate which looks out for discrepancies between speculated carry and carry from the previous sub-adder. If both these values are different, an error flag(fe) is generated, which causes one of the two compensatory techniques either error reduction or correction to be activated [12].

If the output of the XOR gate is 0, the sum directly passes to the final output. Or else, the XOR gate generates "1" which indicates an error has occurred. This error can be either positive or negative depending on the actual carry and speculated carry. Faults can occur only in two right-hand paths. The first LSB of the middle path is correctable, this is mentioned in the Fig. 5 as "S4 corrected". The first LSB of the right path cannot be corrected, thus first MSB of the subsequent sum is flipped to balance the output [12]. All these steps are well depicted in Fig. 5. In the process of balancing, a slight error may be introduced. As a slight error may be introduced due to the use of speculative techniques, this kind of adder is called inexact speculative adder(ISA).

Fig. 4 Speculator block used in inexact speculative adder
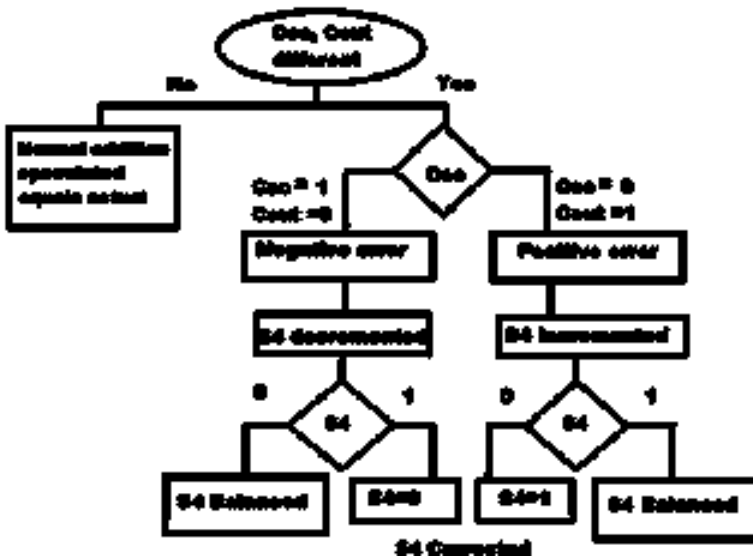
Fig. 5 Speculation bits calculation for inexact speculative adder

The 16-bit inexact speculative adder described above is used to implement the addition of 64-bit HUB floating-point numbers by repeating the blocks. The advantages of both the HUB format and the inexact speculative techniques are thus present in the proposed 64-bit inexact speculative HUB floating-point adder. These results are discussed below.
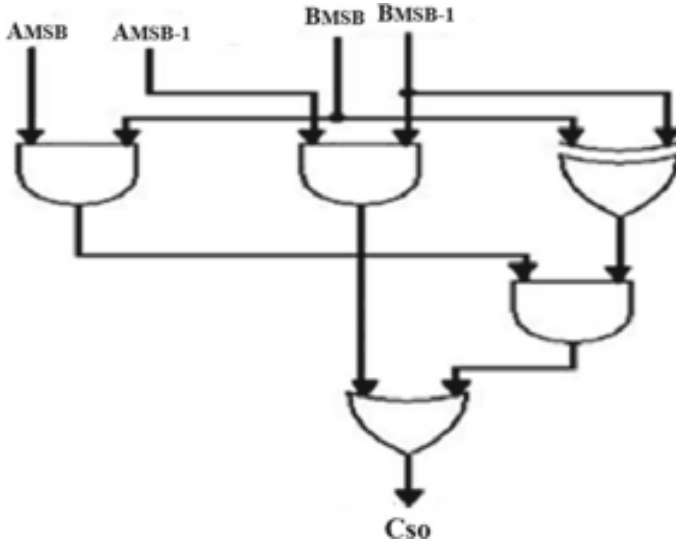
4   Implementation Results

Both conventional 64-bit floating-point adder and 64-bit inexact speculative HUB floating-point adder have been implemented using Xilinx FPGA Nexys A7 board. Table 1 compares the delay of a conventional 64-bit floating-point adder and the 64-bit inexact speculative HUB floating-point adder. Thus, it is seen that the delay of the proposed adder is less than the conventional adder by 81.19%. This change is mainly because of two reasons: usage of the HUB format and implementation of the speculative technique. By using the speculative technique as specified in Sect. 3, the addition of bits given to the sub-adder, and the calculation of the carry for the next sub-adder both steps happen parallelly. This helps in reducing the length of the carry propagation chain. Hence, as the carry for the second sub-adder is speculated before the addition of the second sub-adder starts, the delay in the addition process is reduced. This technique is implemented throughout the 64-bit floating-point adder. Also, the result thus obtained is rounded to the nearest value easily and quickly as the HUB format is used. Both these reasons help in reducing the delay of the 64-bit inexact speculative HUB floating-point adder.

The FPGA implementation of the 64-bit conventional floating-point adder and 64-bit inexact speculative HUB floating-point adder is shown in Figs. 6 and 7, respectively.

Table 1   Comparison of delay between two architectures

Parameter	Conventional 64-bit floating-point adder	64-bit inexact speculative HUB floating-point adder
Delay(ns)	60.741	11.421

Fig. 6   FPGA implemented circuit for 64-bit conventional floating-point adder

Fig. 7 FPGA implemented circuit for 64-bit inexact speculative HUB floating-point adder
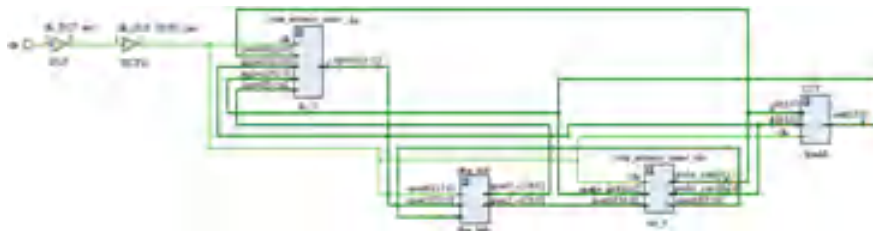
5 Conclusions

In this Paper, inexact speculative technique has been applied to the HUB floating-point adder. Implementation of 64-bit inexact speculative HUB floating-point adder architecture has also been done on the FPGA board. From the results, it is concluded that the 64-bit inexact speculative HUB floating-point adder runs faster as compared to conventional methods but at the cost of slight inaccuracy. The inaccuracy is in the order of 10^{-14}. When high-speed floating-point arithmetic operations are required and a reasonable amount of inaccuracy can be tolerated, this architecture can be used.

References

1. Hormigo J, Villalba J (2016) New formats for computing with real-numbers under round-to-nearest. IEEE Trans Comput 65(7):2158–2168
2. Ramesh S, Saravanavel K (2021) HUB floating-point adder using double path. Int J Scientif Res Sci Technol (IJSRST) 9(1):708–717
3. Shaik A, Fairooz SK (2020) HUB floating-point addition using unbiased rounding. IOSR J VLSI and Signal Process (IOSR-JVSP) 10(1):09–14
4. Lahari M, Agrawal S (2020) Efficient floating-point HUB Adder For FPGA. In: 2020 4th international conference on electronics, materials engineering and nano-technology (IEMENTech), Kolkata, India
5. Villalba-Moreno J, Hormigo J, González-Navarro S (2018) Unbiased rounding for HUB floating-point addition. IEEE Trans Comput 67(9):1359–1365
6. Srinivas Reddy T, Shekar CH, Prabhakar J (2019) Analyze and Implementation of FPGA Implementation of HUB floating-point addition. Int J Recent Technol Eng (IJRTE) 7(ICETESM)
7. Villalba-Moreno J (2016) Digit recurrence floating-point division under HUB format. In: 2016 IEEE 23nd symposium on computer arithmetic (ARITH), Silicon Valley, CA, USA, pp 79–86
8. Hormigo J, Villalba J (2016) Measuring improvement when using HUB formats to implement floating-point systems under round-to-nearest. IEEE Trans Very Large Scale Integra (VLSI) Syst 24(6):2369–2377
9. Swetha Bai K, Vijay Murugan S (2020) High speed inexact speculative adder using carry look ahead adder and brent kung adder. Int Res J Eng Technol (IRJET) 7(3):1006–1010

10. Reethika Rao V, Ragini K (2014) Comparative analysis of 32 bit carry lookahead adder using high speed constant delay logic. Int J Sci Eng Technol Res (IJSETR) 3(10):2768–2774
11. Shrestha R (2017) High-speed and low-power VLSI-architecture for inexact speculative adder. In: International symposium on VLSI design, automation and test (VLSI-DAT), Hsinchu, Taiwan, 24–27 April
12. Camus V, Schlachter J, Enz C (2015) Energy-efficient inexact speculative adder with high performance and accuracy control. In: 2015 IEEE international symposium on circuits and systems (ISCAS), Lisbon, Portugal

Interview Supporting System Using Facial Features

Kondur Datha Vaishnavi, Kokkonda Ranee Prathyusha, A. Sharada, and Sahithi Kalluri

Abstract In this era of technology, recruiting candidates to fit a particular job profile is a crucial task for all the companies. The traditional way of recruiting has changed over time. Determining the confidence of a person in an interview acts as a major factor in the recruiting process. A typical human recruiter can handle up to 20–50 candidates at a given time, but a machine can perform the same task more effectively in a short amount of time. The Interview Supporting System in the field of education acts as a helping hand for students and improves their performance based on the utterances and facial features. The main goal is to develop a system which determines the confidence level of the interviewee based on the number of filler words in their answers and their emotions using HOG features. The confidence levels are classified into very confident, neutral, and poor. This Supporting System helps students to estimate and enhance their skills before attending the interview.

Keywords Emotion recognition · Confidence · HOG

1 Introduction

In recent years, online interviews are increasingly used in the hiring processes as the popularity of video-based job interviews rises, so does the necessity for automated tools to evaluate interview performance.

Lack of feedback:

Due to the recent economic decline, some organizations have been receiving huge number of applications, because of which providing feedback is problematic. However, providing unsuccessful candidates with proactive and timely feedback is

K. D. Vaishnavi · K. R. Prathyusha · A. Sharada (✉)
G. Narayanamma Institute of Technology and Science for Women, Hyderabad, India
e-mail: sharada@gnits.ac.in

S. Kalluri
Accenture Netherlands, Business Arch Integration Team Lead, Almere, Netherlands

beneficial to the candidates. Not all candidates want to receive feedback, except for those that do it's vital they receive the feedback that can help them grow and give them something to build on. 70% of candidates receive no feedback after being rejected post-phone screen, and 54% receive no feedback after being rejected post-interview.

Confidence Levels as a feedback:

Facial expressions are the attributes to portray the emotions of the person. A confident face may be a very important accent in interviews. In some cases, filler words act as an important part, and an interview can be affected by the filler words [3]. How confident was the interviewee? [10] and had to offer a label based on a Likert scale from 1 (Not Confident) to 7 (Very Confident).

2 Literature Survey

The aim of this paper an automated system for the assessment of interview performance through Audio and Emotion Cues Priya et al. [2] is to determine the performance of the interviewee based on the audio and emotion cues. Evaluating the performance of a student and giving feedback in an interview is highly unexplored and also considered as a challenging task. The proposed system evaluates the performance of an interviewee and gives feedback to the users. An input video which consists of an interview process is split into audio and visual frames.

Automated Prediction and Analysis of Job Interview Performance: The Role of What You Say and How You Say It Naim et al. [1] is to automatically estimate the quality of individual workers using an EM-style optimization algorithm, and estimate a weighted average of their scores. For the prediction framework, they automatically extracted 82 features from the interview videos and trained two regression models: support vector regression (SVR) and Lasso. The aim of this training is first, to predict the Turker's ratings on the overall performance and each behavioral trait, and second, to acquire meaningful insights on the relative importance of individual features for each trait. They collected three types of features for each interview video like prosodic features, lexical features, and facial features.

How Confident Are You? Exploring the Role of Fillers in the Automatic Prediction of a Speaker's Confidence Dinkar et al. [9]

Filler words remain unexplored and they are overlooked as a noise. The proposed system works on the prediction of FOAK, which is a quite challenging task for educational applications. They design a set of filler features based on linguistic literature and investigate their potential in FOAK prediction. The goal of this system is to system provides an analysis on how different filler words are interconnected with the confidence.

3 Dataset

Training dataset: In data science or machine learning, a dataset is the information or data that is needed. The information is usually gathered manually. In most systems, there are two or three datasets: a training dataset, a development dataset, or a validation and test dataset. Models are built using training datasets. The Interview Supporting System is mainly focused on the confidence prediction which is based on the audio and facial expressions. The training data need to contain the interviewee's faces which express their emotions. The crop_dataset is the training dataset for predicting emotions which consists of images of five different emotions which are happy, surprise, neutral, sad, and fear which are the collection of few images from the CK+ emotion dataset. While creating the crop_dataset, the images collected from the CK+ dataset are cropped and then saved to the crop_dataset. crop_dataset consists of around 25–150 images for each emotion and consists of a total of 331 images.

The new_data.csv is the training dataset which is used to predict the confidence level. The new_data.csv file consists of different answers and also the information of the number of filler words, ratio, and confidence of each answer.

4 Methodology

Audio and facial cues are extracted from the video [5]. Audio is extracted from the video, and this audio file is again converted into .wav format to perform speech-to-text conversion. Filler words are extracted, and based on the filler words, ratio is calculated. Key image frames are extracted from the video, and emotion recognition is performed. The machine learning model predicts the confidence level of a person for each answer and their emotions from the images. The overall confidence is determined as one of the three categories not at all confident, neutral, and very confident. Resulted confidence is based on the performance of the Interviewee.

5 Confidence Level Prediction Through Speech Recognition

The confidence of person can be recognized through speech and their fluency. So the audio of the interviewee is collected from the video. Confidence levels are determined through speech recognition. Speech-to-text conversion is the main aspect in this particular phase. It is used to convert the audio into text. The audio consists of interviewee answers. This text is further used to determine the confidence of the interviewee based on the number of filler words used by them (Fig. 1).
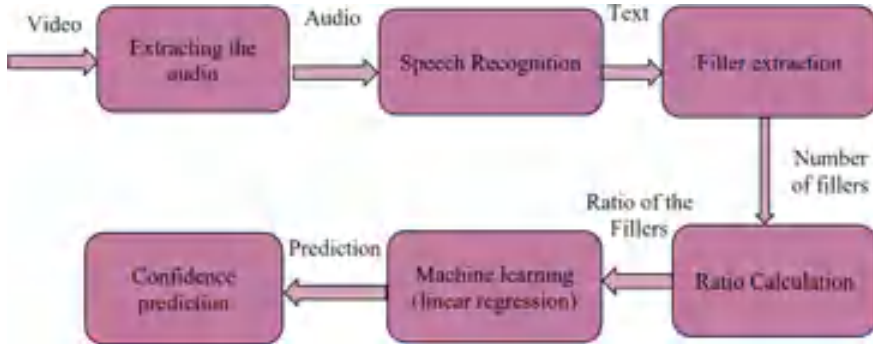
Fig. 1 Workflow of the audio confidence prediction

5.1 Extracting the Audio

MoviePy is a Python library used for editing videos. By using this MoviePy library, the audio (mp3) is extracted from the video (mp4). The audio which is in mp3 format is changed to .wav format using FFmpeg through command prompt. The audio is used in this system to identify the confidence levels, and the video is used for emotion recognition.

5.2 Speech-to-Text Conversion

Speech recognition is an ability of a machine to understand what humans are speaking. Speech recognition identifies words or phrases in spoken language and converts into human-understandable text. In this proposed system, speech-to-text conversion is done in Python using speech recognition library and Google speech recognition.

5.3 Fillers

In speech, filler words are defined as short, meaningless words (or sounds) that are used to fill the little pauses that occur while they decide what one is going to say next. The number of filler words may affect the confidence level. Using excessive filler words can reflect poorly on your communication skills because the person may not be able to understand your idea.

Based on the number of fillers that occur in the person's speech, the ratio is calculated and the level of confidence of the person is determined. The value of the ratio depends on number of filler words to the total number of words. The fillers in

Table 1 Ratio and confidence table

Ratio	Confidence scale	Confidence level
0–0.14	7	Very confident
0.14–0.28	6	Very confident
0.28–0.42	5	Very confident
0.42–0.56	4	Neutral
0.56–0.70	3	Not at all confident
0.70–0.84	2	Not at all confident
0.84–0.98	1	Not at all confident

an answer are identified by the words present in the text file in the path provided, if it identifies filler words in the answer, it counts as 1 and it continues the step until the last word of the answer and finally gives the overall count of the number of filler words present in the answer, if it doesn't find then it gives 0 fillers.

5.4 Calculation of Ratio

After the extraction of the fillers, the ratio is calculated. The ratio which is calculated is based on the number of filler words present in an answer to the number of total words in that particular answer. Based on the value of the ratio, the confidence score is determined as per confidence score [9] (Table 1).

$$\text{Ratio} = \frac{\text{The Number of Fillers in an answer}}{\text{The Number of Words in an answer}}$$

5.5 Confidence Level

The confidence levels are determined on the scale of 1 to 7 based on the confidence score predicted by the machine learning. Linear regression is a machine learning algorithm that is used to predict the confidence levels based on the training data. We are considering confidence of each answer and calculating the average confidence. This average confidence is the overall confidence of the interviewee (Table 2).

Table 2 Confidence determination table confidence score

Confidence score	Confidence level
5–7	Very confident
4	Neutral
1–3	Not at all confident

6 Confidence Level Prediction Through Emotion Recognition

Emotion recognition plays an important role in this system to determine the confidence levels of an interviewee. Five different kinds of emotions are identified such as happy, surprise, neutral, sad, and fear (Fig. 2).

The first two steps of the proposed system are to take video as an input and convert that video into a number of frames and extract the key frame from the sequence of frames (every 25th frame is considered as a key frame) and perform HOG feature extraction and classification of features using multimodal SVM is performed (Fig. 3).

6.1 Key Frame Extraction

The prerequisite of this phase is the video input and video to frames conversion. Analyzing each and every video is a time-consuming process. So to reduce the time consumption, key frame extraction technique is used. This technique is performed on the input video. Every 25th frame is extracted and these frames are used to detect the emotions.

If the input video is of 3 min captured at a frame rate of 25 frames per second, there will be a total of 4500 images. A key frame is extracted for every 25th frame,

Fig. 2 Five different kinds of emotions
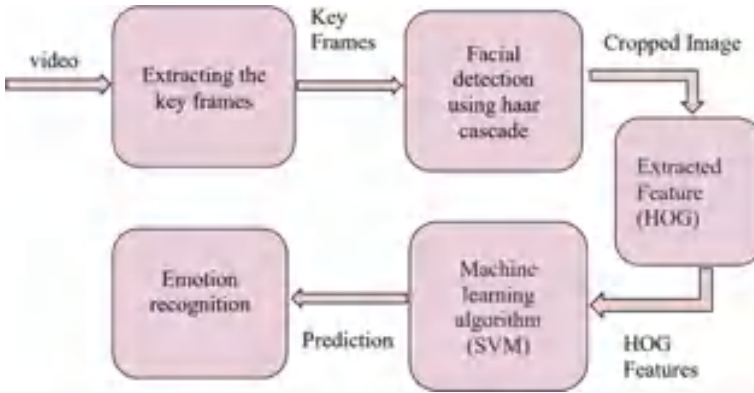
Fig. 3 Workflow of the emotion recognition

and hence at the end of the key frame extraction method, only 180 images are taken for further processing.

6.2 Face Detection

Paul Viola and Michael Jones proposed an object detection using Haar feature-based cascade classifiers which may be used for face detection. It depends on machine learning where positive and negative images are fed to a cascade function for training purpose so that it can be used to detect objects in images.

OpenCV performs face detection using a pre-trained Haar cascade because of which there is no need to train a classifier with positive and negative images.

A pre-trained classifier is to be loaded to detect the faces for example:

filename = 'data/haarcascade_frontalface_default.xml'...

face_cascade = cv2.CascadeClassifier(filename)

The following function call is used by the classifier to detect the faces.

faces = face_cascade.detectMultiScale(img_gray,...)

6.3 HOG

Feature extraction is a process that involves extracting valuable information from an image. It then throws away irrelevant data. The facial features are then extracted to determine the facial expression. The distribution of gradients and directions is the feature in the HOG descriptor [4]. The size of gradients around the edges and

Points scored
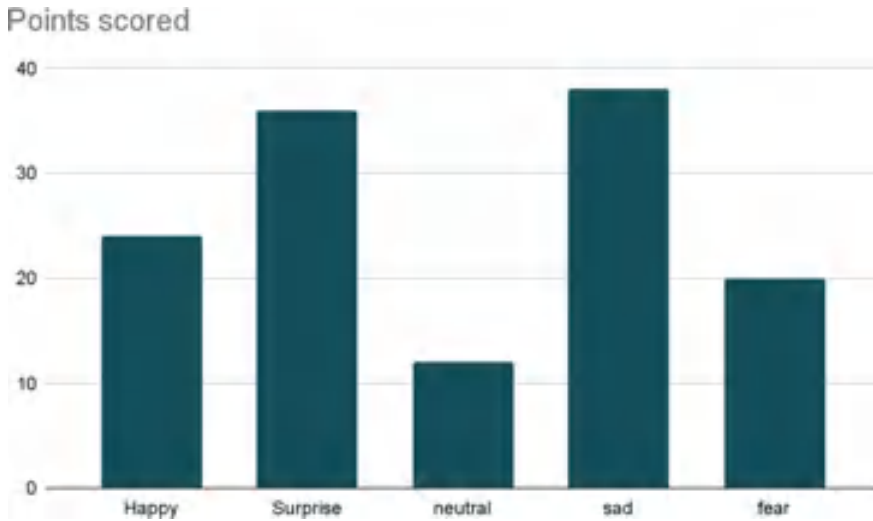
Fig. 4 Emotions recognized from the video

corners is extracted from gradients. As the peak of the edges and corners contains a lot of information about object regions and flat regions, HOG is a vital feature of this application.

6.4 Emotion Recognition

Support vector machine is a machine learning model that is used to classify facial expressions. The SVM is trained using the extracted HOG features [8]. The classified emotions are happy, surprise, neutral, sad, and fear [7].

Emotion is recognized for each and every frame. The recognized highest face emotion from the input video will be determined as the output (Fig. 4).

6.5 Overall Confidence Calculation

Based on the obtained confidence score, the performance of an interviewee is determined. The performance of a candidate in an interview is analyzed by the proposed algorithm. The emotion from facial expression is classified with five classes and confidence of the speech is classified with three classes. Totally 15 possibility classes are used to analyze this performance evaluation. The performance of the candidate is classified into the high, medium, and poor. This is shown in Table 4. The emotion predicted by the SVM and the confidence obtained for the text are fused, and the

Table 3 Labeling of face emotions and confidence of the text

Audio Cues emotions	Very confident	Neutral confident	Not at all confident
Happy	7	6	4
Surprise	6	5	4
Neutral	5	4	3
Sad	3	2	1
Fear	3	2	1

Table 4 Performance rating scale

Performance score	Rating
5–7	High
3–4	Medium
2–1	Poor

overall performance is determined. This system computes the confidence of audio and facial features from the video, and the score is calculated by using the following (Table 3).

7 Results

Emotion recognition from video and speech recognition from audio are the two phases, and the results of these two phases are used to determine the performance of the interviewee in terms of confidence levels.

Three different input videos have been tested, and a dataset is created in which all the crop images are stored which are taken from CK+ emotion dataset [6]; it is trained and tested for emotion recognition from video.

A 30 s video is taken as input from which emotion and confidence level of a person through filler words is predicted. Based on the combined score, the overall confidence of a person is predicted.

Comparison Analysis for Algorithms:

Linear regression and SVM are two algorithms used to predict the confidence of a person, based on the filler words used by them in the interview. Support vector machine is a classification algorithm to predict a continuous variable. While linear regression models minimize the error between the actual and predicted through the line of best fit. Out of the two algorithms linear regression gives better performances with the accuracy of 99% where training data is 80% and testing data is 20% (Tables 5 and 6).

Table 5 Comparison of algorithms used in the system for confidence prediction

Model	Training data (%)	Testing data (%)	Accuracy (%)
Support vector machine	80	20	89.5
Linear regression	80	20	77
K-Nearest neighbor	80	20	29
Naïve Bayes	80	20	55

Table 6 Comparison analysis of algorithms predicting confidence levels based on fillers

Model	Training data (%)	Test data (%)	Accuracy (%)
Linear regression	80	20	99
	70	30	91
Support vector machine	80	20	85
	70	30	80

In Fig. 5: X-axis represents training data and Y-axis represents accuracy of the algorithm.

SVM model has been used to predict the emotions of a person which gives an accuracy of 89%.
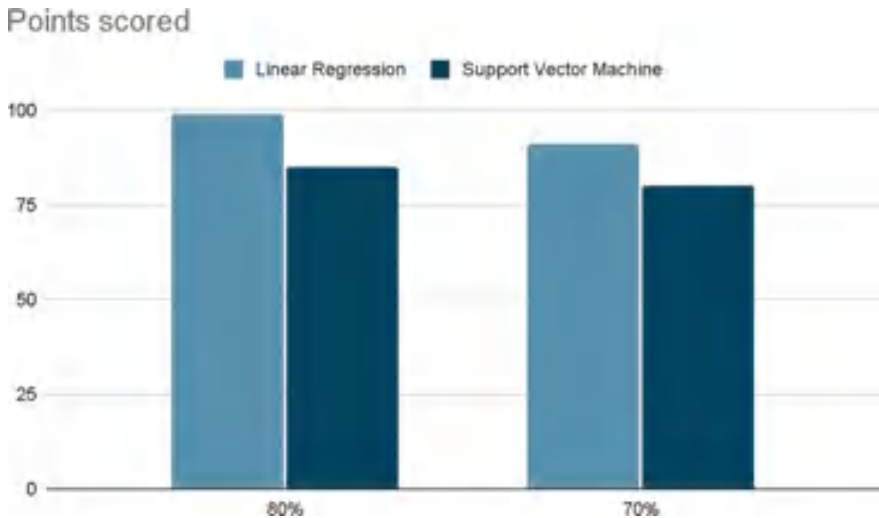

Fig. 5 Comparison of linear regression and SVM

8 Conclusions

This interview supporting system helps students to improve their skills by determining their performance in terms of confidence levels as feedback using facial features and speech of the interviewee. This proposed system provides an automatic analysis to determine the interviewee behavior.

References

1. Singh AP, Kumar A (2019) Robust face recognition system using HOG features and SVM classifier. Int J Inform Sci Appl (IJISA) 11(1) ISSN 0974-2255
2. Evangelista–Pelaez C, Tutor-Del Mundo C, Christopher Dayuta K (2018) In: Conference paper, fillers used in a mock job interview: A cognitive analysis, April 2018
3. Naim MI. Tanveer I, Gildea D, (Ehsan) Hoque M (2015) Automated prediction and analysis of job interview performance: the role of what you say and how you say it
4. Li J, Oussalah M (2010) In: Automatic face recognition system
5. Priya K, Seturaman MG, Kalaivani P, Shanmugavadibu P, Mohamed/An S (2019) In: Conference (ICACCS): automated system for the assessment of interview performance through audio and emotion cues
6. Lucey P, Takeo Kanade JFC, Saragih J, Ambadar Z (2010) Computer society conference on (CVPRW): The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, 2010 IEEE, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA
7. Ekman P (1993) Facial expression and emotion. Amer Psycol 48:384–392
8. Kosti R, Alvarez JM, Recasens A, Lapedriza A (2020) Context based emotion recognition using EMOTIC dataset
9. Dinkar T, Vasilescu I, Pelachaud C, Clavel C (2020) How confident are you? Exploring the role of fillers in the automatic prediction of a speaker's confidence
10. Li X, Zhang X, Wang H (2008) Real-time optimization of Viola-Jones face detection for mobile platforms. In: 2008 IEEE dallas circuits and systems workshop: system on-chip—design, applications, integration, and software

Dragon Fruit Stem Disease Detection Using Image Processing

Y. Rakesh Kumar, P. Satyanarayana Goud, V. Radha Krishna, Chandra Shaker Arrabotu, V. Samhitha Reddy, and G. Sahithi

Abstract Dragon fruit is generally tolerant to major diseases concerned; but due to improper knowledge and poor climatic conditions, there might some diseases which effect the growth of the dragon fruit. Detection of these diseases is more significant to avoid any loss to the farmer. Detection of dragon fruit stem diseases manually for large scale farming is difficult and depends on expertise of the farmer on diseases; so a computer-aided method like image processing technique is required. A color features-based multi-threshold segmentation method is proposed, and diseases classification is done based on the statistical features like entropy, contrast, energy, and homogeneity.

Keywords Dragon fruit stem diseases · Segmentation · Classification · Multi-thresholding · Statistical features

1 Introduction

In the Indian market, dragon fruit is a freshly introduced superfruit. Due to its appealing fruit color, delectable pulp with edible black seeds embedded inside, nutraceutical value, excellent export potential, and highly lucrative nature, it is gaining enormous popularity among growers. It produces a yield from 14 to 16 months after stem-cutting planting and a yield of up to 20 years with a long crop cycle from May to December in various flushes every year. The red-fleshed pitaya Hylocereus costaricensis and the white-fleshed pitaya Hylocereus undatus are the two main species flourishing in India. Various diseases that occur in dragon fruit stems are soft rots, Botryosphaeria dothidea, Colletotrichum gloesporiodes, Bipolaris Cactivora, and Cactus Virus X.

Y. Rakesh Kumar (✉) · P. Satyanarayana Goud · V. Radha Krishna · C. S. Arrabotu · V. Samhitha Reddy · G. Sahithi
G. Narayanamma Institute of Technology and Science (for Women), Hyderabad, India
e-mail: rakeshyacharam@gnits.ac.in

In the Indian market, dragon fruit is a superfruit [1] and is currently extremely well-liked by growers because of its delicious pulp with edible black seeds embedded inside, appealing fruit color, strong export potential, and highly lucrative nature.

Patwary et al. [2] discussed on different growth stages of dragon fruit; Gopalu Karunakaran et al. [3] discussed dragon fruit cultivation growth and reported on growing states in India. Downer's [14] on pitahaya diseases describes the various diseases of pitahaya and the management of these diseases. It explains the causes of the diseases like Soft rot, Anthracnose, etc. Sijam et al. [4] isolated and identified the pathogenic bacteria causing disease on dragon fruitlike soft rot.

Red-fleshed dragon fruit stem canker was identified by Salleh et al. [5] (Hylocereus polyrhizus). To specifically investigate the relationship between the prevalence of disease and environmental conditions, a study was conducted to isolate and identify the pathogenic bacteria causing disease in dragon fruit. It is possible to identify different diseases using physical indicators, genetic traits, and molecular traits.

According to Arya et al. [6], acquiring RGB images is one of the phases in the detection of sick plant leaves, and convert the RGB image to HSI format of the input image, then masked for the green pixels elimination. Otsu's approach is used to divide the parts into segments. Using the color-cooccurrence methodology, texture features are calculated, and then a genetic algorithm is used to categorize the condition.

A study on the application of digital image processing techniques to identify, measure, and categorize plant diseases from Digital Photographs is presented by Ankita Das [7]. Various image processing-based methods for detecting leaf diseases are proposed [8, 9, 12]. The fruit disease detection using image processing is presented in [10, 11]. The feature extraction methods GLCM and color moments are combined in [13], proposed to distinguish three types of dragon stem diseases smallpox, stem rot, and stem sting. The optimal accuracy for the SVM classifier is 87.5% and for KNN is 73.33%.

The primary goal of this effort is to assist dragon fruit growers in determining if dragon stems are healthy or not and identifying the types of spots that may be present. To determine if the plant is diseased or not, the image of the diseased stem is captured, and some operations are then carried out on it. The procedure of segmentation is used to isolate the diseased area, and the stem diseases are ranked by determining the quotient of disease spot. By doing this, the computational load and storage needs are reduced without compromising the quality of the segmentation output. For segmentation, a variety of algorithms is employed. For segmentation, the histogram multi-channel thresholding method is employed.

The process of extracting information classes from a multiband raster image is referred to as image classification. The types of spots on the stems of dragon fruit are detected, and images are categorized as healthy and unhealthy. Statistics like entropy, contrast, homogeneity, and energy are used to categorize images.

2 Proposed Multi-threshold Segmentation and Statistical Parameter-Based Classification Algorithm

Dragon disease detection and classification using computer vision are proposed. To segment the diseases on dragon stem, a multi-threshold algorithm is developed, and then to classify the type of disease a statistical parameter-based algorithm is proposed. The complete flow of segmentation and classification is shown in Fig. 1.

2.1 Image Acquisition

It is the procedure used to convert obtained data into the required output format. To begin, processing an analog image for this application is first transformed into a digital image. In our dataset, various pictures of the stem of the dragon fruit have been taken. Our approach may be used for any other image outside of the dataset as long as its size and format are compatible. Our system's classifier then predicts the disease for that image.

2.2 Image Preprocessing

The processes for picture segmentation, image enhancement, and color space conversion are included in this step. The digital image is first scaled. After that, a channel reduction approach is used to reduce the stem picture. The subsequent analysis uses the image that was obtained.
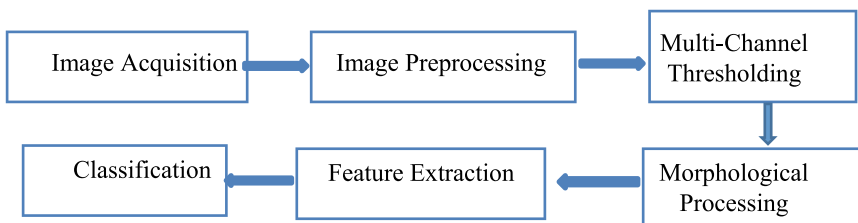

Fig. 1 Block diagram of proposed segmentation and classification of dragon stem diseases

2.3 Multi-threshold-Based Dragon Stem Disease Segmentation

The segmentation of dragon stem disorders is done using multi-channel thresholding. Thresholding is a technique used in image processing to divide an image into smaller chunks, or junks, with at least one color or gray scale value defining their boundary. First, the RGB image is transformed to the green channel component, and the remaining first and second channels of the cropped images of the damaged and healthy parts of the dragon fruit stem are subjected to histogram analysis. The difference between healthy and unhealthy was visible in the R and B channels histogram, which aided in determining the threshold values.

Unhealthy regions in the dragon stem can be segmented using a threshold for R values greater than 60 and a threshold for B channel less than 140, else it is decided as the healthy region.

2.4 Morphological Processing

The most basic morphological operations are Dilation and Erosion. Dilation adds pixels to the boundaries of objects in an image, while Erosion removes pixels on object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. In the morphological Dilation and Erosion operations, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image. The rule used to process the pixels defines the operation as a Dilation or an Erosion. Here, the undesired artifacts are removed using the Erosion process, and only the diseased part is retained. This part is represented using white pixels and the remaining by black pixels.

2.5 Feature Extraction

Statistical features are contrast, energy, correlation, and homogeneity. Entropy is used for the description of the image. These features are calculated using the following formulae.

Energy. It is a gray scale image texture. Homogeneity measures changing and reflecting the gray scale image distribution uniformity of weight and texture, given in Eq. (1).

$$E = \sum_x \sum_y p(x, y)^2, \tag{1}$$

where $p(x, y)$ is the GLCM

Contrast. Contrast is the main diagonal near the moment of inertia, which measures the value of the matrix, which is distributed, and images of local changes in the number, reflecting the image clarity and texture of shadow depth, given in Eq. (2).

$$I = \sum \sum (x - y)^2 \, p(x, y) \tag{2}$$

Correlation. The Correlation Coefficient measures the joint probability occurrence of the specified pairs of pixels, given in Eq. (3).

$$\text{sum}(\text{sum}(x - \mu x)(y - \mu x)p(x, y)/\sigma_x \sigma_y), \tag{3}$$

where μ is mean, σ_x is the variance of event x, and σ_y is variance of event y.

Homogeneity. Closeness is the measure of the elements distribution in the GLCM to the GLCM diagonal, given in Eq. (4).

$$\text{Homogeneity} = \text{sum}(\text{sum}(p(x, y)/(1 + [x - y]))) \tag{4}$$

Entropy. Entropy measures the disorder of an image, and it will achieve its largest value when all elements in the P matrix are equal. When the image is not texturally uniform, many GLCM elements have very small values, which implies that entropy is very large. Therefore, entropy is inversely proportional to GLCM energy, given in Eq. (5).

$$\text{Entropy (ent)} = -\sum \sum P_i \, \log_2 P_i \tag{5}$$

2.6 Multi-threshold Dragon Stem Classifier

For the classification function, statistical feature parameters used are energy, contrast, homogeneity, and entropy. Using these statistical functions, diseases are classified in the image. They also help in detecting the type of disease that the dragon stem may be affected by, such as a dark spot, a white pail, and a medium red spot. These three target classes are considered and a multi-threshold classifier algorithm is proposed.

A multi-threshold dragon stem classifier algorithm is constructed after analyzing the statistical feature values of entropy, homogeneity, and contrast. Threshold points of these statistical features are estimated to classify the types of dragon stem diseases. Disease types are classified using contrast and entropy values. Contrast less than 90 or entropy less than 11 are used to detect the dark spot dragon stem disease. If the entropy is less than 13 and the contrast is greater than 400, the disease is classified as a white pail; otherwise, it is classified as a medium red spot.

3   Results and Discussions

Images of healthy and unhealthy dragon stem images (data set) are collected from about 252 from [14–17] to analyze and evaluate the performance of the proposed method in segmenting and classifying dragon stem diseases. Simulation is carried out in MATLAB environment for segmentation and classification. From segmentation, histograms of cropped unhealthy and healthy parts of dragon stem are generated and analyzed for estimating multi-thresholds on HSV images, respectively. Based on features calculated for diseased image and non-diseased image, the type of disease is predicted finally. The original images of diseased and healthy dragon stems are shown in Figs. 1 and 2.

After preprocessing and the green channel subtracted diseased dragon stem, the image shown in Fig. 3. Then, multi-channel thresholding applied in Fig. 3 results in Fig. 4. After thresholding, the diseased part is extracted; but along with it, we can also see that some artifices are misleading as diseased regions. To overcome these artifacts, morphological operations are applied, and the resultant image is shown in Fig. 5.

Fig. 2 Diseased dragon fruit stem

Fig. 3 Healthy dragon fruit stem

Fig. 4 Preprocessed image

Fig. 5 Image after thresholding
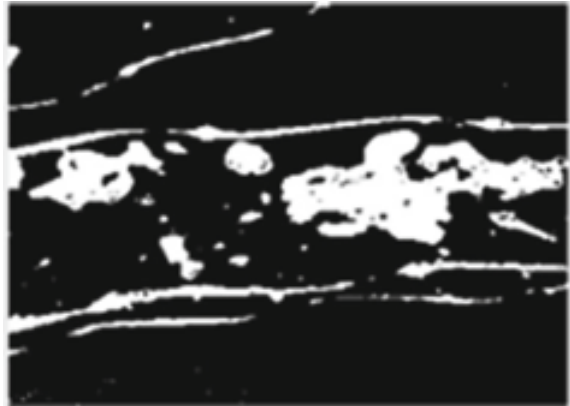
Classification results after segmenting the disease region in a given dragon stem diseased images are shown in Figs. 6 and 7. The GLCM features contrast, entropy, and homogeneity threshold values, which helped to classify the different types of dragon stem diseases like a white pail, medium red spot, and dark spot. Out of these diseases, it works well for dark spot and medium red spot, but it fails for some of the white pail diseases images and for some of the backgrounds portions in the image.

Accuracy is found using Eq. 6 and is defined as the total number of diseased dragon images correctly classified as diseased/non-diseased images to that of the total number of images considered for the test. Accuracy for the proposed method is found to be 97.22% for correctly classified images 245 to that of total images 252 (Fig. 8).

$$\text{Accuracy} = \big[\text{total correctly classified images/total number of images}\big] \times 100 \quad (6)$$

Fig. 6 Image after
morphological operations

Fig. 7 MATLAB output
showing unhealthy
classification and also type
of disease is displayed as
medium red spot
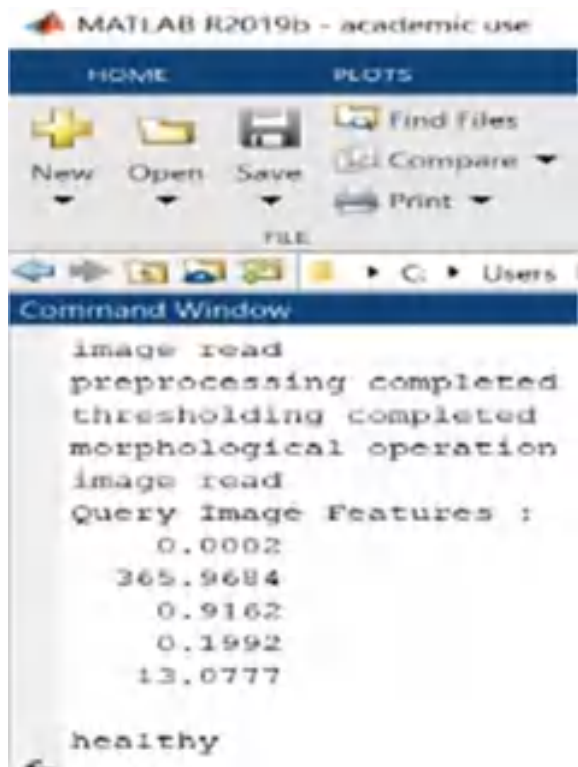
Fig. 8 MATLAB output showing healthy classification
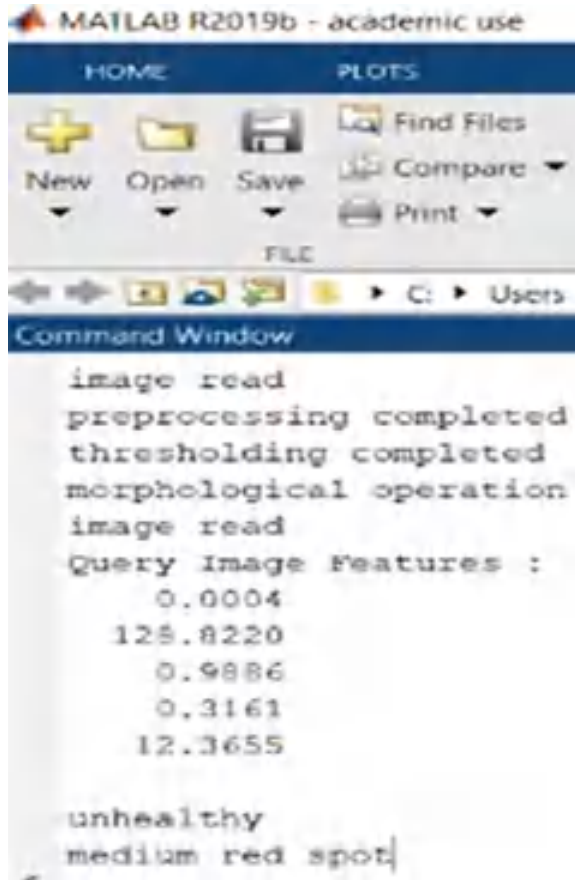
4 Conclusion

In this paper, a color-based and statistical features-based image processing technique is proposed to segment and classify dragon fruit stem diseases. Diseases on the dragon stem leads to a reduction in both the quality and quantity of dragon fruits and loss to the farmer. A solution to this issue is by taking proper care in managing diseases on the dragon stem. This requires expertise in disease symptoms and continuous monitoring of the farm by the farmer, but this is not possible for all farmers because of a lack of awareness of diseases and also expert knowledge may differ from person to person. So, a computer-aided method is required to estimate dragon storm disease correctly. In this process, a computer-aided method is developed which can detect dragon stem diseases. It shows good segmentation results with an accuracy of 97.2% in classifying. This work can be extended as an app so that it is readily available for the use of the framers and also can be extended to an Agro Robot for farming the dragon fruit plants.

References

1. Perween T, Mandal KK, Hasan MA (2018) Dragon fruit: an exotic super future fruit of India. J Pharmacogn Phytochem 7:1022–1026
2. Patwary MA, Rahman M, Barua H, Sarkar S, Alam MS (2013) Study on the growth and development of two dragon fruit (Hylocereus undatus) genotypes. The Agriculturists. 11:52–57. https://doi.org/10.3329/agric.v11i2.17487
3. Karunakaran G, Arivalagan M, Sriram S (2019) Dragon fruit country report from India
4. Masratul Hawa Mohd BS, Zakaria L (2019) First report on bacterial soft rot disease on dragon fruit (Hylocereus spp.) caused by enterobacter cloacae in peninsular Malaysia. Int J Agric Biol 11:659–666
5. Mohd MH, Salleh B, Zakaria L (2013) Identification and molecular characterizations of neoscytalidium dimidiatum causing stem canker of red-fleshed dragon fruit (Hylocereus polyrhizus) in Malaysia. J Phytopathol 161:841–849. https://doi.org/10.1111/jph.12146
6. Arya MS, Anjali K, Unni D (2018) Detection of unhealthy plant leaves using image processing and genetic algorithm with Arduino. In: 2018 International conference on power, signals, control and computation (EPSCICON). IEEE, pp 1–5. https://doi.org/10.1109/EPSCICON.2018.8379584
7. Das A, Dey AK, Sharma M (2018) Leaf disease detection, quantification and classification using digital image processing. Int J Adv Managem Technol Eng Sci 8:423–432
8. Shire A, Jawarkar U, Manmode M (2015) A review paper on: agricultural plant leaf disease detection using image processing. Int J Innov Sci Eng Technol 2:282–285
9. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: 2015 International conference on computing communication control and automation. IEEE, pp 768–771. https://doi.org/10.1109/ICCUBEA.2015.153
10. Patil D (2018) Fruit disease detection using image processing techniques. Int J Res Eng Appl Managem 128–131
11. Sharma R, Singh VP (2017) A research on automatic detection of defects in mango fruit through image processing and machine learning techniques. J Adv Scholarly Res Allied Educ 12:1258–1268
12. Saradhambal G, Dhivya R, Latha S, Rajesh R (2018) Plant disease detection and its solution using image classification. Int J Pure and Appl Mathem 119:879–884
13. Hakim L, Kristanto SP, Yusuf D, Shodiq MN, Setiawan WA (2021) Disease detection of dragon fruit stem based on the combined features of color and texture. INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi 5:161–175. https://doi.org/10.29407/intensif.v5i2.15287
14. Downer J (2022) Pitahaya Diseases. https://ucanr.edu/sites/sdsmallfarms/files/248991.pdf. Last Accessed on 27 Oct 2022
15. Dragon Fruit Pitaya (2022). https://www.specialtyproduce.com/produce/Dragon_Fruit_Pitaya_1439.php. Last Accessed on 27 Oct 2022
16. Raquel J (2022) Dragon fruit stem rot—3 effective treatments. https://fruitinformation.com/dragon-fruit/dragon-fruit-stem-rot/. Last Accessed on 27 Oct 2022
17. Thomas Osborne: Dragon Fruit Diseases (2022) https://tastylandscape.com/2015/09/05/dragon-fruit-diseases/. Last Accessed on 27 Oct 2022

IoMT Assisted Monitoring and Voice-Based Food Recommendation System Using Deep Learning Model

Shridevi Soma and Susmita Dyapur

Abstract Many individuals in the contemporary world are afflicted by a wide range of illnesses and diseases. Providing a diet recommendation on short notice is notoriously challenging. An AI-powered, cloud-based medical automation system has the potential to extend human life, prevent the spread of illness, and improve people's general level of health. Using real-time data from biomedical sensors on 50 patients, this study presents a deep learning approach to construct an IoMT-aided health information system capable of autonomously detecting which food should be supplied to individuals. Provided information is sent to the cloud, and the user will get updates on their health status from the cloud. With the user's current health situation in mind, the system formulates a voice-based dietary suggestion for maximum effectiveness. The suggested technique is improved upon by using deep learning algorithms such as recurrent neural networks (RNNs), multilayer perceptrons (MLPs), and Long Short-Term Memories (LSTMs). It was determined by comparing the precision, recall, accuracy, and F1-measures of a number of different deep learning methods that the LSTM methodology is the most effective. Using an LSTM deep learning model, we were able to get an accuracy of 89.9%. For the permitted class, we get an F1-measure of 0.86 s, recall of 0.87, and accuracy of 0.89.

Index Terms Food recommendation · IoMT · Deep learning · Neural networks · LSTM · and MLP

1 Introduction

Systematically, tracking a patient's health to provide recommendations about what they should eat from a wide range of options is the targeted outcome of the recommendation system in question. A carefully executed recommendation system with the aim of encouraging patients to consume nutritional supplements, food that is

S. Soma · S. Dyapur (✉)
Department of Computer Science, PDA College of Engineering, Kalaburgi, India
e-mail: sushmitadyapur@gmail.com

thought to be more suited to patient's health requirements and preferences. Chronic illnesses, diabetes, vitamin deficiencies, and poor food quality are widespread. This project's goal is to develop a system for recommending foods that will have a positive impact on an individual's health. Many people with ailments might benefit from food recommenders as a kind of preventive medication. To raise the patient's standard of living, foods rich in vitamins, minerals, antioxidants, protein, and fat promote improved health and are essential for optimal physical performance. Patients whose sick bodies can't handle a healthy diet may also be at risk.

A voice-based recommendation helps people who are physically challenged, and those who do not understand the system parameters, they can make use of the voice recommendation to know their body condition and appropriate suggestion as quick as possible. In addition to developing a patient-facing information system, the authors of this research want to realize and incorporate a comprehensive body of nutritional theory in IoT infrastructure. In this work, we use a deep learning model (LSTM, MLP, etc.) to solve the problem.

1.1 Internet of Medical Things (IoMT)

Connected devices and systems that share data through the internet are collectively referred to as the "Internet of Things" (IoT). There are already roughly 10 billion linked IoT devices, and that number is expected to expand to nearly 25 billion by 2025, after being initially mentioned by Ashton in 1999. On a technical level, this means maximizing the efficiency with which information is sent between linked devices and stored in a safe cloud server, from which users may access their information from any number of their own computers. Smart gadgets built on the Internet of Medical Things (IoMT) are having an increasingly noticeable influence all over the world, especially in the current pandemic stage. Health care is expected to be one of the most difficult domains for IoMT because of the sheer scale of problem.

Connected by electrical or electronic means, wired or wirelessly, most IoMT systems include a wide variety of components. As considered to patient diseases or illness, the main reasons are unhealthy diets, food patterns, and not aware of their body conditions, and these conditions may lead to further diseases or bigger problems; in order overcome these issues, our proposed system come up with a solution which integrate the IoMT and deep learning model in one system to recommend the correct food, to monitor patients body condition and voice-based recommendation.

1.2 Deep Learning

Machine learning techniques based on artificial neural networks and representation learning include deep learning. One may engage in supervised, semi-supervised, or

unsupervised learning. Computer vision, speech recognition, NLP, machine translation, bioinformatics, drug design, medical image analysis, climate science, material inspection, and board game programs are just some of the areas where deep-learning architectures like neural networks, belief networks, reinforcement learning, recurrent neural networks, and convolutional neural networks have been applied and shown to produce results on par with human experts. At each successive stage of deep learning, the system learns to abstract and composite its input data in a little different way. One possible representational layer for an image recognition program would abstraction the pixel and encode edges; another would compose and encode configurations of edges; a third would code an eyes and nose; as well as a fourth would detect that the picture comprises a face. An essential aspect of deep learning is that it can figure out for itself how to best organize features across different tiers.

2 Related Work

Literature survey is base for any challenging research work. A thorough study of existing system results to ide Literature survey includes many existing systems considering IoMT assisted health monitoring, diet recommenders and food recommendation system. Some of the researchers consider the user preferences, and some of them only focused on athletes or physically challenged people. Researchers have used machine learning algorithms for the implementation, and some of them have used DNN model which includes various algorithm, and they have carried out the performance evaluation Following are the recommended papers to understand the existing work in the design of intelligent IoMT assisted food recommendation system.

A survey includes some of the existing systems related to IoMT-based health monitoring system [1] which uses sensored medical data and product features to recommend a good diet for patient by implementing deep learning methods like LSTM, MLP, GRU achieved a good accuracy rate for LSTM and GRU with 94% of accuracy. IOT-based health monitoring system [5] to reduce the coronary illness of a patient using machine learning techniques like SVM with 78% of accuracy, random forest classifier has an accuracy of 83%, naïve bayes has an accuracy of 83%. A deep learning-based health monitoring system [7] use big data analytics with recommender system, the study includes Restricted boltzman machine RBM and CNN focusing on people choices and their preferences on social media platform. Researchers have developed a health monitoring system [2] for athletes using wearable sensors and deep learning methods like Gradient Descent algorithm, as per the study system raises some critical issues including computation cost and complexity. An intelligent food recommendation system [6] using product reviews and rankings gives a good product recommendation for the user. A voice-based health monitoring system [4] to monitor patient health and consider heart rate as a primary parameter. System alerts user using Thing talkcloud if any irregularity in the heart rate. Some of the researchers designed a system to help elderly/blind patient's health through wearable sensors, and system will remind patients caretakers to follow the medications suggested by

the doctors using telegram bot, to store the voice message have used APR module to instruct the blind patient.

Reviewing various existing papers regarding food recommendation and health monitoring, proposed contents of the paper can be helpful for people by giving the solution in real time. A proposed system includes IoMT and deep learning methods to produce the result in the form of voice.

3   Proposed System Design and Methodology

The proposed system is introduced to provide the features like voice-based recommendation, continuous monitoring of patient health condition, and appropriate food suggestions which are not present in the existing system. The proposed model consists of three modules: One is IoMT which takes the input data, and it will be stored in database, and another module includes deep learning classifiers, and the other module is the output module which includes food recommendation, monitoring, and voice-based recommendation.

Figure 1 depicts the architectural design of a recommendation system. A system takes input data from biomedical sensors like heart rate and spo2 sensor, temperature sensor, blood pressure sensor, and body hydration level sensor. The collected data will be send to the Blynk cloud platform, in the next step deep learning model uses classifier's like LSTM and MLP it does performance analysis. In the next step, food recommendation is given in the form of voice. Implementation of a system includes both hardware and software. A hardware consists of IOT sensors, LED display, and a nodemcu microcontroller. A software part consists of Python programming for implementing various deep learning algorithms.
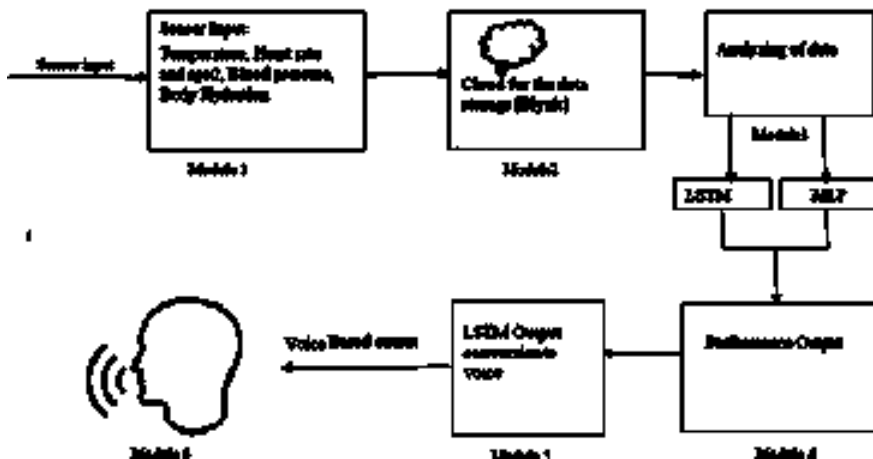

Fig. 1 Architectural design for food recommendation system

This study presents a system design for tracking vital signs such a patient's heart rate, fluid levels, temperature, and blood pressure. The system's progress lies in its ability to save settings in the cloud and to assist in communicating important information to the patient's loved ones. For this to work, we must link the sensor to the microcontroller, which will then serve as our data generator. At this point, the microcontroller takes the labeled data it has received from sensors node and uploads it to the cloud/database, where it may be utilized for further analysis.

3.1 Working of LSTM Network

To address the issue of long-term dependencies in RNNs, LSTM networks were developed. In contrast to standard feedforward neural networks, LSTMs may learn from their past mistakes thanks to their feedback connections. This feature allows LSTMs to handle full sequences of data without separating out individual points and instead using knowledge of prior data to guide processing of subsequent data.

A "cell state," or memory cell, plays a vital function in an LSTM model by retaining its state across time. One metaphor for this is a conveyor belt on which data is passed unaltered. In LSTM, the cell state may have information added to or withdrawn from it according to the rules of gates. Information may enter and leave the cell selectively via these gates. The technique is supported by pointwise multiplication operation and sigmoid neural network layer. Inherently, LSTMs are geared for capturing long-term dependencies. The memory units in an LSTM are trained to learn what information is most important, in while also learning general weights to prioritize that information. How the input affects a next internal state is set by the input gate, and how the previous internal state affects the future internal state is set by the forget gate. How the network's internal state affects the output is set by the output gate. In certain cases, it's beneficial to store away relevant information for later use (Fig. 2).

Instead of the hidden units seen in regular recurrent networks, this one uses cells as its recurrent building blocks. A standard artificial neuron unit is used to calculate a characteristic of the input. As long as the sigmoidal input gate permits it, its value may be added to the state. The forget gate regulates the strength of the linear self-loop in the state unit. The cell's output is controlled by an output gate, which may be closed to prevent further signals from leaving the cell. Input unit may have any nonlinearity that squashes, whereas gating units all have a sigmoid nonlinearity. Gating units may take the state unit as an additional input.
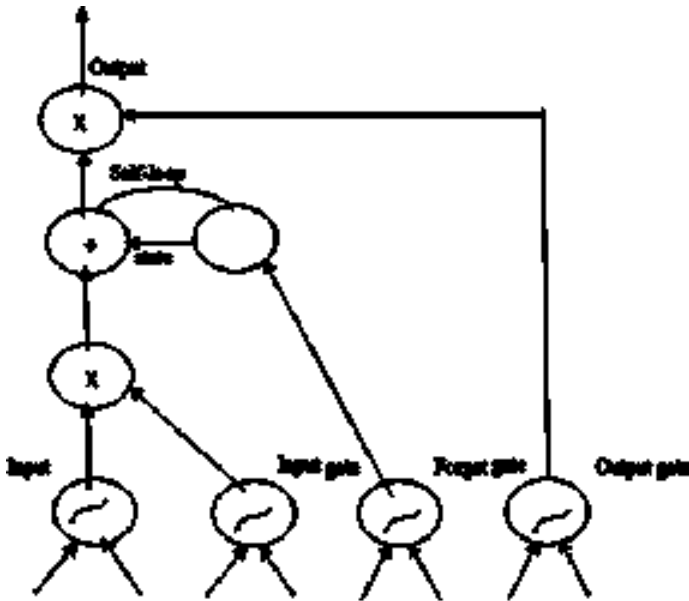
Fig. 2 Block diagram of the LSTM recurrent network "cell"

3.2 Working of Multilayer Perceptron (MLP)

MLP learning process is as

- The data should be sent forward from the input layer to the output layer. The forward propagation stage has now arrived.
- Determine the error based on the output. An effort should be made to reduce the size of the mistake.
- The mistake should be sent backward. The model should be updated by finding its derivative in regard to each of network's weights.

The input layers of a multilayer perceptron (MLP) are represented by neurons that represent the available data, here the multispectral picture band values; the hidden layer illustrates the training process; and the output layer is bathymetric information. Figure 3 depicts a hypothetical multilayer perceptron (MLP) ANN with four input layers, five hidden layers, and a single output layer (4-5-1).

MLPs are a kind of neural network model that may be used as a universal approximator, providing an approximation for any continuous function. Specifically, MLPs are made up of neurons that we'll refer to as perceptions. Thus, the general structure of a perceptron would be discussed before the general structure of MLPs. Each of the n characteristics sent into a perceptron ($x = x_1, x_2, ..., x_n$) is given a weight, as illustrated in Fig. 4. Numeric characteristics are required for input. As a result, a
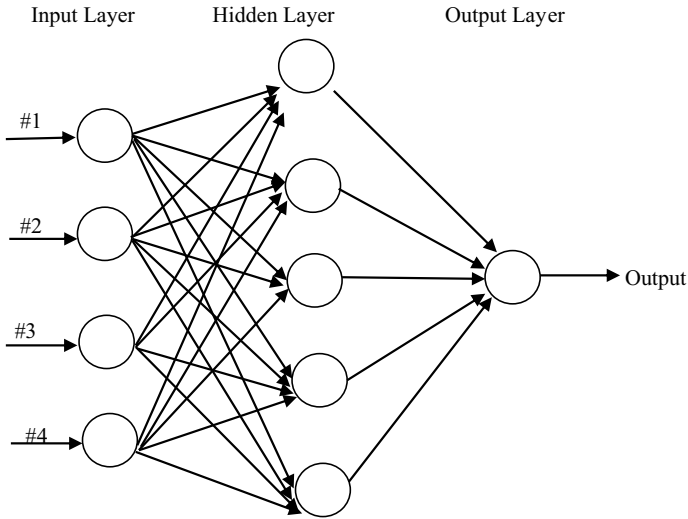
Fig. 3 Multilayer perceptron network

perceptron requires a transformation from non-numerical input properties to numerical ones. A categorical feature having p potential values may be transformed in p input features that indicate presence/absence of such values.

An input function u receives the input features and returns weighted sum of features:

$$u(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i. \tag{1}$$

The perceptron's output is then generated by applying an activation function f to outcome of this computation. The activation function in the classic perceptron is a step function:
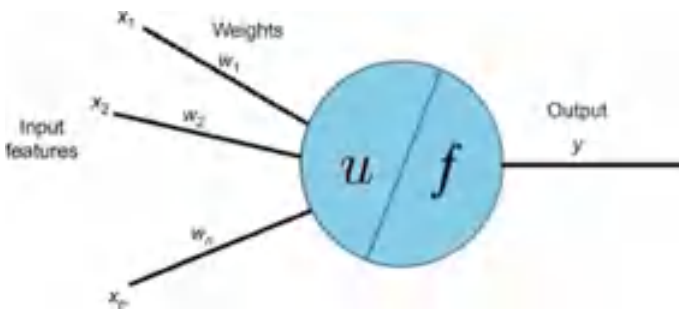
Fig. 4 Scheme of perceptron with *n* input features

$$y = f(u(\text{x})) = \begin{cases} 1, & \text{if } u(\text{x}) > \theta \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where θ is a threshold parameter. An example of step function with $\theta = 0$. Thus, we can see that the perceptron determines whether $w_1 x_1 + w_2 x_2 + \cdots + w_n x_n - \theta > 0$ is true or false. The equation $w_1 \times w_2 \times + \cdots + w_n x\text{n} - \theta = 0$ is the equation of a hyperplane. When an input point is above the hyperplane, the perceptron returns 1, and when the input lies on or below hyperplane, the perceptron returns 0. This is why the perceptron is sometimes referred to as a linear classifier; specifically, it is best used with linearly separable data. For a perceptron to learn, one must first tweak the weights until a hyperplane is found that effectively classifies the training data.

4 Experimental Results

Experimental results of a designed system includes a hardware setup of a sensors connected to the nodemcu microcontroller which is connected to Wi-Fi. The microcontroller uses an Arduino Uno to display the readings on an output window. Initially, all the readings of sensor set up to zero; once a user or a patient checks his body conditions by placing his or her hand on a particular sensor, the LED display will show the current status of patient health and shows the real-time data on the Arduino output window. In the next step, collected data from sensors are connected to cloud to send notification to the user.

4.1 Dataset

A patient dataset consists of 50 patient details like user id and food preferences like veg or non veg, nutrient, disease, and kind of diet which he used to prefer. The details are mentioned in Table 1.

A Food dataset consists details like food product name, nutrient, category, description of a food product. The details are mentioned in Table 2.

Table 1 Patient dataset

1	User_id	Veg_Non	Nutrient	Disease	Diet
2	User_id1	Nonveg	Chloride	Anemia	High_ protein
3	User_id2	Veg	Chloride	Goiter, hypertension	High_ fiber
4	User_id3	Veg	Magnesium	Cancer, hypertension	High_ fiber
5	User_id4	Veg	Magnesium	Anemia	High_ Protein

Table 2 Food dataset

Meal id	Name	Category	Nutrient	Veg-non	Disease	Diet
Mealid_1	Squash salad	Salad	Fiber	Veg	Diabetes, hypertension	Alkaline diet
Mealid_2	Chicken salad	Salad	Fiber	Nonveg	Anemia	Low fat diet
Mealid_3	Tricolor salad	Salad	Fiber	Veg	Hypertension	Alkaline diet
Mealid_4	Sweet chilli almonds	Chilli	Vitamin a	Veg	Obesity, hypertension	Alkaline diet
Mealid_5	Gluten free cake	Cake	Vitamin a	Veg	Goiter, kidney disease	High protein diets

Table 3 Precision, recall, and F1 scores classification report

Classifier name	Precision	Recall	F1-measure
LSTM	0.89	0.87	0.86
MLP	0.78	0.67	0.65

4.2 Evaluation Metrics

The effectiveness of our suggested approach is measured in many ways. There are some of them listed below. Tp + tn Fp + tn Tn The purpose of precision is to rank True Positives (TP) against False Positives (FP). Extreme Accuracy = Tp Tp + Fp Recall is used to compare correctly classified items (True Positives, TPs) with those that were incorrectly labeled, or False Negatives, FNs. There is a reference to the mathematical form of memory recall in; Memory = TP + Fn If a mining method has poor recall but great precision, for example, a different algorithm is required for a proper performance evaluation. There is also the issue of which algorithm performs best. The F1-measure, which provides a mean recall and accuracy, is the key to solving this issue. The formula for determining F1 is as follows. Scale = 2* Accuracy *Recall Accuracy + Memory (Table 3).

Considering recall, precision, and F1-measure results mentioned in Fig. 5 and Table 1 result proves that LSTM technique performs better than the MLP with respect to accuracy, recall, precision, and F1-measure.

5 Conclusions

An automated IoMT assisted system could grow endurance, protection against sickness in the future, also enhance living standards. A system evaluates performance of the numerous deep learning classifiers and provides the best solution for the problem. It also monitors patient health conditions continuously to give the better
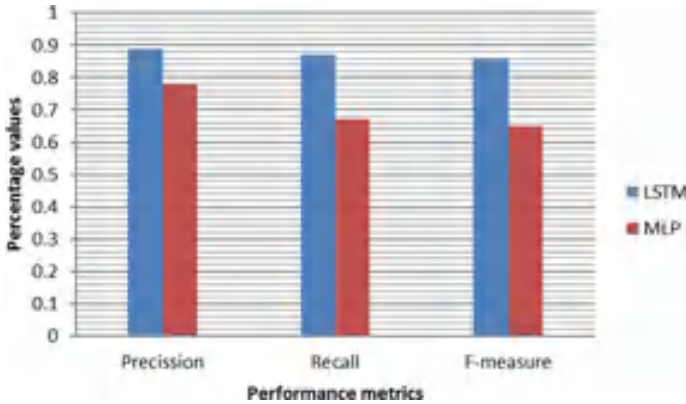
Fig. 5 Comparison of performance of LSTM and MLP

health. A voice-based recommendation system will fulfill the people health needs in a better way. Proposed system comes up with a solution which integrates the IoMT and deep learning model in one system to recommend the correct food, to monitor patient's body condition and voice-based recommendation. In order to fill the gap in the existing system, the research incorporates the new features in to the system to provide features like voice-based recommendation, continuous monitoring of patient health condition, and also medicine suggestions of Ayurveda which are currently not present in the existing system. It was determined by comparing the precision, recall, accuracy, and F1-measures of a number of different deep learning methods that the LSTM methodology is the most effective. Using an LSTM deep learning model, we were able to get an accuracy of 89.9%. Similarly, for the permitted class, we get an F1-measure of 0.86, recall of 0.87, and accuracy of 0.89.

References

1. Iwendi C, Henry Anajemba J (2020) Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. In: IEEE Special section on deep learning algorithms for internet of medical things. vol 8
2. Wu XD, Lijun C, Muhammad Bilal W (2021) Internet of things-enabled real-time health monitoring system using deep learning. 15 Sept 2021, Springer
3. Telkar S, Taj N, Khan F (2020) Monitoring of elderly/blind patients health remotely through wearable sensors. Int J Eng Res Comput Sci Eng (IJERCSE) 7(9)
4. Pandey H, Prabha S (2020) Smart health monitoring system using IOT and machine learning techniques. IEEE Xplore
5. Hari Kishore K, Surendra Nath KV, Harikrishna Kvn, Pavan Kumar D (2019) IOT based smart health monitoring alert device. Int J Innov Technol Explor Eng (IJITEE) 8(6S). ISSN: 2278-3075
6. Sahoo A, Pradhan C (2019) DeepReco: deep learning based health recommender system using collaborative filtering. Computation
7. Ashok Naik P (2020) Intelligent food recommendation system using machine learning. 5(8)

8. Li W, Chai Y, Khan F (2021) A comprehensive survey on machine learning- based big data analytics for IoT-enabled smart healthcare system
9. Kale A, Auti N (2015) Automated menu planning algorithm for children: Food recommendation by dietary management system using ID3 for Indian food database. Proc Comput Sci 50:197–202
10. Alian S, Li J, Pandey V (2018) A personalized recommendation system to support diabetes self- management for American Indians. IEEE Access 6:73041–73051
11. Yang CC, Jiang L (2018) 'Enriching user experience in online health communities through thread recommendations and heterogeneous information network mining. IEEE Trans Comput Soc Syst 5(4):1049–1060
12. Yera Toledo R, Alzahrani AA, Martinez L (2019) A food recommender system considering nutritional information and user preferences. IEEE Access 7:96695–96711
13. Agapito G, Calabrese B, Guzzi PH, Cannataro M, Simeoni M, Care I, Lamprinoudi T, Fuiano G, Pujia A (2016) DIETOS: a recommender system for adaptive diet monitoring and personalized food suggestion. In: Proceedings IEEE 12th international conference wireless mobile computing, network communication (WiMob), New York, NY, USA, Oct 2016
14. Mu R (2018) A survey of recommender systems based on deep learning. IEEE Access 6:69009–69022
15. Iwendi C, Zhang Z, Du X (2018) ACO based key management routing mechanism for WSN security and data collection. In: Proceedings IEEE international conference industrial technology (ICIT), Lyon, France, Feb 2018

Framing of Quality Questions for Quality Code Snippets

Divya Kumari Tankala and **T. Venu Gopal**

Abstract Nowadays, developers or programmers depend on the reusability of code to implement software as it reduces time and effort. Therefore, most developers depend on code bases or on Question and Answers (Q&A) sites to reuse the existing code. They look for exact code snippets by simply giving questions to Q&A websites. But the problem is when programmers give code snippets with low-quality question titles leads to retrieving irrelevant answers. Hence, our approach helps users (who are providing code snippets to Q&A sites) in the forming of high-quality questions by automatically generating question titles for a code fragment using the token-based approach with LSTM deep neural network model with 96% accuracy which is a better than the existing model. Our model evaluated three programming languages (e.g., Java, Python, C#) datasets collected from GitHub.

Keywords Question generation · Quality question · Code snippet · LSTM · Deep learning

1 Introduction

Our major focus in this paper is to frame correlated and quality questions for code snippets as developers are highly dependent on the reusability of code, and there is a wide range of platforms providing code snippets for related questions. Always knowledge-gathering communities or forums help developers in the implementation of software, maintenance of software, fixing bugs in software, etc. So, it became a habit for programmers to look for the code in search engines. As soon as the developer gives code as an input query, it suggests various web application links as

D. K. Tankala (✉)
Department of CSE, G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India
e-mail: divyatankala@gnits.ac.in

T. Venu Gopal
JNTUH UCE, Department of CSE, Jagityal, Telangana, India

503

a result. Sometimes, he gives natural language queries to request a code snippet, but he left with poor results as the question title for the code snippet is inappropriate. On this website or forum, the developer posts question to request code or to solve the error in the software program, and the community of Internet users can share code snippet as an answer to the question. Users are allowed to vote on questions and answers to assess the quality. The potential benefits of such Q&A websites are an immediate solution to the problem faced by programmers or developers. Of course, this need can be satisfied by various websites such as Stack Overflow, Quora, Yahoo answers, GitHub, etc. But the genuine problem faced by such platforms is the lack of quality questions for a code snippet. Formulating the question for an appropriate code snippet is highly needed to attract more users. Most of our work is based on existing work with few modifications [13]. Figure 1 shows example code pair which have a poor correlation between code snippets and questions from the Stack Overflow web application. Therefore, improving the quality of questions for a code snippet helps in a better way to the user. Of course, previous studies provided some sort of solution for the acquisition of code snippets by analyzing question titles [7] using natural language processing techniques, few studies [9] provided enormous methods to address this issue by adopting the voting system, the ranking also provided for answer by many users, and text similarity. Another study advised keeping bit code in question for better retrieval of code snippets [8]. The prior research work on Question and Answers communities was successful in many aspects to enhance the question quality such as by understanding the user behavior while editing questions for a code snippet [1] to reduce the gap between requester and answer. Researchers worked based on metrics to strengthen the question title of the code snippet to boost the requester. Another study [2] investigated classifying the questions into 'very good,' 'good,' 'bad,' and 'poor.' In this paper [3], they aim to provide intuitional observation into the quality of a question by inspecting the code fragments in a question. Together with the code-to-text ratio and readability metrics, metrics for these constructs were used in a classifying algorithm.

As per my knowledge still, there is challenge in generating question titles for code snippets, which may not understand the behavior of logic to acquire the semantic nature of code snippets. Our major contribution is to frame a question with respect to code fragments so that it effectively relates to the exact code snippet. Section 2 presents related techniques used in existing research work on framing question titles for code snippets, Sect. 3 describes the approach of our work then evaluation and results, and later discussed the conclusion and future enhancement of our work.

2 Related Work

In the field of software engineering, due to the huge need for existing codes for developers, there is rapid growth in web service tools to serve programmers. Deep learning plays an important role in all the research fields to automate processes.

Fig. 1 Sample code snippet and question pair from stack overflow [17]
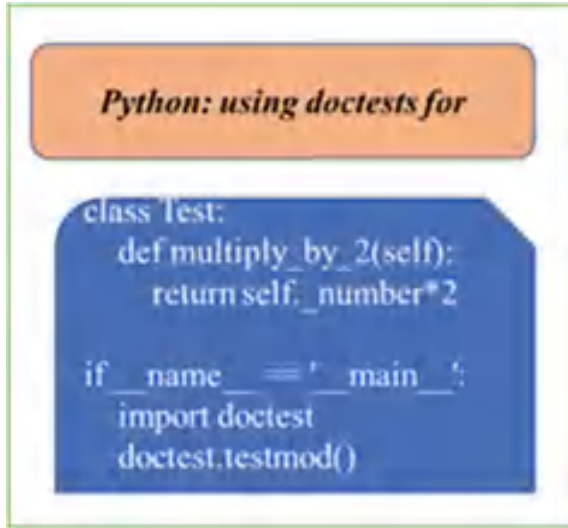
In this section, we give a survey of work related to our approach and then various techniques used in our work.

2.1 Deep Learning in Code Analysis

Understanding code fragments results in effective performance. Numerous techniques proposed by various authors to analyze code [3, 4, 14, 15]. Neural networks are used to convert code into a vector with the help of intermediate transformations such as Abstract Syntax Tree or graphs and check for similarity. Machine learning techniques also gave acceptable analysis in code summarization. Here [4], the main aim is to build a deep learning system that bridges the language gap by directly quantifying query-questions semantic similarity across languages using word embeddings [12], CNN, and sampling SVM. This research work [5] has given an elaborative machine learning framework and built a predictive model to identify a 'closed' question at the time of question creation. The other work [6] is to identify the code clones and code-related bugs code, which helps to improve user confidence in the reliability of code.

2.2 Question Generation for the Code Snippet

While searching for code snippets, programmers may give natural queries or a combination of natural language and code queries and expect an exact code snippet as a

Table 1 Dataset statistics

Language	#Cod2e tokens	#Question tokens	Avg. code length	Avg. question length
Python	2,367,148	109,329	84.7	11.2
Java	3,371,946	123,994	103.2	10.8
C#	2,340,202	100,178	82.1	11.0

solution. To address this issue, the existing work suggested various techniques either generating a quality question for code snippets [2, 3, 5, 9] or mapping the terms of the query to code snippets using NLTK [16]. In technical Q&A sites, how to write a question for a code snippet to retrieve better results for information seekers with the help of observational investigation [10, 13] advised to avoid unnecessary terms and keep code snippets in question.

3 Approach

In this section, experimental work is described. Most of our work is like the existing model [13]. Our main aim is to provide a novel approach to assist developers in posting trait questions for the appropriate code snippet. Such question titles can be generated for a given code snippet automatically with the help of our web service tool. We first describe the dataset then the methodology to develop a model to retrieve a quality question for a given code fragment.

3.1 Datasets

Our dataset consists of code snippets and their corresponding question titles. As we know most software developers get stuck at some or the other point while writing code (Table 1).

For assistance related to technical errors, developers usually approach the Internet and most specifically a Q&A site to get their issues resolved. Three datasets were collected from GitHub which contains Python, Java, C# codes, and questions. The proposed model was evaluated on these datasets. Figures 2 and 3 are showing the sample screenshots of Python codes and the titles of respective codes in the dataset.

3.2 Methodology

Our approach can generate informative questions by analyzing code, which is automatically processed with code tokens and question tokens. Hence, the programmer
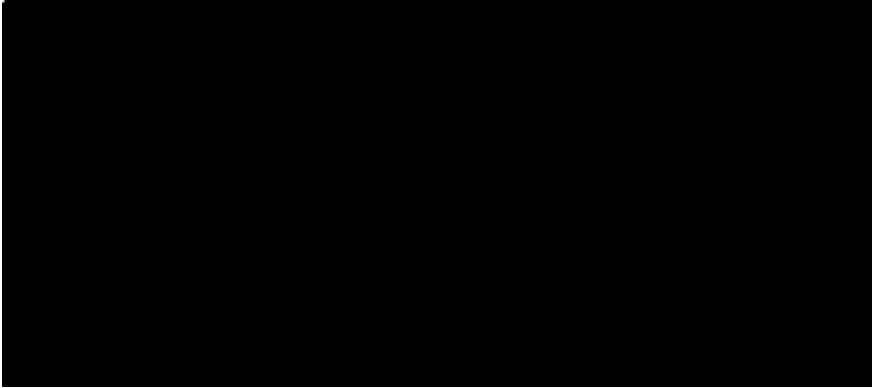
Fig. 2 Sample screen to represent code snippets of the dataset

Fig. 3 Sample screen to represent questions of the dataset

can put less effort to edit questions. As shown in Fig. 4, a Question Decoder and Source code encoder are built with the Long Short-Term Neural Network (LSTM) as it depends on the previous hidden state and Seq2Seq [11] technique used to mine code snippets and formulates the question from them. The Source code encoder has a code embedding network with bi-directional LSTM, which helps in the transformation of code into vector. The Question Decoder abides with the LSTM network, while decoding the question it copies a few terms of code snippets in the question. LSTM is a text-mining recurrent neural network that encodes source data (Code) and target data (Questions) and then trained a model. This trained model can accept any
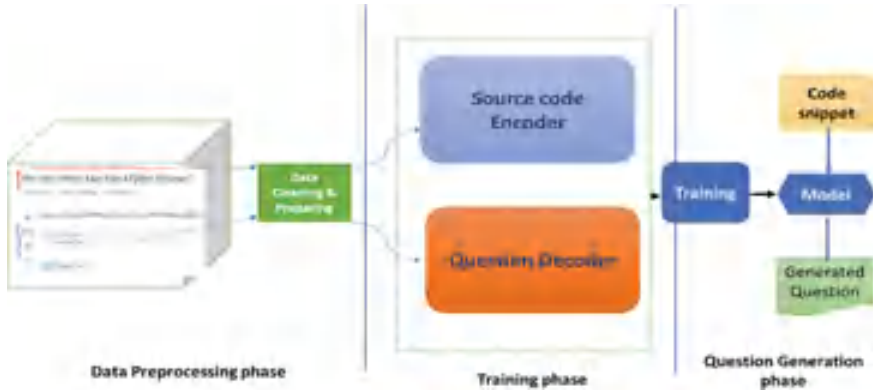
Fig. 4 Proposed framework to frame a question for the code snippet

coding data and then the decoder will predict questions from that code. LSTM will encode both source and target data based on sequence 2 sequence mapping [11].

Preprocessing: Data preprocessing is a very crucial step to be performed in the implementation of any neural model as the model cannot understand raw data such that it applies necessary steps to convert it into a vector representation. Algorithm 1 shows the step-by-step procedure to understand the preprocessing of the dataset. To achieve this phase, tokenization of code and question titles are performed to make sure the lengths of code snippets and question title padding are applied.

Training: In Fig. 5, given the code snippet and the respective question as the training dataset for the model. The training phase consists of two components. *Source Code Encoder* converts the source code into sequential tokens by iterating through the entire code snippet. The token sequences were modified by adding START and END tokens at the beginning and end of the training sequence. It has a code embedding network with bi-directional LSTM (forward pass f, backward pass b) as shown in Eqs. (1) and (2).

$$\text{(forward pass) } f_t = \text{LSTM}\,[C_t,\ h_t - 1] \tag{1}$$

$$\text{(backward pass) } b_t = \text{LSTM}\,[C_t,\ h_t - 1] \tag{2}$$

At time stamp t, the concatenated form at the last layer is represented as $S_t =$ [f_t, b_t]. It takes sequential tokens of source code using the Seq2Seq model to get a vector representation called context vector. The words used in the code snippet were analyzed to get the context of the code snippet. In a similar way, the *Question Decoder* decodes the question by transforming it into tokens. These tokens are provided as inputs for training the model. After training the model, the user can provide the code snippet as input. The model will be generating the question as per the context of the code snippet, important words can be copied to the question title.
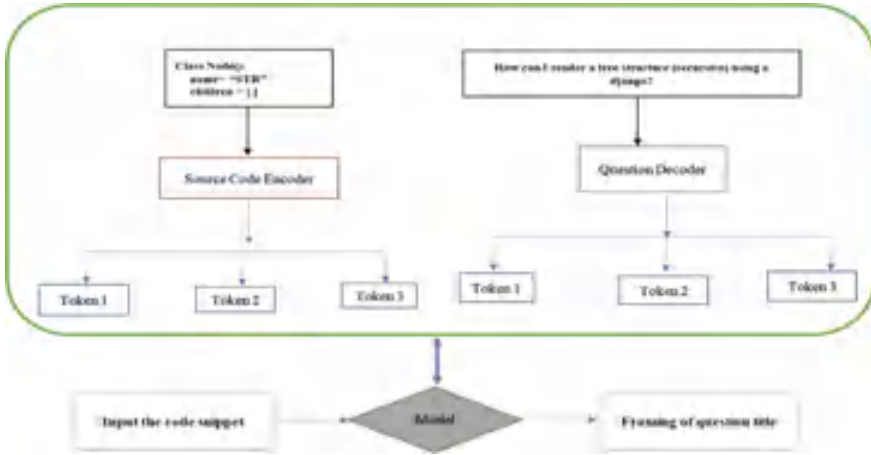
Fig. 5 Workflow of our model

Table 2 LSTM pre-padding versus LSTM post-padding (%)

	LSTM-4 pre-padding	LSTM-4 post-padding
Train	80.072	49.977
Test	80.321	50.117
Epochs	9	6

The hyperparameters of the proposed model chosen are the softmax activation function for hidden layers, the model compiled with sparse_categorical_crossentropy loss function, Adam optimizer. Source code tokens and question tokens Algorithm 1 give the steps to preprocess the dataset and take word vectors, or distributed representations of words. Algorithm 2 provides the elaborated steps to train and evaluate the model. The input sequences are padded to the maximum length during testing and training as shown in Table 2.

Our model experimented with basic parameters, if we increase the number of epochs and changing of hyperparameters may lead to better performance. In the decoding process, it reads the code embeddings of the source code encoder to frame a question title. Instead of having only raw text as a question for a code snippet, some of the technical terms can be copied into the question while generating a title for it. Thus, the programmer can search for exact results for the intended code. Hence, deep insight has been taken into analyzing the code and question titles of the dataset. Mostly, the mechanism of our approach is data-driven.

Algorithm 1: Preprocessing the dataset.

Input: Dataset with code snippets and question titles

Output: Code tokens and question tokens
Function preprocess (dataset)
Step 1: Declare global variables like max_code_len, code_vocab, max_ question_len, question_vocab,code_pad_sentence, question_pad_sentence, question_text_tokenizer code, question, code_text_tokenizer, code_token, question_token, code_length, question_length
Step 2. text.delete("1.0,END) to clear the text element
Step 3. Input the code snippets to the code tokenizer
Step 4. Input the questions to the text tokenizer
Step 5. Calculate the max length code_text_tokenizer and question_text_ tokenizer and store their length
Step 6. Do the padding of both code_text_tokenizer and question_text_ tokenizer to make sure all lists the sequence have the same length.
Step 7. Outputs code tokens and question tokens.

3.3 Evaluation of Model

The LSTM and CNN take sequential inputs of equal length. Hence, all the inputs should be padded to make the lengths of the inputs equal. Here, we consider a common task for both CNN and LSTM and analyze the effect of padding on them, the task being Sentiment Analysis. There are two types of padding, namely pre- and post-padding. Since LSTMs and CNNs take inputs of the same length and dimension, input images and sequences are padded to maximum length while testing and training. This padding can affect the way functioning of neural networks and can make a great deal when it comes to performance and accuracy.

Algorithm 2: Evaluation of the model.

Input: Code tokens and Question tokens
Output: Generates question
Step 1: Convert the tokens into vectors using embedding ()
Step 2: Apply the LSTM() and encode the embedding
Step 3: RepeatVector() is used to repeat the input max_question_len times to generate 3D tensor of space
Step 4: Apply the LSTM() to decode the Repeat_Vector
Step 5: logits = TimeDistributed() is used to slice the decoded input
Step 6: Create the model using initial input, softmax activation function, and logits

Step 7: compile the model using sparse_categorical_crossentropy loss function, Adam optimizer and accuracy metrics

Then summarise the model, Train the model (here epochs we is 500). write the model in JSON file.

Convert the model into a pickle file, Find the accuracy of the model

The metrics used to evaluate the model are accuracy, Mean Squared Error (MSE), Mean Absolute Error (MAE), and mean absolute percentage error (MAPE). Table 3 is showing the values while evaluating the model. The LSTM model gives 96% accuracy, and the risk function is used to reduce the loss of the model.

Accuracy: It is a metric used to evaluate the classification models, basically used to get the predictions right with the fraction. Formally, the formula is given in Eq. (3)

$$Accuracy = (TP + TN)/(TP + FP + TN + FN), \tag{3}$$

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

Mean Squared Error: It is a metric to calculate the loss function corresponding to the expected value of the squared loss. The formula is mentioned in Eq. (4)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2, \tag{4}$$

where MSE = Mean Squared Error, n = number of data points, Y_i = Observed values, $\hat{Y}_i$ = Predicted values.

Mean Absolute Error: It refers to the metric, which is the magnitude of difference between the prediction values and the true values. The formula mentioned in Eq. (5)

$$MAE = \frac{\sum_{i=1}^{n} |Y_i - X_i|}{n}, \tag{5}$$

where MAE = Mean Absolute Error, Y_i = Prediction, X_i = True value, n = total number of data points.

Table 3 Values obtained in the evaluation of model

Metric	Value (%)
Accuracy	96%
MSE	10
MAE	12
MAPE	11.6

Mean Absolute Percentage Error: It measures the average magnitude of error produced by a model.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{AI} \right| * 100,$$

where MAPE = Mean absolute percentage error, n = Sample size, A_i = actual value, F_i = forecast value.

The loss of the model can be measured with the metrics mentioned above; accuracy of the model increases proportionally to the epochs or iterations. Of course, the loss decreases with iteration count.

Figure 6 is showing the graph between loss/accuracy and epochs/iterations. Tkinter is the package of Python used to develop interfaces to make use of models. Here, Fig. 7 shows a sample screen to retrieve the question title for a given code snippet. Hence, quality questions can boost the user's productivity.
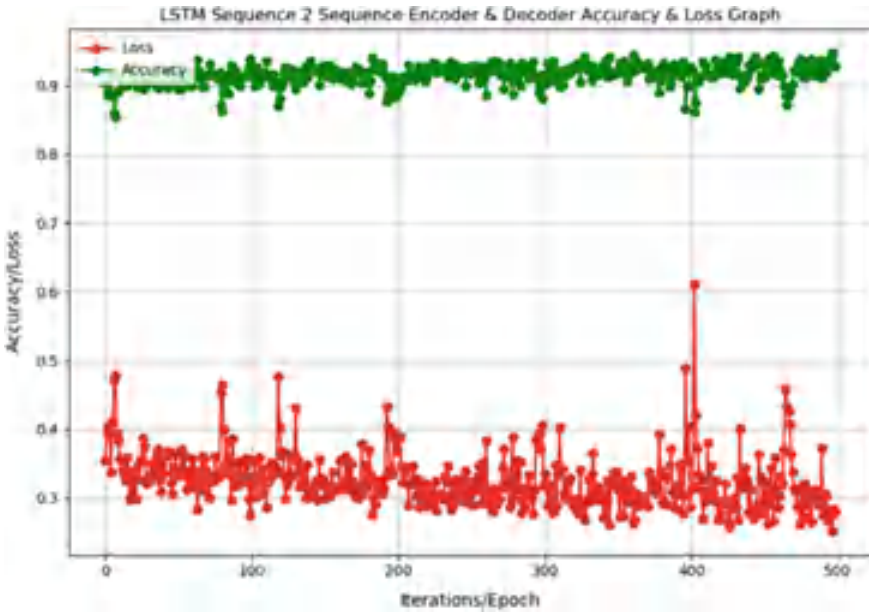

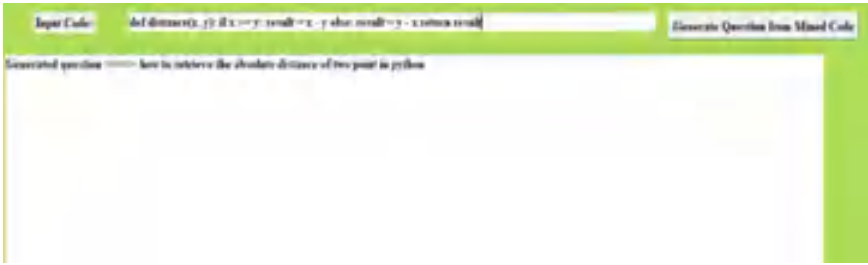**Fig. 6** LSTM accuracy/loss versus epoch/iteration graph

Fig. 7 Sample screen to retrieve questions for given code

4 Conclusion

There are so many Q&A websites, among them, Stack Overflow is very popular. The low-quality questions may show unrelated code snippets as answers. Uploading incomplete, improper questions cause a hindrance in the learning process because such questions cannot be answered by the developers who can answer the questions or can be used by the people who can clear their doubts with these questions. Therefore, addressing this problem LSTM neural networks gave a better solution with 96% accuracy, which is better than the model [5], and evaluated the model with various metrics. Still, the model can be trained to work with multiple programming languages, whereas the current model can perform well for only three languages. It can be extended to different languages and investigated with other metrics.

References

1. Yang J, Hauff C, Bozzon A, Houben G-J (2014) Asking the right question in collaborative Q&A systems. In: Proceedings of hypertext 2014, pp 179–189
2. Ponzanelli L, Mocci A, Bacchelli A, Lanza M (2014)Understanding and classifying the quality of technical forum questions. In: Proceedings of QSIC 2014, pp 343–352
3. Duijn M, Kucera A, Bacchelli A (2015) Quality questions need quality code: classifying code fragments on stack overflow. In: 2015 IEEE/ACM 12th working conference on mining software repositories, 2015, pp 410–413. https://doi.org/10.1109/MSR.2015.51
4. Chen G, Chen C, Xing Z, Xu B (2016) Learning a dual-language vector space for domain-specific cross-lingual question retrieval. In: 2016 31st IEEE/ACM international conference on automated software engineering (ASE), 2016, pp 744–755
5. Correa D, Sureka A (2013) Fit or unfit: analysis and prediction of 'closed questions' on stack overflow. In: Proceedings of the first ACM conference on Online social networks (COSN '13). Association for Computing Machinery, New York, NY, USA, pp 201–212. https://doi.org/10.1145/2512938.2512954
6. Zhipeng G, Vinoj J, Lingxiao J, Xin X, David L, John G (2019) SmartEmbed: a tool for clone and bug detection in smart contracts through structural code embedding
7. Tapan H, Aryak S, Anirban G (2015) Analysis of titles from the questions of the stack overflow community using natural language processing (NLP) TECHNIQUES. 17:2278–661. https://doi.org/10.9790/0661-17411728

8. Squire M, Funkhouser C (2014) A bit of code: how the stack overflow community creates quality postings. In: 2014 47th Hawaii international conference on system sciences, pp 1425–1434. https://doi.org/10.1109/HICSS.2014.185

9. Wang X, Huang C, Yao L, Benatallah B, Dong M (2018) A survey on expert recommendation in community question answering. J Comput Sci Technol 33(4):625–653

10. Calefato F, Lanubile F, Novielli N (2018) How to ask for technical help? Evidence-based guidelines for writing questions on stack overflow. Inf Softw Technol 94(C):186–207

11. Chen Z, Kommrusch S, Tufano M, Pouchet L-N, Poshyvanyk D, Martin M (2021) SequenceR: sequence-to-sequence learning for end-to-end program repair. IEEE Trans Software Eng 47:1943–1959

12. Gao Z, Jiang L, Xia X, Lo D, Grundy J (2021) Checking smart contracts with structural code embedding. IEEE Trans Softw Eng 47(12):2874–2891. https://doi.org/10.1109/TSE.2020.297 1482

13. Gao Z, Xia X, Grundy J, Lo D, Li Y-F (2020) Generating question titles for stack overflow from mined code snippets. ACM Trans Softw Eng Methodol 29, 4:37. https://doi.org/10.1145/ 3401026

14. Wan Y, Zhao Z, Yang M, Xu G, Ying H, Wu J, Yu PS (2018) Improving automatic source code summarization via deep reinforcement learning. In: Proceedings of the 33rd ACM/IEEE international conference on automated software engineering (ASE 2018). Association for Computing Machinery, New York, NY, USA, pp 397–407. https://doi.org/10.1145/3238147. 3238206

15. Zhao G, Huang J (2018) DeepSim: deep learning code functional similarity. In: Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering (ESEC/FSE 2018). Association for Computing Machinery, New York, NY, USA, pp 141–151. https://doi.org/10.1145/3236024.3236068

16. Bird S (2004) NLTK: The Natural Language Toolkit. ArXiv, cs.CL/0205028

17. Anon (2008) Stack overflow. Available at: https://stackoverflow.com/

Domestic IoT Smart Home Its Future Roles and Human Behavior Through Interaction

Suresh Kallam, C. Sreedhar, Nayan Rai, Moturi Sirisha, Sadu Venkatesu, and P. Srinivasa Rao

Abstract Although objects can be abstract in their meanings, they will be referred to in their physical sense, i.e., physical objects or material objects. Also, the objects described here are inanimate. An object is a "thing consisting of matter." Objects have physical representations that enable human beings to experience them through the senses, to see them, touch them, taste them, smell them, and hear them. Interpretations of objects are often limited to their views as equipments or instruments to fulfill tasks like work or survival. However, objects are more than that. This paper shall present some important roles of objects. Physical inactivity has increased significantly over the past years, and the advancement of technology has contributed to it. Paradoxically, domestic IoT shapes human behavior through human interaction. As everyday objects become a part of the Internet of Things (IoT), this paper aims to investigate how the IoT devices and everyday objects can collaborate with humans to address growing physical inactivity. Using a speculative and critical design approach, design proposals in the form of physical and video prototypes are constructed and discussed in a series of workshops. Participation in the workshops moves the participants from being passive consumers of technology to citizens that actively debate and design their own future. The outcomes of the paper are themes that critically address the implications of domesticating technology and its future roles and functions. Also,

S. Kallam (✉)
School of Computer Science and Engineering, Jain (Deemed-to-be University), Bengaluru, Karnataka, India
e-mail: sureshkallam@gmail.com

C. Sreedhar
Department of Computer Science and Engineering, G. Pulla Reddy Engineering College (Autonomous), Kurnool 518007, AP, India

N. Rai · M. Sirisha
Department of CSD, GNITS, Hyderabad 500104, Telangana, India
e-mail: nayan@gnits.ac.in

S. Venkatesu
IVY Software Development Services Pvt Ltd, Hyderabad, Telangana, India

P. Srinivasa Rao
Department of ISE, Jain (Deemed-to-be University), Bengaluru, Karnataka, India

a set of characteristics is outlined to illustrate desirable, undesirable, and preferred characteristics of networked technologies that may encourage physical activity.

Keywords Domestic Internet of Things · Smart objects · Speculative design · Physical activity · Motivation · Domesticating technology

1 Introduction

Technological evolution has brought extreme convenience and inevitable drawbacks to society. Innovations such as smart devices have enabled humans to access each other and things regardless of distance and have even automated tasks that would be fulfilled by humans through physical efforts and means. Through networked technologies in the Internet of Things (IoT), motion sensors cause lights to turn on and doors to open by detecting movement without requiring excessive use of the body. This advancement although beneficial has negatively impacted physical activity and caused a rise in the sedentary lifestyle of humans. Paradoxically, the trajectory of Internet of Things research shows the potential and possibility of everyday objects being augmented with technological abilities and contemporary technology collaborating to positively influence sedentary behavior and human lifestyle in general.

Electronic objects and everyday objects can influence human practices and their lifestyle. Traditional design practice of solving the problems of users with technology tends to overlook the influence of these objects.

In a network where humans exist with other non-human actors like these electronic objects and everyday objects, all actors must equally be examined. This paper aims to enhance the understanding of the limitations and possibilities of designing a network where these actors collaborate to encourage physical activity within the household.

2 Literature Review

Objects mediate social relationships [1]. For instance, gifts are used to express or convey emotions and intentions. Humans exchange gifts to express satisfaction with a deed or simply as a mark of appreciating a person or their role in one's life. By giving or receiving a gift, social relationships can be formed, defined, or even strengthened [2]. Giving an object or receiving one may equally signify a loss or an end of a relationship. Also, through this symbolic act, an object given may become an extension of the giver and trigger memories or stoke certain feelings when the receiver sees the object or thinks of it [3].

The objects one owns may signify their emotions [4]. In an extreme example, a person may use weapons to demonstrate anger or to stoke fear. A person may use an object out of fear of an attack. Objects may also affect one's emotions by providing

comfort or mental relaxation. They may even be acquired as a consolation for loss or a reward for an achievement, or they may simply be acquired for the pleasure of the human [5]. People may acquire certain objects they believe possess a certain power or they may attribute this power or capabilities to certain.

(A)   Problem Statement

Although objects can be abstract in their meanings, they will be referred to in their physical sense, i.e., physical objects or material objects. Also, the objects described here are inanimate. An object is a "thing consisting of matter." Objects have physical representations that enable human beings to experience them through the senses, to see them, touch them, taste them, smell them, and hear them. Interpretations of objects are often limited to their views as equipments or instruments to fulfill tasks like work or survival. However, objects are more than that. This shall present some important roles of objects.

(B)   Existing Systems

From influencing emotions to mediating the relationships between humans, enabling the formation of identity, and conditioning behavior to motivate humans, the significance of objects is apparent. With that said, it is worth noting that this study does not suggest that the potential of inanimate objects is a product of their existence or intrinsic materiality, but rather something that emerges out of their relationship with humans. Also, this study does not suggest that objects are equal to humans in the sense that they can sense or that they experience sensations, but "they do have an existence and agency" that exists outside their relationship with humans which can be built upon through IoT designs [6]. This study sympathizes with this view and asserts that the design of IoT devices may benefit from understanding the use and influence of everyday objects.

3   Energy Model for IoT

This paper follows research through design method (RtD) [8]. The authors define design research as a contribution of knowledge and propose four criteria for assessing the quality of knowledge contribution [9]. The criteria are process, invention, relevance, and extensibility. Figure 1 shows the Process–Sufficient details on the methods and justification for the chosen methods must be provided.

Invention—Details on the designs produced must be communicated clearly. Relevance—How the contribution can be applied to the real world and justification for its importance [7].

Extensibility—The extent to which the contribution can be used by others.

Fig. 1 Architecture of the
proposed model
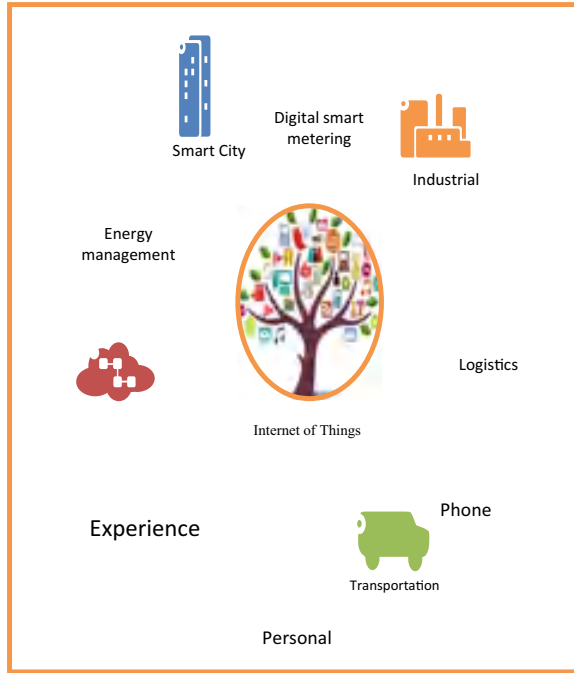
4   Trimming Model for Security Attacks

Utopianism describes an imaginary society in which everything is perfect. Dystopia
is the opposite of utopia and describes a bad and imperfect society [10] recommend
not striving to build utopias or dystopias but to use them to weigh the current state
of things and explore better and preferable states [11] also suggests using utopia and
dystopian scenarios to create discussions on not only future states people do want
but also those they do not want [13]. Figure 2 shows the engaging the participants in
discussions about utopian and dystopian concepts provided insights for the study on
ways humans may prefer smart domestic technology to motivate them to do physical
activity.

5   Experimemtal Analysis

So far objects in the study, objects home have been explored with their conventional
forms and function. However, more abstract forms of the objects were investigated
in the next iteration. The primary motivation was to broaden the scope of the design
exploration and explore potential designs for new objects. Inspiration was also taken

Fig. 2 Architecture of the trimming mechanism for denial of service attacks
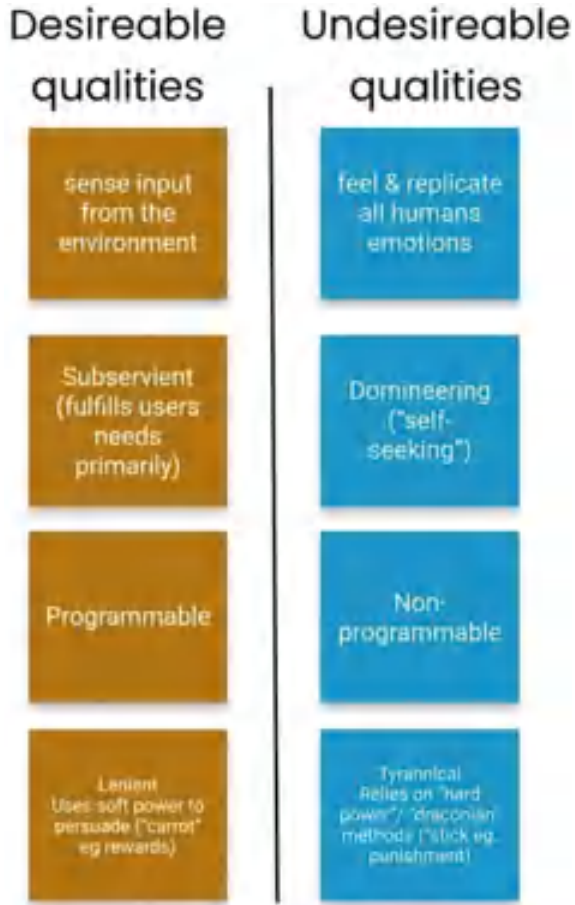
from Technological Dreams Series, [12] write that they aimed to avoid "fuss about functionality" and thus abstracted the forms of objects to reimaging their functions.

Network Lifetime:

One of the first questions asked was "what if the user gets injured on the way?" and must return home without completing the workout? The participants discussed and agreed that under this circumstance, the mood dial would be beneficial. A seemingly difficult question was how the technologies could detect deception in activities like running without fitness trackers like Fitbit and Apple Watch. There was speculation on how the machines can calculate the distance over time between landmarks like bus stops and how check-in systems could be placed between these landmarks. "Another person can do the exercise for you and scan the QR-code," was another remark about detecting deception. Embedding fingerprint sensors on the equipment and or

Fig. 3 Energy consumption
of devices
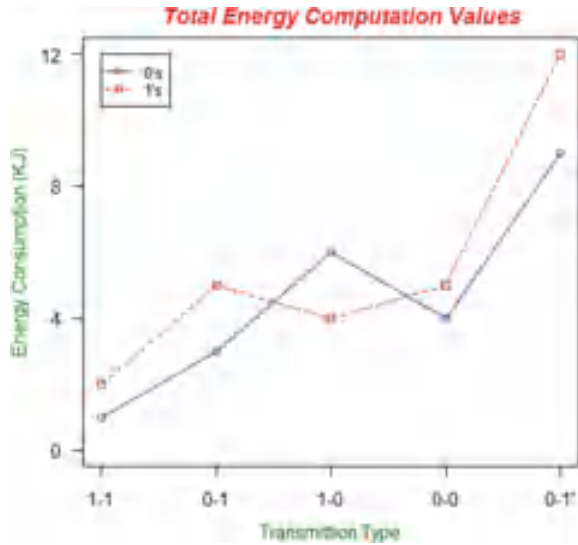
a facial recognition system to indicate that the user is doing the activity were some
suggestions (Fig. 3).

There are several reasons why a person might be deceitful, and the fear of judgment
or punishment is some of these reasons [13]. Sometimes people can and want to, and
nothing can be done for that specific reason.

Studies [14] conducted have shown some strength in the use of computer vision
to detect deception through facial and eye movement. This could be explored in the
future iterations of the concept. In this context, however, deception might indicate a
lack of trust. It may arise due to the fear of judgment, shame, or punishment. In that
case, the mood dial might be a way to overcome this. With the mood dial, the user
may be honest without fear of judgment.

Another point of discussion was if the machine had to always decide the kind
of activity the user must do? The discussion on this point indicated that it must be
flexible enough to give the user multiple varieties of workouts to do. Also, it needs
to enable the user to input other activities they may have done.

6 Conclusion

The initial aim of this investigation was an exploration of designing things to
encourage physical activity in the household from a human perspective. Research
on artificial intelligence and the IoT in the study demonstrates the possibility of
augmenting everyday objects as IoT technologies. However, this opportunity also
has some challenges, particularly how meaningful interaction would be supported
between humans and electronic objects in the given context. Therefore, an extra

dimension was added to examine important qualities from a perspective of Things that networked technologies need to support meaningful interactions with humans.

References

1. ACE Fitness, American Cancer Society, & American Society of Clinical Oncology (n.d.) Exercise versus physical activity. penn state college of medicine. Retrieved May 6, 2022, from https://research.med.psu.edu/oncology-nutrition-exercise/patient-guides/exercise-vs-activity/
2. American Psychological Association (n.d.) APA dictionary of psychology. American Psychological Association. Retrieved February 15, 2022, from https://dictionary.apa.org/motivation
3. Dunne A, Raby F (2007) Dunne & Raby. Dunne & Raby. http://dunneandraby.co.uk/content/bydandr/13/0
4. Auger J (2014) Living with robots: a speculative design approach. J Human-Robot Interact 3(1):20. https://doi.org/10.5898/jhri.3.1.auger
5. Auger J (2012) Why robot? Speculative design, the domestication of technology and the considered future. Royal College of Art, London
6. Auger J (2013) Speculative design: crafting the speculation. Digital Creativity 24(1):11–35. https://doi.org/10.1080/14626268.2013.767276
7. Raffaella B, Tom M (2003) Dark horizons. Psychology Press
8. Boubekri M, Cheung IN, Reid KJ, Wang CH, Zee PC (2014) Impact of windows and daylight exposure on overall health and sleep quality of office workers: a case-control pilot study. J Clin Sleep Med 10(6):603–611
9. Brookhouse O (2022) Can artificial intelligence understand emotions?—think big. Telefónica February 7. https://business.blogthinkbig.com/can-artificial-intelligence-understand-emotions
10. Cherry K (2021) What is extrinsic motivation? Verywellmind. https://www.verywellmind.com/what-is-extrinsic-motivation-2795164/
11. Cila N, Smit I, Giaccardi E, Kröse B (2017) Products as agents: Metaphors for designing the products of the IoT age. In: Conference on human factors in computing systems—proceedings, May, pp 448–459
12. Citrogene (n.d.) Wearable micro component health sensors. Retrieved March 1, 2022, from https://www.citrogene.com/wearable-micro-component-health-sensors/
13. Coulton P, Lindley JG (2019) More-than human centred design: considering other things. Design J 22(4):463–481. https://doi.org/10.1080/14606925.2019.1614320
14. Csikszentmihalyi M, Halton E (1981) People and things. The meaning of things: domestic symbols and the self. Cambridge University Press, Cambridge, pp 1–19

Comparative Analysis of Crude Oil Price Prediction Using Various Machine Learning Models

Shubh Gupta and Aayush Jadhav

Abstract Crude oil is arguably the most important resource on the planet right now. Fluctuations in its prices affect every commodity due to the direct effect on transportation. Getting ahead of the uncertainty around crude oil prices can prove to be a game changer for businesses. It is extremely challenging to predict the price of crude oil because of its high volatility and its dependence on several external factors. People have been trying to make models using different machine learning algorithms and appropriate datasets to make the best price predictions. This paper is a survey paper which does a comparative analysis between three models available for crude oil price prediction, namely SVM, ANN, and GARCH-GED. The vital information from the WTI unrefined petroleum market dataset is used in all these models and they are evaluated on the basis of the RMSE value obtained.

Keywords Crude oil · Artificial neural network (ANN) · Generalized autoregressive conditional heteroscedasticity (GARCH) · Support vector machine (SVM) · Root mean square error (RMSE)

1 Introduction

Natural petroleum products like crude oil are made up of deposits of hydrocarbons and other organic elements. Crude oil, a form of fossil fuel, is refined to produce meaningful products like gasoline, diesel, and various other petrochemical products. It is a limited resource since it is non-renewable, which indicates that it cannot be regenerated easily and naturally at the rate humans consume it. Crude oil is a vital commodity with a significant influence on the global economy. Since crude oil and natural gas account for about two-thirds of the world's energy consumption, they play a significant role in the international economy. Crude oil production and effective usage are critical to the global economy's operation. The reasons why the

S. Gupta (✉) · A. Jadhav
Thadomal Shahani Engineering College, Mumbai, India
e-mail: shubhgupta3@gmail.com

price of crude oil fluctuates are several. Natural disasters, geopolitical tensions, and numerous other unforeseen occurrences could impact the price fluctuations of crude oil. Forecasting the price of crude oil can be advantageous to refiners, hedge funds, and consumers because they are frequently exposed to crack spread.

Crude oil price volatility has a significant impact on the macroeconomy, social security, and prosperity of entire nations [1, 2]. In [3] author contends that oil shocks have been the single most influential factor in causing recessions. For these reasons it has become vital to have a reliable method to forecast the price of crude oil. The importance of prediction in crude oil prices has led to different prediction mechanisms. While the belief network has been used in [4, 5] has a semiparametric methodology that uses the GARCH properties of crude oil prices to predict the price distribution over a short-term horizon. In [6] vector autoregression (VAR) and ANN are explored to forecast oil prices in the US. Reference [7] uses error correction models (ECM) to predict crude oil prices to check crude oil and product price dynamics. More recently, [8] used a model which is based on ensemble empirical mode decomposition (EEMD) and long short-term memory to try and predict the price of crude oil (LSTM).

This paper assesses three approaches for forecasting crude oil prices: support vector machine (SVM) [9], artificial neural network (ANN) [10], and generalized auto regressive conditional heteroskedasticity (GARCH) [11] and compares them with the RMSE value. Section 2 examines the available resources about the referred models. In Sect. 3, we discuss the three models in detail. Section 4 analyzes the models and compares, showcasing the results. Section 5 concludes the paper.

2   Literature Survey

Support vector machine has been a frequent and reliable method to predict that has been used in various fields like wind speed prediction for renewable energy production [12], for prediction of the levels of air pollutants based on the monitored air pollutant database in an advancing time series [13]. The method has implemented various times for forecasting the crude oil prices, [14, 15] proposed a new method for crude oil price forecasting based on a support vector machine (SVM).

In the current environment, when technology is taking over our lives and efforts are being made to reduce the need for human labour, the artificial neural network technique has emerged to be among the most useful techniques for forecasting the behaviour of nonlinear systems such as economic and financial systems. For example, in [16] using backpropagation neural networks, the option prices of Nikkei 225 index futures are predicted. Similarly, these [17–19], employ neural network architecture to model and/or predict the changes in the dynamic exchange rate. Reference [10] proposes a modern and novel approach for predicting crude oil prices using an artificial neural network.

The classes of generalized autoregressive conditional heteroscedasticity models are better able to characterize the mean or variance of financial time series, (GARCH)

family models have been extensively used in recent forecasting research. It makes more sense for us to create short-term forecasts using a univariate GARCH-type model. Asymmetric and long-memory GARCH classes of models may ultimately outperform other forecasting models as the perturbation term of the self-correlation coefficient decays with time with a hyperbolic rate. Reference [20] investigates the efficacy of a volatility model using CGARCH, FIGARCH, and IGARCH models. Then [21] concludes that linear GARCH-class models can accurately predict long-memory and asymmetric volatility better than linear ones, especially when predicting over longer time horizons. Reference [22] forecasts the market volatility using both univariate and multivariate GARCH-class models. Reference [11] implements GARCH-N, GARCH-t, and GARCH-G to forecast crude oil spot prices.

3   Proposed Methodology

3.1   GARCH

The autoregressive conditional heteroscedastic (ARCH) model was first developed by Engle, then Bollerslev generalized it to create generalized autoregressive conditional heteroscedastic model (GARCH). It was created to solve the problem of predicting volatility in prices of assets that play an important role. The terms "conditional" and "autoregressive" refer to the degree of connection on the previous sequence of observations and the feedback process that integrates past data into this [23].

The key characteristics of asset returns, such as the fat-tailed nature of the return distribution and volatility clustering, which are categorized as different facts, are well captured by the GARCH model. However empirical research had frequently found that the GARCH model falls short in explaining the leptokurtosis in return distributions. It was advised to use a different distribution for the qt error that had a fatter tail than the standard distribution to get around this. For the terror term, a number of distributional models with fatter tails than the normal distribution is used, including the student's t-distribution and generalized error distribution (GED) [24].

The GARCH (p, q) model, a more inclusive variation of the ARCH model, was first proposed by Bollerslev. The formula for the GARCH (p, q) model is

$$y_t = \sigma_t \, \epsilon_t \tag{1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{2}$$

where αi, βi and ω are GARCH model parameters. In order for the model to match the output data, these parameters were chosen. The GARCH parameters are limited to $\omega > 0$, $\alpha i > 0$ and $\beta\_i > 0$ to assure a positive volatility because the volatility σ_t^2

should be positive. ε_t is a separate normal error after N (0, 1), and yt provided the return time series. The GARCH (1, 1) model, where p and q are equal to 1, and the volatility process is provided by

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3}$$

and from this point forward, the GARCH model just refers to the GARCH (1, 1) model. Additionally, we refer to the GARCH model that has normal errors as the GARCH-N model. It is challenging to describe the original time series' leptokurtic (high kurtosis and heavy-tailed) features using GARCH (p, q) models using a normal error distribution, despite the literature's extensive use of non-normal error distributions. Hansen utilized the skewed t-distribution for calculating the time series' skewness and excess kurtosis. Reference [25] assessed the GARCH models' volatility prediction using the student-t and SGED distributions.

The papers used for this survey uses the WTI dataset from 1st January 1986 to 30th September 2006. The data contained 5237 data points of the price of WTI prices for the days mentioned above. The dataset was later divided into in-sample (training period) from 20th May 1987 to 31st December 2002 and out-of-sample (testing period) from 1st January 2003 to 30th September 2006. In [11] the author has implemented GARCH-GED where they conducted multiple tests like Jarque and Bera test, augmented Dick Fuller test and Phillips Perron test to find the impact of volatility to the model. The model was able to achieve an RMSE value of 0.7667.

3.2   SVM

Support vector machine (SVM) is a supervised machine learning algorithm used for regression and classification. SVM has the benefit of minimizing the issues with overfitting or local minima, in contrast to neural networks and other supervised learning methods to which SVM is typically compared [26]. This is because learning in neural networks is based on the empirical risk minimization principle, whereas learning in SVM is based on the structural risk minimization principle. SVMs are based on a linear machine in a high-dimensional feature space, nonlinearly related to the input space, which has allowed the development of somewhat fast training techniques, even with many input variables and big training sets [27].

The dataset that has been used for the SVM model compared, includes 5237 observations from the weekly and monthly datapoints from West Texas intermediate records because of the variety of missing points in their daily data. This data was then normalized and transformed into the range [0, 1]. After this, the dataset was split into training and testing sets. For training, there are no concrete guidelines for determining the parameters other than a trial-and-error way. The training's outcomes help in deciding the architecture. RBF Kernel has been used because it often offers decent performance under standard softness assumptions. The formula that was used to normalize the data is mentioned below.

$$X_n = (X_i - X_{\min})/(X_{\max} - X_{\min}) \tag{4}$$

3.3 ANN

Computing systems based on artificial neural networks, connectionist systems, often known as artificial neural networks, are conceptually related to neural networks found in the human body but differ from them in certain important ways. An ANN completes tasks without programming or task-specific rules by using examples [28]. A neural network's job is to create or develop an output pattern from an input pattern. A parallel-distributed artificial neural network (ANN) has a high number of nodes (neurons) and connections [29]. In order to achieve the best performance, we use the backpropagation learning technique, which involves changing and manipulating the network's weights to propagate the error signal through the network, backward. Up until the network can produce the desired outcomes, the process is iterated.

The ANN model development steps for our prediction are shown in Fig. 1.

Since the closing price of crude oil, the only dependent variable in the proposed model, is a time series, we conducted the experiments in accordance with the model for general time series prediction and presented the findings as follows [28]:

$$Y_t = f(X') \tag{5}$$

where X' is a vector of lagged variables $\{xt - 1, xt - 2, \ldots, xt - p\}$. The input variables depicted in Fig. 2 are the lagged variables. To closely approximate the function is the estimation problem. To accomplish this, the modelling methodology weights can be iteratively adjusted. Figure 2 shows a diagram of the proposed model.

As an example, in this case, the four stages that involved the creation of this proposed predictive model were as follows:

(a) Data collecting: They obtained the data on the price of crude oil from investing.com. The collected data spans a time period of 20 years, 9 months. About 5200 data were retrieved, and the closing price of crude oil was compromised.

(b) Data normalization: Prior to the commencement of the training process, data is normalized. Using the following calculation, the normalization range of the closing price is [0.001, 0.005].

$$P^{\wedge'} = (p - \text{Min})/(\text{Max} - \text{Min})(l - m) + m \tag{6}$$

where the normalized value has been used; p stands for the value to be normalized; the minimal value of the series that requires normalization is called Min; the series' maximum value, Max, is what will be normalized; the range's minimum value is m; l is the range's highest value [30].

(c) Activation function: The activation function also sometimes referred to as the transfer function, establishes the connection between a network's input and output

Fig. 1 Structure of artificial
neural network
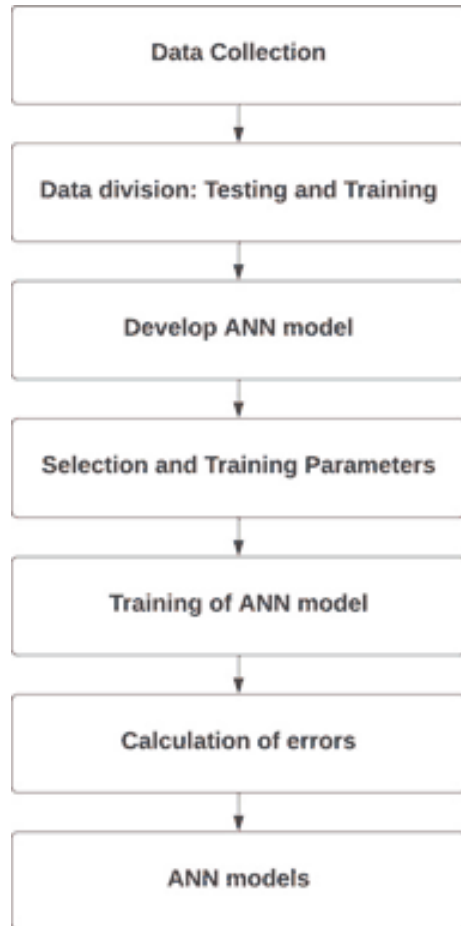
nodes. For this task, the sigmoid function shown below was utilized.

$$f(x) = 1/\left(1 + e^{\wedge}(-x)\right) \tag{7}$$

(d) Training algorithm: A neural network requires extensive training. The arc weights of the neutral network are modified iteratively to minimize the total mean or overall squared error between the desired and the actual output values for the output nodes' overall input patterns It is an unconstrained nonlinear problem of minimization (change in output is not proportional to the change in input). The conventional BP technique, which employs the steepest descent gradient approach to train the model and eliminates mistakes, defines the error function E.
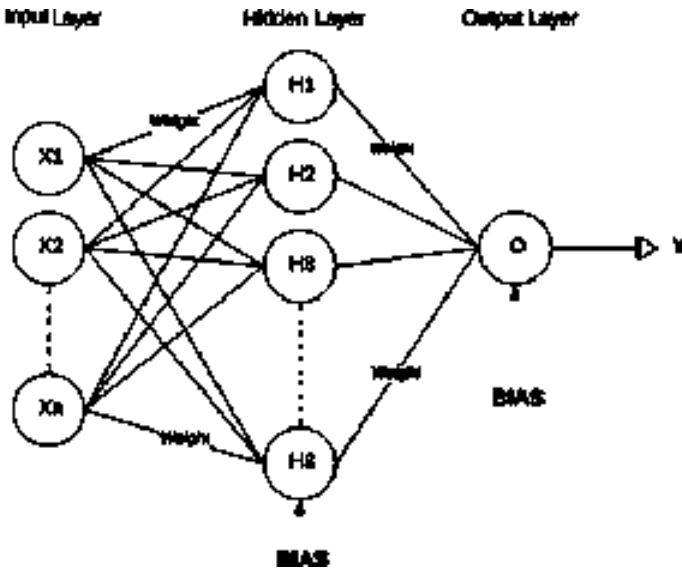
Fig. 2 Model development diagram

$$E = \frac{1}{2N} \sum_{p=1}^{n} (y_p - y_p^d)^2 \tag{8}$$

where Yp is considered to be the network's output and y_p^d is the desired output for the pTH input pattern. The modified rules for the weights and biases of this model are produced using the following equations: here, the steepest descent gradient technique is used, and the partial derivative is calculated using the chain rule.

$$w_i^{\text{new}} = w_i^{\text{old}} + \Delta w_i \tag{9}$$

$$b_i^{\text{new}} = b_i^{\text{old}} + \Delta b_i \tag{10}$$

$$\text{Where } \Delta w_i = -\eta \frac{dE}{dw_i} \tag{11}$$

$$= -\eta \frac{1}{n} \sum_{p=1}^{n} \left((y_p - y_p^d) y_p (1 - y_p) \frac{u}{w_i x_i + b_i} x_i \right) \tag{12}$$

$$\Delta b_i = -\eta \frac{dE}{dbi} \tag{13}$$

$$= -\eta \frac{1}{n} \sum_{p=1}^{n} \left((y_p - y_p^d) y_p (1 - y_p) \frac{u}{w_i x_i + b_i} x_i \right) \qquad (14)$$

where η is used as the learning parameter that governs the rate of convergence.

Database used for the model in comparison was the daily West Texas Intermediate (WTI) decomposed into subseries by the Mallat algorithm. Input combinations evaluated by them are (i) y_{t-1}, (ii) y_{t-1}, y_{t-2}, (iii) y_{t-1}, y_{t-2}, y_{t-3}, (iv) y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, (v) y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}, and (vi) y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}, y_{t-6}. All data points in the input neuron were normalized and were transformed within $[-1, 1]$. Activation function for the hidden and output layer was done by employing hyperbolic tangent function. The learning rate had initially been set to 0.001 and the momentum coefficient was originally set to 0.9. Until the alteration above yields a decreased performance number, the learning rate was increased by a factor of 10 and the decrease step is of 0.1. The model had trained for 1000 epochs.

3.4 RMSE

It is required to offer certain forecasting assessment criteria in order to assess the performance of the predictions. Root mean square error (RMSE) and direction statistics (Dstat) are introduced as the study's two primary assessment criteria. The RMSE is determined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \qquad (15)$$

where N is the number of testing data sets, $\hat{y}_t$ are the projected values, and y_t is the actual value. Evidently, the RMSE indication indicates how much the estimations are off from the actual numbers. From the perspective of practical applications, a change in trend in oil price forecasting is more significant than the goodness-of-fit accuracy level. Trading that is based on a forecast that has a tiny forecast error could not be as beneficial as trading that is based on a precise prediction of the movement's direction. Hence, we use this as the performance measure to compare the different machine learning models.

4 Result and Analysis

a. Root mean square error (RMSE), support vector machine (SVM), generalized auto regressive conditional heteroscedasticity (GARCH), artificial neural network (ANN).

Table 1 Evaluation of forecasting results for WTI crude oil prices

Methodology	RMSE value
ANN	0.9452
SVM	0.8684
GARCH-GED	0.7667

The final evaluation is given in Table 1. The generalized auto regressive conditional heteroscedasticity—generalized error distribution model outperforms the artificial neural network and support vector machine models, as given in Table 1. In terms of RMSE, the GARCH-GED model had a forecast error of 0.77, whereas the support vector machine had a forecast error of 0.87. The artificial neural network outperformed the other models, which had a forecast error of 0.9452.

5 Conclusion

This study examines the three methods—SVM, GARCH-GED, and ANN models—that are used to forecast the price of crude oil. The well-known RMSE method of error evaluation was used to compare the models. The thoroughly examined findings demonstrated that the GARCH-GED approach outperformed the others. As a result, it is advised that future crude oil price forecasts be made using the GARCH-GED approach.

References

1. Sadorsky P (1999) Oil price shocks and stock market activity. Energy Econ 21(5), no. 1:449–469
2. G a Wu Y-JZ (2014) Does China factor matter? An econometric analysis of international crude oil prices. Energy Policy 72:78–87
3. Evans M (2000) Don't oil prices matter any more? In: Industry Week, vol 249, Endeavor Business Media
4. Abramson B, Anthony F (1991) Using belief networks to forecast oil prices. Int J Forecast 7(3):299–315
5. Morana C (2001) A semiparametric approach to short-term oil price forecasting. Energy Econ 23(3):325–338
6. Mirmirani S, Hsi CL (2004) A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. In: Applications of artificial intelligence in finance and economics, Emerald Group Publishing Limited
7. Lanza A, Matteo M, Massimo G (2005) Modeling and forecasting cointegrated relationships among heavy oil and product prices. Energy Econ 27(6):831–848
8. Wu Y, Wu Q, Zhu J (2019) Improved EEMD-based crude oil price forecasting using LSTM networks. Physica A 516:114–124
9. Abdullah R, Shabri AB (2014) Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines. Am J Appl Sci 11(3)
10. Shabri A, Ruhaidah S (2014) Daily crude oil price forecasting using hybridizing wavelet and artificial neural network model. Math Prob Eng 2014

11. Ahmed RA, Shabri AB (2013) Fitting GARCH models to crude oil spot price data. Life Sci J 10(4):654–661
12. Mohammad MA, Talal HO, Rehman S, Hussain AA (2014) Support vector machines for wind speed prediction. Renew Energy 29(6):939–947
13. Wang W, Men C, Lu W (2008) Online prediction model based on support vector machine. Neurocomputing 71(4–6):550–558
14. Kaijian H, Lai KK, Yen J (2009) Crude oil price prediction using slantlet denoising based hybrid models. In: 2009 International Joint Conference on Computational Sciences and Optimization vol 2, pp 12–16
15. Wen X, Yu L, Xu S, Wang S (2006) A new method for crude oil price forecasting based on support vector machines. In: International conference on computational science, Springer, Berlin, Heidelberg
16. Jingtao Y, Li Y, Tan CL (2000) Option price forecasting using neural networks. Omega 28(4):455–466
17. Gioqinang Z, Hu MY (1998) Neural network forecasting of the British pound/US dollar exchange rate. Omega 26(4):495–506
18. Ahmad B, Peyhani HM, Akbari M (2014) Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization. Expert Syst Appl 41(14):6235–6250
19. Gonzalez JAA, Rey CMO, Garcia JCF (2012) A self-organizing neural fuzzy system to forecast the price of Ecopetrol shares. In: 2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), pp 1–6
20. Kang SH, Kang S-M, Yoon S-M (2009) Forecasting volatility of crude oil markets. Energy Econ 31:119–125
21. Wang Y, Yu W, Huang D (2010) Forecasting crude oil market volatility: Further evidence using GARCH-class models. Energy Econ 32(6):1477–1484
22. Wang Y, Wu C (2012) Forecasting energy market volatility using GARCH models: can multivariate models beat univariate models? Energy Economics 34(6):2167–2181
23. Laux P, Vogl S, Qiu W, Knoche HR, Harald K (2011) Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. Hydrol Earth Syst Sci 15(7):2401–2419
24. Chen TT, Takaishi T (2013) Empirical study of the GARCH model with rational errors. J Phys: Conf Ser 454(1):012040
25. Lee Y-H, Pai T-Y (2010) REIT volatility prediction for skew-GED distribution of the GARCH model. Expert Syst Appl 37(7), no. 24:4737–4741
26. Khashman A, Nwulu NI (2011) Intelligent prediction of crude oil price using support vector machines. In: 2011 IEEE 9th international symposium on applied machine intelligence and informatics (SAMI) 25, pp 165–169
27. Akay MF (2009) Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 36(2), no. 26:3240–3247
28. Haykin SS (2008) Neural networks and learning machines 28
29. Lakshmanan I, Ramasamy S (2015) An artificial neural-network approach to software reliability growth modeling. Procedia Comput Sci 57(29):695–702
30. Kiran NR, Ravi V (2008) Software reliability prediction by soft computing techniques. J Syst Softw 81(4), no. 30:576–583

Glaucoma Retinal Image Synthesis Using the GAN

Yerrarapu Sravani Devi and **S. Phani Kumar**

Abstract The human eye is an organ that responds to pressure and light. The eyes and associated structures may be affected by a variety of diseases, abnormalities, and ageing-related changes. One of the eye diseases is glaucoma. Without treatment, it may harm the visual nerve and cause blindness. The early detection of glaucoma minimizes the risk of vision loss. The proposed model synthesizes highly realistic controllable fundus images to obtain precision in detecting glaucoma through a deep learning model. Generative adversarial network (GAN) is a strategy for unsupervised machine learning that can be used to improve datasets and yield the collected images to be indistinguishable from the real-world data. The deep convolutional GAN (DCGAN), another variant of GAN, outlines the model's architectural limitations needed to develop high-quality generator models in an efficient manner. The enhanced dataset which is obtained from data augmentation as well as the original ACRIMA dataset of fundus images are separately given to CNN classification model for detection of glaucoma disease. The proposed research provides effective high-resolution processing of fundus images. The synthesized data using DCGAN enhances the model's ability to identify glaucoma at its earliest stages.

Keywords Glaucoma · GAN · Data augmentation · DCGAN · ResNet · Transfer learning · Eye disease

1 Introduction

Among the most clinically significant glaucoma diagnostic techniques is optic nerve head (ONH) screening. Glaucoma is a persistent condition of the eyes that can irreversibly vision loss and gradually weakens the optic nerve. By manually measuring the ONH geometric structures, one can separate glaucomatous patients from healthy ones based on the ONH assessment. In this method, clinical criteria for glaucoma

Y. Sravani Devi (✉) · S. Phani Kumar
GITAM Deemed to Be University, Hyderabad, Telangana, India
e-mail: y.sravanidevi@gnits.ac.in

screening [1, 2] are provided using measurements such the vertical cup-to-disc ratio (CDR). Many forms of glaucoma have no symptoms at all. A vision alteration might not be noticeable to you until the condition has progressed to an advanced stage because the effect is so gradual. Glaucoma-related since vision loss is persistent, it's important to get regular eye exams that monitor your eye pressure to detect the disease early and start the right course of treatment. Early glaucoma detection can reduce or even prevent lifelong vision loss. Glaucoma gradually develops resulting in blindness if neglected. Within 20 years, glaucoma patients lose the use of at least one eye, even with therapy. Glaucoma image and healthy image are shown in Figs. 1 and 2.

Fig. 1 Glaucomatous eye

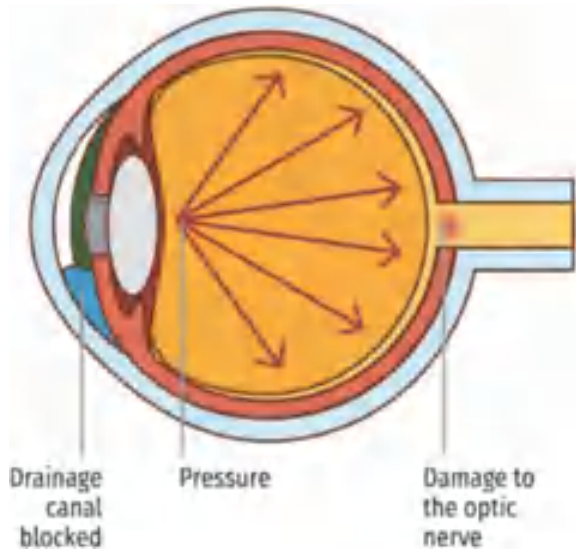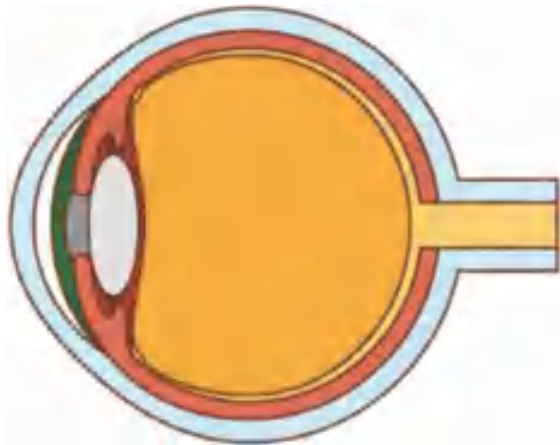Drainage canal blocked   Pressure   Damage to the optic nerve

Fig. 2 Healthy eye

Several deep learning models could be trained to detect glaucoma. Big data improves its performance since it allows it to acquire and detect a wider range of features from the training dataset, producing a system that is extremely accurate and dependable.

Deep learning-based object detection [3] algorithms like fast region-based convolutional neural networks (FR-CNN), you only look once (YOLO), etc. could be used as an automated localization and segmentation approach for detection. Another alternative approach is to use CNNs. The most often used type of DL is the convolutional neural network (CNN). It selects utilizing the annotated training dataset to determine the best discriminative deep features for image recognition, analysis, and classification. The creation of an accurate and quick automated glaucoma diagnosis method is required [4, 5]. This can be facilitated through DL models because there is no requirement to have a manually developed set of features for classification because DL networks can learn for themselves while being trained [6]. For medical images, analysis domains do not have access to a large, annotated dataset needed to train a DL model. Its performance may be substantially reduced by a small number of images. Data augmentation is a solution to the data shortage issue.

Data augmentation is the process of increasing the volume of data used to train a model. For correct predictions, deep learning models typically need a lot of training data, however this is not always possible. As a result, new data are added to the already collected data to enhance the generalized model. An approach to generative modelling using convolutional neural networks and other deep learning techniques is referred to as GANs, or generative adversarial networks.

1.1 Online Data Augmentation

The training data's level and kind of variance were both increased using this data augmentation technique. In image classification problems, data augmentation is frequently employed and can produce models that perform better, are more generalizable, and are invariant to specific kinds of image changes and fluctuations in image quality.

The earliest examples of data augmentations' success come from straightforward changes like random cropping, colour space augmentations, and horizontal flipping.

1.2 Data Augmentation Using GAN

The two networks are the generator model and the discriminator model that make up the fundamental GAN architecture [7]. Because the two networks are trained concurrently and are pitted against one another, much like in a zero-sum game like chess, GANs receive the name "adversarial." The generator model generates original images. The generator seeks to create images that look so authentic they fool the

discriminator. The input is typically random noise in the most basic GAN architecture for image synthesis [8–10], and the output is a synthesized image. A binary image classifier, which you are probably already familiar with, serves as the discriminator. Identifying if an image is real or fake is part of its job.

Combining all these together in the architecture of GAN represented, the generator creates false images, additionally, we give the discriminator several batches of real and fake images. Following that, the discriminator determines whether an image is real or faked. By incorporating feedback from the discriminator, the generator component of a GAN learns to make fake data.

The following elements make up the generator training section of the GAN: random input generator network that transforms it into a data instance, discriminator network that categorizes the created data, discriminator output, and generator loss that restricts the generator when it is unable to fool the discriminator. The discriminator of a GAN is only a classifier. It attempts to distinguish between data that is collected and data that is generated. Any kind of network structure that is appropriate for classifying the data may be employed.

While the discriminator is being trained, the generator is not. While it generates samples for the discriminator to learn from, its weights remain constant. It follows the following rules in particular: strided convolutions (discriminator) and fractional strided convolutions should be used without any pooling layers (generator). In the discriminator and generator, use batchNorm, for deeper designs, remove fully connected hidden layers. Every layer in the generator employs ReLU activation aside from the output; the output uses tanh, utilizing leaky ReLU activation across all layers of the discriminator.

To create stronger deep learning models, the key contributions of this research are to increase the number and quality of training datasets. The application of augmentation methods based on the variant of GAN, i.e. deep convolutional generative adversarial network (DCGAN) and build classification models to detect glaucoma, using datasets enhanced from DCGAN and traditional data augmentation techniques and compare the accuracies.

2 Related Work

Agarwal et al. [11] have developed a technique to identify glaucoma in retinal fundus pictures using cup-to-disc ratio (CDR) and rim-to-disc ratio. This is a noise- and image-quality-independent adaptive threshold-based approach. In addition, it is observed that rim-to-disc ratio gives higher accuracy in glaucoma detection when paired with CDR.

A classification method based on SVM for the identification of glaucoma was explored by Narasimhan et al. This model discussed the use of a local entropy thresholding strategy for glaucoma and was based on the extraction of the blood vessels found in the optic. It is observed that maximum accuracy of 95% accuracy

was obtained by utilizing the SVM classifier. Hussain and Holambe also applied the SVM classifier for the fundus images for detection and classification of glaucoma.

Few researchers by name Mittapalli and Kande [12] glaucoma was examined utilizing the optic cup's structural and grey scale features. QAzThe sample size used for this approach was 59 retinal images and the calculated 89% F-score with SVM classifier. An automated glaucoma detection method has been created by Bock et al. using a probabilistic two-stage classification approach that extracts the affordable digital colour fundus camera's photos; a trustworthy, competitive, and probabilistic glaucoma risk index was created (GRI). GRI is compared with medical relevant glaucoma parameters using and observed 88% accuracy.

Various researches [13] contributed nationally and internationally for the detection of eye diseases like DR and ARMD. The datasets that can be used to create the models are collected in a clinical environment and relate to few diseases. Very minimal dataset is available for glaucoma. Different machine learning techniques like SVM, decision trees, the use of ANN demonstrates highly accurate, sensitive, specific, and AUC classification between glaucoma and healthy eyes. Detection of glaucoma characteristics and measuring the severity is challenging using portable devices in the fields by the ophthalmologists. So, there is a need for a comprehensive and automated screening tool using deep learning techniques for early prediction as even with routine care, the majority of patients progress slowly and could cause significant losses.

The ophthalmologist only reviews clinical evaluations when glaucoma is suspected, according to Yuki Hagiwara in utilizing fundus photos to assist in glaucoma diagnosis. There is a need for an automated glaucoma diagnostic system.

3 Glaucoma Image Synthesis Using DCGAN

Deep learning networks can learn for themselves while being trained, minimizing the requirement for manually developed sets of categorization characteristics. Applying methods like convolution neural networks (CNN), [14] residual networks (ResNet-50), etc. using the fundus images to identify glaucoma requires huge datasets to provide excellent accuracy and performance. Manually collecting the images to form a huge dataset is tedious. A small dataset doesn't provide good results and will lead to misclassifications.

The proposed architecture is shown in Fig. 3. An architecture for generative adversarial networks is called deep convolutional GAN (DCGAN) [15]. Convolutional and convolutional transpose layers are utilized by DCGAN for the generator and discriminator, respectively. Radford et al. proposed in their article unsupervised model learning using deep convolutional generative adversarial networks. There are two steps in DCGAN training: (1) training a discriminator network (2) training in generator networks. The discriminator's primary objective is to accurately identify whether to teach the generator to produce fake images more skillfully, regardless of whether the input image is fake or real. First, the discriminator is trained to compute

log D on a batch of original images (x). Second, the generator creates a collection of fake images, and using this collection of fake images, the discriminator is trained to determine $\log(1 - D(G(z)))$. A total of 200 epochs of the DCGAN are used to train the model, batch size is 64 and learning rate is 0.0002, and the DCGAN was able to create images that, across 60 epochs, represented retinal images. 500 epochs later, the quality of the generated images continued to increase. Figure displayed a sample grid of real and fake images.

Generator
The generator model takes as input a point in the latent space and outputs a single 64×64 RGB image. The generator architecture is given in Table 1.

Discriminator
The discriminator model receives a single 64×64 RGB image as input and produces a binary prediction of whether the image is real (class = 1) or fraudulent (class = 0). It uses a convolutional neural network as its implementation. The discriminator architecture is given in Table 2.

The generator is the initial neural network in a DCGAN [16]. It begins with a random input1 and repeatedly produces data that is close to the calibre of data from
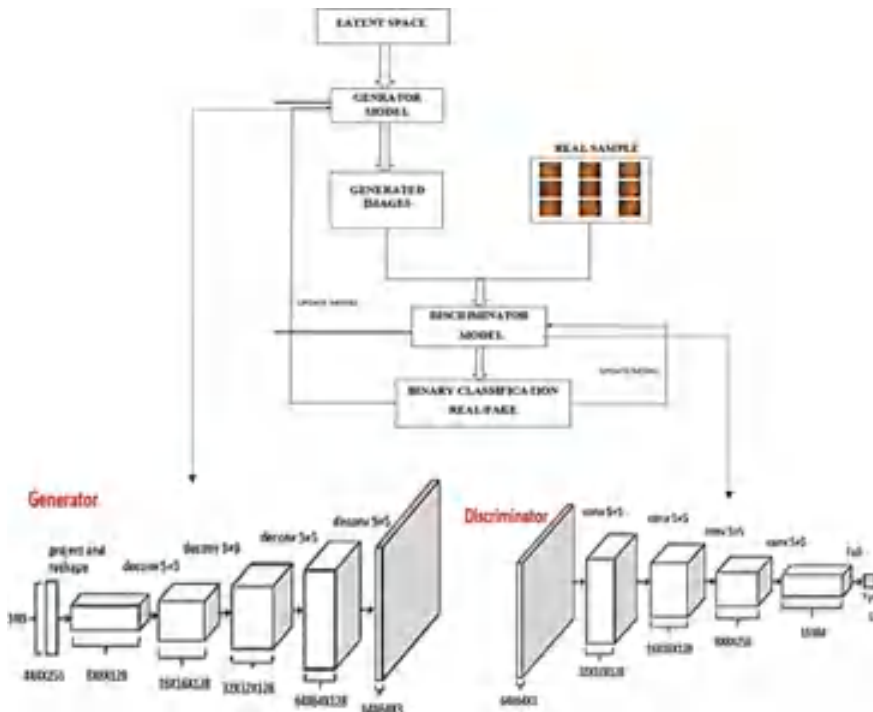

Fig. 3 Proposed architecture

Table 1 Generator architecture

Layer type	Output shape	Param #
(Dense)	(None, 4096)	413,696
Reshape (reshape)	(None, 4, 4, 256)	0
Conv2DTran	(None, 8, 8, 128)	295,040
Leaky_re_lu_1 (LeakyReLU)	(None, 8, 8, 128)	0
Conv2DTran	(None, 16, 16, 128)	147,584
Leaky_re_lu_2 (LeakyReLU)	(None, 16, 16, 128)	0
Conv2DTran	(None, 32, 32, 128)	147,584
Leaky_re_lu_3 (LeakyReLU)	(None, 32, 32, 128)	0
Conv2DTran	(None, 64, 64, 128)	147,584
Leaky_re_lu_4 (LeakyReLU)	(None, 64, 64, 128)	0
Conv2DTran	(None, 64, 64, 3)	3459

Table 2 Discriminator architecture

Layer type	Output shape	Param #
Conv2DTran	(None, 32, 32, 128)	73,856
Leaky_re_lu_1 (LeakyReLU)	(None, 32, 32, 128)	0
Conv2d_2 (Conv2D)	(None, 16, 16, 128)	147,584
Leaky_re_lu_2 (LeakyReLU)	(None, 16, 16, 128)	0
Conv2d_3 (Conv2D)	(None, 8, 8, 256)	295,168
Leaky_re_lu_3 (LeakyReLU)	(None, 8, 8, 256)	0
(Flatten)	(None, 16,384)	0
(Dense)	(None, 1)	16,385

the real world. To achieve this, it sends its output to the discriminator, a different neural network that uses training data to gradually increase its classification accuracy. The discriminator then sends the generator its output back. In terms of implementation, the discriminator's loss function contains the generator's loss function, whereas the generator's loss function is present in both the discriminators and the discriminator's loss function. The loss functions contain the output (classification), and during training, the generator updates its weights by backpropagation.

4   Results and Discussion

4.1   Dataset Description

The dataset is a csv file consisting of 650 rows and 5 columns. It has images belonging to two classes, i.e. glaucomatous and non-glaucomatous. The five attributes in the csv file are: filename, CDR (cup-to-disc ratio), eye (left or right), set (training set or validation set), glaucoma classification (class label)). There are 650 images in the dataset, in that 168 glaucoma and 482 non-glaucoma images.

4.2   Data Augmentation Using GAN

After the data augmentation using GAN is completed, an enhanced dataset is produced. GAN works as a data augmentation tool here and helps generate real-like images which can further be used to train some DL models for detection later. This model generates 200 images for each class while training. As it can be seen by the image samples below, ranging from 0th epoch to 200th epoch and it shows the change from plain noise to a relatively good resembling fundus image (Figs. 4 and 5).

4.3   Performance of Model

One approach of evaluation is to use the quality of the generated synthetic images as a benchmark for the model which is one method of evaluation. Manual examination of the resulting images provides a clear understanding.

Discriminator Loss
The discriminator classifies both the real data and the false data from the generator while it is being trained. It penalizes itself for mistakenly classifying a real instance as fake or a fake instance (created by the generator) as real by maximizing the following function.

$$\Delta_{\theta d} 1/m \sum_{i=1}^{m} [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$$

Maximizing $\log(1 - D(G(z)))$ would help the generator label the fake image it creates more accurately. The likelihood that the generator classifies the real image correctly is represented by the function $\log(D(x))$.
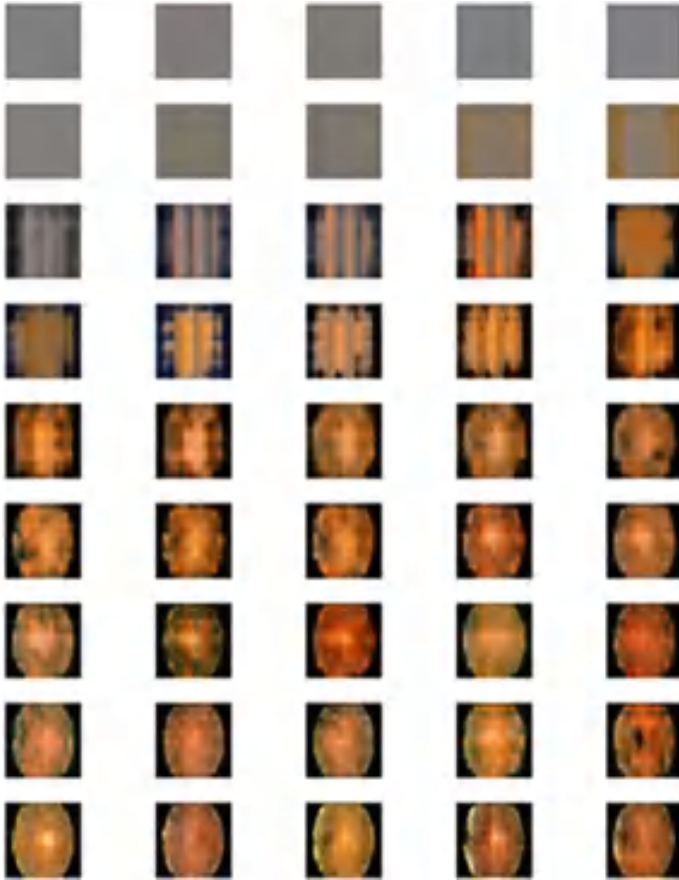
Fig. 4 Glaucoma synthesized images

Generator Loss

As it is taught, the generator samples random noise and outputs something from it. The discriminator then evaluates whether the output is "real" or "fake" based on how effectively it can distinguish between the two. The discriminator's classification is then used to calculate the generator loss; if it is successful in fooling the discriminator, it is rewarded; if not, it is penalized. Reduce the value of the following equation to train the generator:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^{(i)})))$$

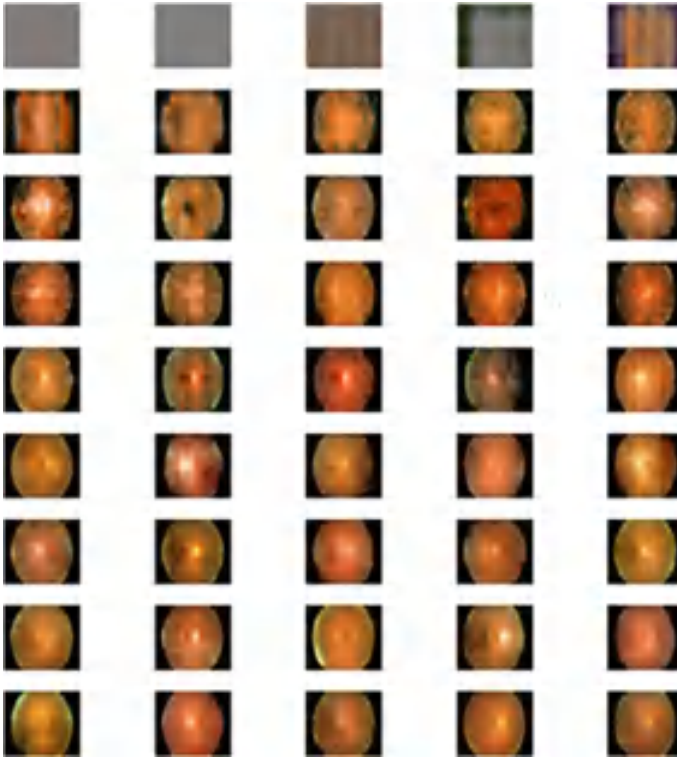The min–max loss is another name for the conventional GAN loss function.

Fig. 5 Non-glaucoma synthesized images

$$E_x[\log D(x)] + E_z[\log(1 - D(G(z)))]$$

D is the discriminator's estimate of the probability that data instance x actually exists (x). E_x denotes the expected value for all instances of real data, $G(z)$ denotes the output of the generator when given noise z, $D(G(z))$ denotes the discriminator's estimation of the probability that a fake instance is real, and E_z denotes the expected value for all random inputs to the generator (effectively, the expected value across all generated fake instances $G(z)$). This function is attempted to be minimized by the generator and maximized by the discriminator. Each time the training loop iterates, the console receives reports on the discriminator's loss, accuracy as well as the generator model's loss given in Table 3.

Figure 6 showing the discriminator loss for real images, the generator loss for generated fake images, and the generator loss for generated fake photos (green).

Initially the losses are inconsistent. As the epochs increase losses become stable after the 175th epoch. Typically, the generator loss is higher and may be in the range of 1, 1.5, 2, or even higher.

Figure 7 showing the A line plot illustrating the discriminator's performance when trained on actual (blue) and fake (orange) images.

Table 3 Generator and discriminator losses and accuracies

Epochs	Discriminator loss	Generator loss	Accuracy for real images (%)	Accuracy for fake images (%)
40	0.279	2.644	97	100
80	0.612	1.843	80	99
120	0.855	1.212	65	100
160	0.599	1.054	75	100
200	0.702	1.150	44	100

X-axis: No. of Epochs, Y-axis: Loss function value
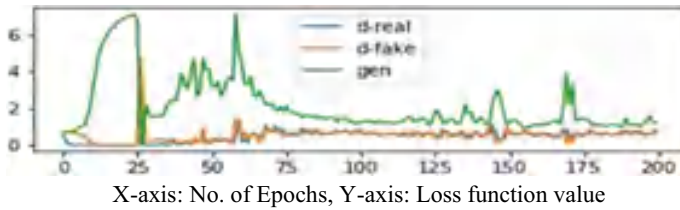
Fig. 6 Representation of generator and discriminator loss graph

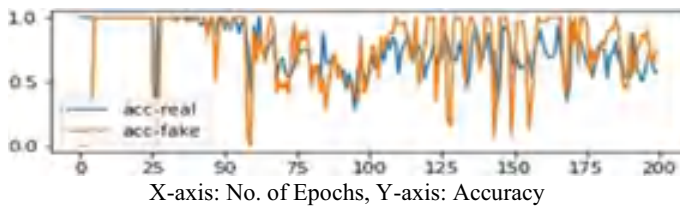X-axis: No. of Epochs, Y-axis: Accuracy

Fig. 7 Representation of real and fake images accuracies

The discriminator loss of a stable GAN will range from 0.5 to 0.8. The discriminator should produce a probability of 0.5 indicating that the generated data is real, and the generator should simulate the real data. In other words, the discriminator loses equivalent certainty regarding the reliability of new data coming from the generator.

5 Conclusion

The utilization of data augmentation, the proposed model enhances the glaucoma dataset by generating synthetic images, balancing all classes of the dataset. The generated images from the GAN model (output) could be combined with the original dataset assessing the problem of class imbalance. In computer vision, creating synthetic images from the actual (real) dataset is one of the main tasks. Harnessing the power of GAN as a data augmentation technique results to high-resolution real

fundus images. Size and quality of the dataset have been amplified. In the area of medical imaging and the identification of retinal diseases, DL has already demonstrated incredibly promising success. The creation of a greater number of high resolution and quality images can improve the classification accuracy. In terms of annotated data collection, time, and energy usage, training images can be expensive. In the proposed DCGAN model, it generates the high-quality synthetic images using the loss functions of generator and discriminator and real and fake images accuracies.

References

1. Damms T, Dannheim F (1993) Sensitivity and specificity of optic disc parameters in chronic glaucoma. Invest Ophth Vis Sci
2. Xu Y, Lin S, Wong TY, Liu J, Xu D (2013) Efficient reconstruction-based optic cup localization for glaucoma screening. MICCAI, 2013
3. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikainen M (2020) Deep learning for generic object detection: a survey. Int J Comput Vis 128(2):261–318
4. Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. Med Image Anal 58:101552. [CrossRef] [PubMed]
5. Niu Y, Gu L, Lu F, Lv F, Wang Z, Sato I, Zhang Z, Xiao Y, Dai X, Cheng T (2019) Pathological evidence exploration in deep retinal image diagnosis. In: AAAI, 2019
6. Moreno-Barea FJ, Jerez JM, Franco L (2020) Improving classification accuracy using data augmentation on small data sets. Expert Syst Appl 161:113696. [CrossRef]
7. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: NeurIPS, 2014, pp 2672–2680
8. Nie D, Trullo R, Petitjean C, Ruan S, Shen D (2016) Medical image synthesis with context-aware generative adversarial networks. MICCAI 10435:417–425
9. Ruan Y, Li D, Marshall H, Miao T, Cossetto T, Chan I, Daher O, Accorsi F, Goela A, Li S (2020) Mb-fsgan: joint segmentation and quantification of kidney tumor on ct by the multi-branch feature sharing generative adversarial network. Med Image Anal, p 101721
10. Xu C, Xu L, Ohorodnyk P, Roth M, Chen B, Li S (2020) Contrast agent-free synthesis and segmentation of ischemic heart disease images using progressive sequential causal gans. Med Image Anal, p 101668
11. Agarwal A, Gulia S, Chaudhary S, Dutta MK, Travieso CM, Alonso-Hernandez JB (2015) A novel approach to detect glaucoma in retinal fundus images using cup-disk and rim-disk ratio. In: 2015 4th International work conference on bioinspired intelligence (IWOBI). https://doi.org/10.1109/iwobi.2015.7160157
12. Zhou T, Fu H, Chen G, Shen J, Shao L (2020) Hi-net: hybrid-fusion network for multi-modal mr image synthesis. IEEE Trans Med Imaging
13. Seetha M, Kalyani N, Sravani Devi Y (2022) An ensemble CNN model for identification of diabetic retinopathy eye disease. In: Satapathy SC, Bhateja V, Favorskaya MN, Adilakshmi T (eds) Smart intelligent computing and applications, volume 2. Smart innovation, systems and technologies, vol 283. Springer, Singapore. https://doi.org/10.1007/978-981-16-9705-0_19
14. Devi Y, Kumar S (2022) A deep transfer learning approach for identification of diabetic retinopathy using data augmentation. IAES Int J Artif Intell 11:1287–1296. https://doi.org/10.11591/ijai.v11.i4.pp1287-1296

15. Sravani Devi Y, Kumar SP (2020) A scoping review of diabetic retinopathy detection techniques using deep learning: taxonomy, methods, and recent developments. High Technol Lett 26(11):392–406
16. Devi YS, Phani Kumar S (2022) DR-DCGAN: a deep convolutional generative adversarial network (DC-GAN) for diabetic retinopathy image synthesis. Webology (ISSN: 1735-188X) 19(2)

Identifying Power Line Faults Using Fuzzy-Based Intelligent Control

K. Shravani, A. Srinivasula Reddy, and S. S. Tulasiram

Abstract The research article details a new method used for fault classification in real time in a system comprising power transmission following a multi-criteria-based method based on fuzzy logic. Only a three-line current detects faults like LLG, LL and LG. This document also presents a combination of wavelet fuzzy real-time approaches for digital forwarding. This proposed fault location algorithm uses wavelet transforms with fuzzy logic, unlike conventional algorithms based on deterministic computations on well-defined models, which must be protected. Next, identify the type of fault by comparing the sudden swings of the three-phase MRA. The effects of obstacle distance, obstacle initiation angle and obstacle impedance were studied, and a fault classification routine was designed to overcome those effects. The wavelet transforms use wavelet MRA coefficients to capture dynamic properties of non-stationary transient fault signals and define fault lines.

Keywords Wavelet transform · Fuzzy inference system · Fault location · Fault classification · MRA

1 Introduction

A well-specified system model to be safeguarded is the foundation for all theories produced in the area of electrical transmission line protection systems. This is challenging because the system model is complicated, its parameters are unknown, there is a lot of information to analyse, and once rules are established, it is difficult to explain

K. Shravani (✉)
UCE, JNTUH, Hyderabad, Telangana, India
e-mail: shravani773@gmail.com

A. Srinivasula Reddy
CMR Engineering College, Hyderabad, Telangana, India

S. S. Tulasiram
GNITS Autonomous, Hyderabad, Telangana, India

547

system deviations (i.e. rules cannot be adapted). This leads to the proper judgements being made when signals are unclear and the dynamic adjustment of system operating conditions. Neural network applications have recently been developed to solve most of the issues above. Problems involving uncertainty are also solved using fuzzy set theory. The main advantage of fuzzy logic is that representation of knowledge is explicit and uses simple relationships: "if–then". Some applications of fuzzy logic and neural networks in the protection of networks are included in every situation, which is not characterized by a well-defined and simple deterministic mathematical model that is easy to be used with fuzzy set theory, simple membership functions and simple rules for deriving accurate results [1]. Fuzzy sets are generally considered more accurate at handling different factors representing knowledge than neural networks. However, neural networks are not enabled to learn from past examples. The fuzzy logic system is inflected into three components: fuzzy inference, fuzzification and defuzzification. In fuzzfied layer, the membership degree is calculated per the if–then rules.

The decisions are based on input on linguistic variable forms deriving from membership functions. A membership function is an expression used to calculate the fuzzy set, value belonging to and membership degree within a set. These variables are matched against specific linguistic rules of if–then, and answer of every rule is attained with fuzzy implication. To run a compound inference rule, each rule's answer is weighted per the degree or impedance of input membership. The centroid of this answer is estimated to produce accurate output. These applications to the problem of fault classification in a system fault are found. Although the present method requires a sequence component for solving the problem of fault classification, training is conducted using data from both the designer's experience and a dataset sample. Another limitation of the approach is that several fuzzy rules exponentially grow with input, resulting in 17 rules being boxed for three inputs. In the present paper, an application for a new method is proposed for processing three-line currents for determining the fault and its type. An example of a three-phase power system is also presented with the help of EMTP software [2]. Processing of line current is done with an online wavelet transform algorithm. The final model's performance is tested with a test set.

2 Wavelet Analysis and MRA

See Fig. 1.

The localization criteria for both frequency and time may be satisfied using wavelets. Wavelets must oscillate; rapidly decrease to zero, with an average value of zero as a sufficient and necessary condition. Additionally, the wavelets are orthogonal to one another for the discrete wavelet transform that is being studied here [3]. In terms of frequency (via dilation) and time (via translation), the waves are localized. Wavelets may provide greater temporal and frequency resolution. By employing a very limited number of components, wavelet analysis may successfully reproduce the
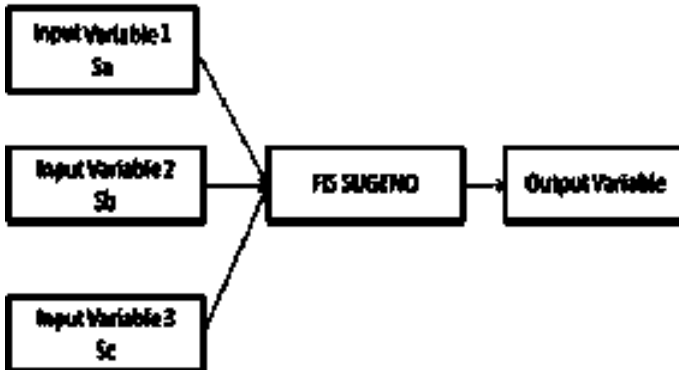
Fig. 1 Fuzzy inference system

signal. The wavelets that have been analysed are referred to as "mother wavelets", and their enhanced and translated variants are referred to as "daughter wavelets". The discrete wavelet transform (DWT) is its equivalent that may be applied digitally. To collect sub-band data from the simulated transients, the resulting waveforms are analysed using multi-resolution wavelet techniques. Since signal matching is crucial in wavelet applications, the Daubechies-8 wavelet is utilized in this study since it resembles the signal which is being processed. Additionally, Daubechies wavelets' effectiveness is based on precisely reconstructing the transient signals from a power system as defined in and applicability of Daubechies-4,8: D-4 and D-8 wavelets from the Daubechies wavelet family for analysing power system transients is outlined in Daubechies-8 wavelets. Additionally, in comparison with the D-4 wavelet, the D-8 wavelet is more localized, or supported compactly in time, making it better for quick transient analysis [4]. It also offers almost flawless reconstruction. Because it is better and smoother, the Daubechies-8 wavelet is considered to be more suited for describing transient signals than the Daubechies-4 wavelet. Oscillatory nature is another characteristic of transient signals. In this study, discrete signal decomposition and Mallat's algorithm-based reconstruction are used to extract the wavelet coefficients of the signal using matrix equations.

3 MRA-Based Fault Detection Algorithm

As was mentioned in the section above, MRA may be used to retrieve sub-band data from the source signal. This sub-band data may include relevant fault signs for a malfunctioning power supply. This article aims at fault detection and fault classification using precise MRA output signals [5]. This task requires the specification of a variety of parameters, including the sampling rate, the wavelet type and the number of MRA filter banks. Frequency of sampling must not be high for minimizing the computing load on the program. Meanwhile, it has to be sufficiently high to record

the fault's distinctive features. In this work, we choose a 600 Hz sampling rate and an MRA filter bank with a single stage. This defect can be detected and classified using just the first-level MRA signal, as shown by the simulation results. The kind of wavelet used is another significant variable. After investigating many wavelets, it is discovered that the Daubechies D-4 wavelet, which has just four coefficients for the band-pass and low-pass $g(n)$ and $h(n)$, has a cheap computational cost and high performance. The "Alternative Transient Programs" (ATP) simulated the power system depicted in Fig. 2 to provide the algorithm with error information. Line P2C, with a length of 174.4 km was used for the fault simulation. It was believed that the relay would be located on Line P2C, namely at bus PR. For eliminating the high-frequency noise, a digital low-pass filter with 300 Hz cut-off was utilized. According to simulation findings, the sampled signals after low-pass filter operation contain a significant amount of fault signature for the MRA-based method. The twofold down-sampling in Fig. 1 has the effect of reducing the MRA representation redundancy. This is crucial for certain applications since it may reduce the memory requirements for the computer, but it is not essential for detecting fault as it will increase the algorithm's processing time. Consequently, the suggested method does not include this procedure. The error is immediately found by using the band-pass filter ~ (n) output from convolution with the source signal. The outcomes for the detailed signals for varied fault kinds with fault distance $D_f = 0.5$ pu and fault impedance $R_f = 0$ fi are shown in Figs. 3 and 4. As can be observed, during a power system's typical operation, the values of the MRA detail signals are relatively low. A significant number of MRA detail signals are produced when a failure arises. As a result, the dramatic fluctuations in the detail signals may be seen as the distinguishing characteristics of fault. These abrupt fluctuations enable accurate fault identification [6].

The method for fault detection using MRA has the following programme structure:

(1) Create a zero initial value for k_a, k_b and k_c.
(2) "The detail signals $D_{ij}(n)$, $D_{ahf}(n)$ and $D_{lof}(n)$ of three-phase currents are derived by convolving sampled signals with a band-pass filter ~ (n) after each sampling instant.
(3) Use the equation $SV_p(n) = ID_{lj}(n)\text{-}D_{lj}(n\text{–}N)l$, $p \epsilon$ a, b, c to compute the sharp variation value of 3 detail signals, wherein N represents the size of sample points/power cycle.
(4) If $SV_p > G_o$, then $k_p = k_p - l - 1$; otherwise, go to step first, here $p = \{a, b, c\}$ and G_o indicates threshold level determined by the detail values during normal & fault operations".
(5) If any of the k_a, k_b or k_c values approach 3, a fault is declared.

L-G, L-L-G, L-L and L-L-L faults are the fault types taken into account in the analysis. The simulations indicate that the fault onset angle has a significant influence on the phase samples of current and thus also on the O/P of the wavelet transformation of the signals after the fault. It is adequate to research the effects of the origin angle in the region of 0°–180° since the waves are periodic. The total of the level 3 output for three-phase currents has been determined as the parameter for categorization after extensive investigation [7]. If S_a is the total of the level 3 values for phase "a" of the

Fig. 2 ATP's flowchart for providing the algorithm with fault data
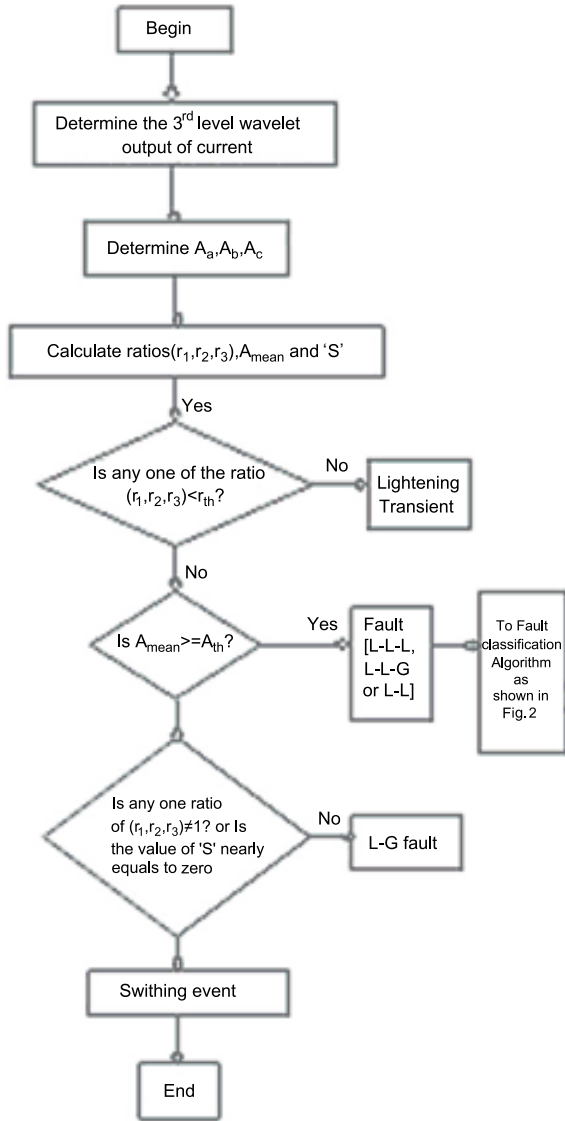
current; S_b is the total of the level 3 values for phase "b" of the current and; S_c is the total of the level 3 values for phase "c" of the current. Furthermore, if "$S_1 = S_a/S_b$; $S_2 = S_a/S_c$ and $S_3 = S_b/S_c$". If $S_a + S_b + S_c \sim = 0$, then the fault is categorized as an L–L–L fault. In this instance, all three-summation values S_a, S_b and S_c have similar magnitudes.

A model for fault localization is based on MRA wavelet coefficients. Even though the sum of MRA coefficients like S_a, S_b and S_c includes all the information on
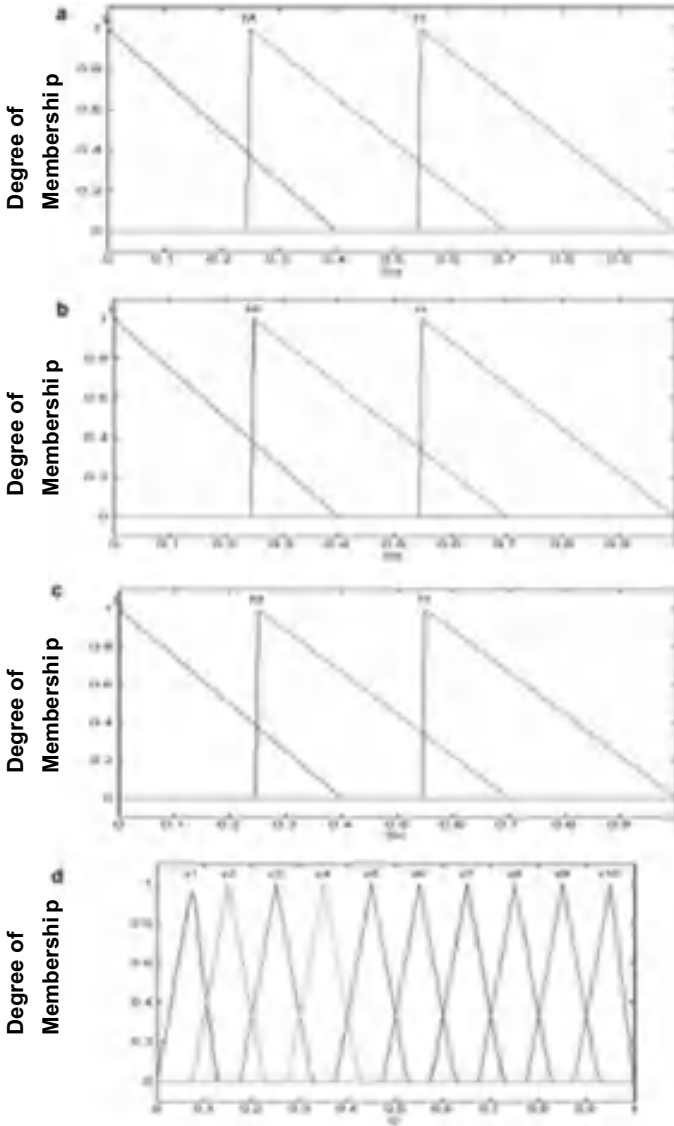
Fig. 3 Graphical representation of fuzzy variables and its corresponding membership degree functions: **a** S_a; **b** S_b; **c** S_c; **d** D

the frequency components connected to any transmission line problem, it is only available as raw data. Therefore, as with any form of raw data, these coefficients are inherently imprecise. The investigation begins with a fundamental fuzzy logic-based uncertainty model to express this uncertainty in a system.
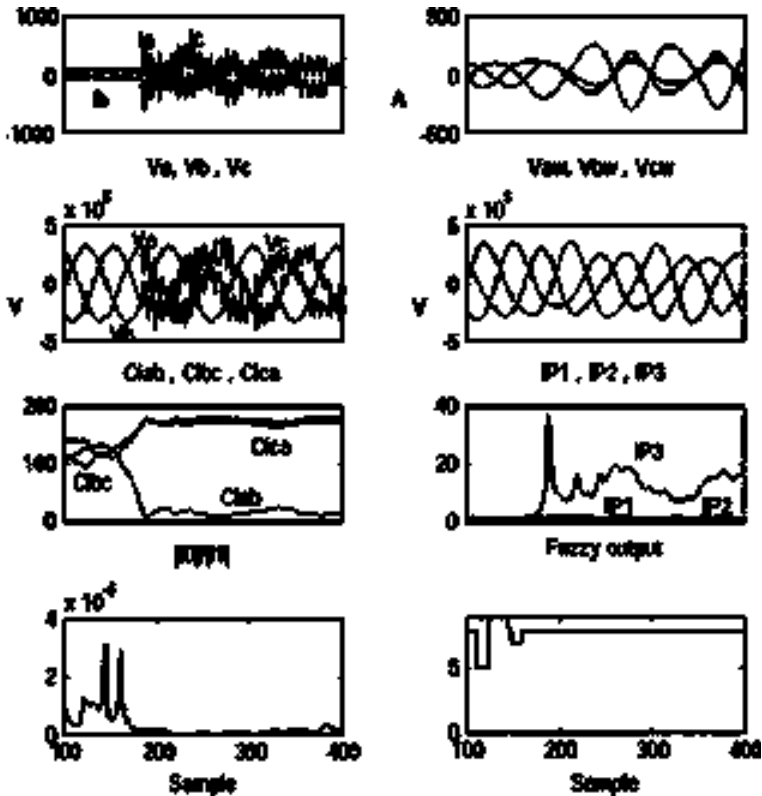
Fig. 4 Classifier output for fuzzy criteria (CA fault, 180 km, $R = 1{:}0$)

A rule-based expert system is created as a result of such analysis, which effectively collects information from the given data to provide a consistent result [8]. To acquire through results, expert knowledge is crucial, and this information is expressed by specific rules to reduce total uncertainty. The rules are linked together in what is referred to as a FIS, as in any "rule-based" system. As illustrated below, this article employees FIS to identify transmission line defects. As the foundation for fault localization, the wavelet coefficients of MRA for the three phases—S_a, S_b and S_c consistent with the three post-fault phase currents are added, and as a result, these numbers are inputs into the FIS. The fault's separation from the source's (D) termination is regarded as the output. Triangular fuzzy numbers, also known as standard membership fuzzy numbers, are employed to symbolize uncertain parameters, including the input (S_a, S_b, S_c) and output variable (D). As a result of human beings' initiate intellect and understanding, which includes contextual and semantic information as well as language truth values regarding the issue, these membership functions are obtained from human ability. This form more closely resembles human thought, which describes probability as a range as opposed to a single point [8]. For instance,

a linguistic statement like S_a would not be less than 0 or more than 0.4, and the best approximation is 0.25, which would be converted to a triangular fuzzy [0, 0.25, 0.4] number rather than the precise value 0.25. Typically, S_a, S_b and S_c decline with increasing distance of the fault from the end of the source. The linguistic variables are a collection of fuzzy sets that represent the universe of discourse for every fuzzy variable [9]. Since normalized values are employed in the calculations for the fuzzy I/P variables S_a, S_b and S_c, the range of discourse for these variables is 0 to 1 (Table 1).

In MATLAB, the Fuzzy Logic Toolbox was used to create rules for the FIS editor (Mamdani version) that evaluates the output range (D) based on expert views [10]. Centroid defuzzification is used to obtain the most important value for D. Fuzzy products ranging in complexity from Z1 to Z10 are generated by combining the

Table 1 Fuzzy rules for fault location

1	If (S_a is L) and (S_b is L) and (S_c is L), then (D is z_{10})
2	If (S_a is L) and (S_b is L) and (S_c is M), then (D is z_{10})
3	If (S_a is L) and (S_b is L) and (S_c is H), then (D is z_9)
4	If (S_a is L) and (S_b is M) and (S_c is L), then (D is z_9)
5	If (S_a is L) and (S_b is M) and (S_c is M), then (D is z_8)
6	If (S_a is L) and (S_b is M) and (S_c is H), then (D is z_8)
7	If (S_a is L) and (S_b is H) and (S_c is L), then (D is z_9)
8	If (S_a is L) and (S_b is H) and (S_c is M), then (D is z_7)
9	If (S_a is L) and (S_b is H) and (S_c is H), then (D is z_7)
10	If (S_a is M) and (S_b is L) and (S_c is L), then (D is z_9)
11	If (S_a is M) and (S_b is L) and (S_c is M), then (D is z_6)
12	If (S_a is M) and (S_b is L) and (S_c is H), then (D is z_6)
13	If (S_a is M) and (S_b is M) and (S_c is L), then (D is z_5)
14	If (S_a is M) and (S_b is M) and (S_c is M), then (D is z_5)
15	If (S_a is M) and (S_b is M) and (S_c is H), then (D is z_4)
16	If (S_a is M) and (S_b is H) and (S_c is L), then (D is z_6)
17	If (S_a is M) and (S_b is H) and (S_c is L), then (D is z_6)
18	If (S_a is M) and (S_b is H) and (S_c is H), then (D is z_3)
19	If (S_a is H) and (S_b is L) and (S_c is L), then (D is z_8)
20	If (S_a is H) and (S_b is L) and (S_c is M), then (D is z_5)
21	If (S_a is H) and (S_b is L) and (S_c is H), then (D is z_3)
22	If (S_a is H) and (S_b is M) and (S_c is L), then (D is z_3)
23	If (S_a is H) and (S_b is M) and (S_c is M), then (D is z_2)
24	If (S_a is H) and (S_b is M) and (S_c is H), then (D is z_2)
25	If (S_a is H) and (S_b is H) and (S_c is L), then (D is z_1)
26	If (S_a is H) and (S_b is H) and (S_c is M), then (D is z_1)
27	If (S_a is H) and (S_b is H) and (S_c is H), then (D is z_1)

inputs with "if–then" rules provided by experts in the FIS editor (Sugeno version). For instance, if S_a, S_b along with S_c is L, thus error should be found in Z10. Similar to superclassification, there are many distinct kinds of faults. These include (L-L-G) faults, three-phase symmetrical (L-L-L), double line (L-L-G) faults, double line (L-L) faults and single-to-ground (L-G) faults. The initial angle of the fault as well as location on the transmission line values are extensively investigated. The method has been shown to reliably locate the proper position with respect to the fault line. About the actual distance, the maximum error in estimating that value is 6.5%. Fault zones are directly proportional to rules, making rule optimization more time-consuming and difficult. Since ANFIS "Artificial Neuro-Fuzzy Inference System" is being looked into for its potential to improve accuracy by dividing the line length into a greater variety of fault regions, and since ANFIS's rule optimization appears to be less cumbersome than FIS's, the two systems are often compared and contrasted [4].

4 Conclusion

This research implements a multidimensional fuzzy logic-based approach in transmission networks for fault classification. It was shown that, when using a deterministic approach to a problem, a basic fuzzy technique may be used to solve a problem that would normally need a more sophisticated solution. This strategy is fast, efficient and dependable, with good results over many system situations. This method is less complicated than the standard and ANN-based methods for identifying and categorizing temporary and transient disruptions. In this paper, it is proposed to use MRA wavelet theory to detect and classify power line disturbances. Wavelets, which can resolve several frequencies and times, are useful for this issue.

For this reason, the authors propose a new generic algorithm that does not depend upon fault location, impedance or incidence angle. The algorithm is robust, simple as well as generalizable. It applies to transmission lines of any voltage and may be used to identify high impedance faults.

References

1. Fault classification based on wavelet transforms. In: Proceedings of the IEEE T&D Conference, Atlanta, GA, Oct 28–Nov 2 2001, paper 01TD069
2. Kasztenny B, Rosolowski E, Saha MM, Hillstrom B (1997) A self-organizing fuzzy logic based protective relay—an application to power transformer protection. IEEE Trans Power Deliv 12:1119–1127
3. Youssef OAS (2003) Online applications of wavelet transforms to power systems relaying. IEEE Trans Power Deliv 18:1158–1165
4. Dalstein T, Kulicke B (1995) Neural network approach to fault classification for high speed protective relaying. IEEE Trans Power Deliv 10:1002–1011

5. Dash PK, Mishra S, Salama MMA, Liew AC (2000) Classification of power system disturbances using a fuzzy expert system and a Fourier linear combiner. IEEE Trans Power Deliv 15:472–477
6. New algorithm to phase selection based on wavelet transforms. IEEE Trans Power Deliv 17:908–914 (2002)
7. A wavelet-based technique for discrimination between faults and magnetizing inrush currents in transformers. IEEE Trans Power Deliv 18:170–176 (2003)
8. Dash PK, Pradhan AK, Panda G (2000) A novel fuzzy neural network based distance relaying scheme. IEEE Trans Power Deliv 15:902–907
9. Chen WH, Liu CW, Tsai MS (2000) On-line fault diagnosis of distribution substations using hybrid cause-effect network and fuzzy rule-based method. IEEE Trans Power Deliv 15:710–717
10. Lee HJ, Park DY, Ahn BS, Park YM, Prk JK, Venkata SS (2000) A fuzzy expert system for the integrated fault diagnosis. IEEE Trans Power Deliv 15:833–838

Robotic ARM for Effective Environmental Cleaning

P. Sunitha Devi, K. Suma, K. Srujana Reddy, Ch. Mandakini, and R. Pallavi Reddy

Abstract Often workers find it difficult to clean the places where heavy objects are dumped. In the present situation, many dump yards are found with heavy material which humans cannot lift, in such scenarios these robotic arms can be used to clean up the place. These robotic arms are controlled by humans sitting in a central hub station manually. A manual operator in the central hub can view the place through the camera fixed to robotic arm, which helps them to control the arm. With this method, there is no need to go to each place to clean it. From the central station, we can clean up the city. UART communication protocol is used to establish communication between the central hub and robotic arms. By using this protocol wireless communication can be established, where the central hub acts as a server and the robotic arm acts as a client. These robotic arms can be further updated and used for cleaning sewage blocks and underground drainage pipes. As robotic arms can work well under extreme conditions, it helps in reducing manpower and ensures the safety gear of workers.

Keywords Raspberry Pi · UART communication · Servomotor · Degree of freedom · Wireless communication · Gripper

P. Sunitha Devi · Ch. Mandakini · R. Pallavi Reddy
Narayanamma Institute of Technology and Science, Hyderabad, India
e-mail: sunitha@gnits.ac.in

Ch. Mandakini
e-mail: mandakinireddy@gnits.ac.in

R. Pallavi Reddy
e-mail: reddygaripallavi@gnits.ac.in

K. Suma · K. Srujana Reddy (✉)
Texas A&M University, College Station, TX 77840, USA
e-mail: srujana1998@tamu.edu

K. Suma
e-mail: ksuma@tamu.edu

1 Introduction

Robotic arm is a handler with some functions to human arm which is programmed easily. Joints allow rotating motion or linear displacement which is attached to the links like manipulator. The relation between manipulator and chain of kinematic is considered. The end effects are considered the chain of manipulator which is similar to human arm. Based on the purpose, the end effectors can design to perform any function like welding and spinning. The arms of robot will be controlled to perform many functions with more accuracy. It will be used in industries and also for home appliances [1].

Robotics is an area where we have plenty of opportunities for creativity and something new to do. It is a very dynamic environment with lots of technical variations. The distance between machines and humans is reduced with the advent of new technology to improve living standards with each passing day. Robotic arms are used in the study on robotics, with special features and design parameters. Most of the robotic arms are operated using an accelerometer sensor and an intelligent artificial algorithm. The robotic arms can attain an accurate virtual interaction. User-friendly robot control is an important open area in robotic science. Nowadays, robotic devices are essential components in virtually all industries. The biggest advantage of these weapons is their ability to operate in dangerous environments and in places that cannot be reached by humans. In these circumstances, without a substantial amount of safety measures, such as disposal of hazardous waste, toxic material, remote management of explosive devices and remediation and hostage cases arm robots can be used. In these hazardous environments, robots can operate safely.

These robots ensure human health and eliminate massive human workers. This can also be implemented for technology, manufacturing field and mega market field of operations, defense purposes, medical research, cutting, trimming, collecting, positioning small objects, etc. These are portable robotic machines and are very effective. Development in cleaning the sewage dumps using robots would be an excellent idea. Cleaning intoxicated areas requires manual power which might cause the person entering the dump yard to face several potential health risks [2, 3].

In these days, robots are incorporated into working tasks to change people's particularly to accomplish repetitive tasks. These are generally separated into two field's robotics for industry and operation. The International Robotic Federation (IFR) defined a service robot as semi-self-employed robot for efficiency, manufacturing the services with benefits of people and equipment.

2 Literature Review

Prototypes were developed and technically verified the concept of robotic vehicle using remote controller. The concept of managing unidentified objects in the environment could be dangerous and jobs would be easy for people to do without contrast.

This approach would handle difficult and complicated operations faster and more accurately [4]. The innovation of robotic arm is based on existing gesture recognition to maximize biomechanics skills by supplying multitouch signals that are helpful in triggering correct gestures. This system is used to recognize human gestures. Wide implementation scope in vulnerable environments as well as in free time. Multitouch devices should be compact, lightweight, safer, and convenient to use. Multitouch technology enables real-time inter-connectivity with virtual entities [5].

The power of robotic arm is based on degrees of freedom (3-DOF) for activity identification by haptic technology which concentrates mainly on how it works on disabled people. More advanced features such as obstacle detection and the way the method of image recognition is used in robotic arm can be regarded as future purposes [6]. The robotic arm was designed to perform various activities like grinding, carrying, and rotating. The robotic arm can be fitted with any required activities like soldering and fastening. Based on the specification, the arm of the automated assembly line will support different activities like heating elements and rotating components and placement. For example, in an automobile production line, robotic arms execute several activities, like shifting and component rotation while production [7]. The robotic arm was proved to be client-friendly, and the incorporation of the sensors was very beneficial in obtaining information on the location of the arm. In real-time scope, the LabVIEW was used as a source for object orientation. Picking or putting activity is however provided via a LabVIEW monitor. The robot incorporates inverse kinematics to calculate the required movement when the parameters have been reacquired [8].

3 Proposed Solution

See Fig. 1.

3.1 Transmitter

In the transmitter section, we use a radio frequency (RF) module to establish communication with the robotic arm. This transmitter section is placed at the Center Hub of GHMC, and a trained person would be operating the robotic arm by visualizing the captured video from the camera on robotic arm at the receiver side. A trained person at the central hub can operate the robotic arm through commands like move forward, move backward, left, right, bend the arm, raise the arm, move arm toward left, move arm toward right, and hold the object.
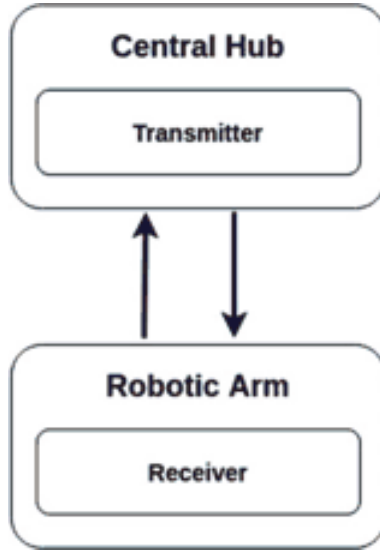
Fig. 1 Basic architecture of module

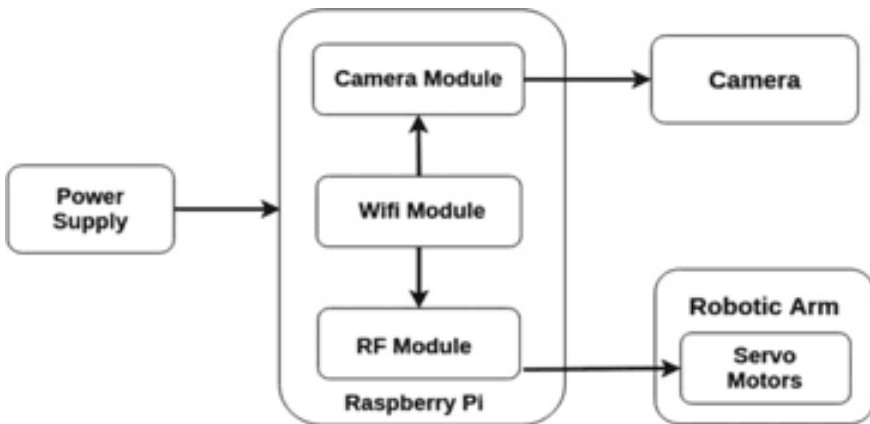## *3.2   Receiver*
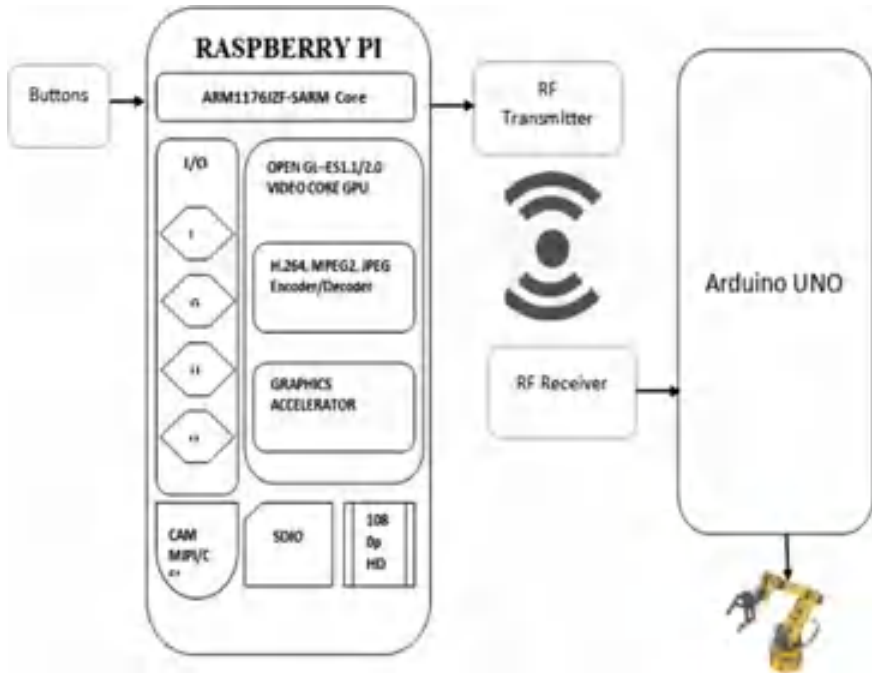
See Fig. 2.

Fig. 2 Receiver module

Fig. 3 Raspberry Pi module

3.3 *Raspberry Pi*

It is a single board device with low-cost credit card. It has ample power to run sports, word processors such open offices, image editors such as GIMP and other related programs. Since the SD card reader does not have storage hard drive and OS picture can be shot using an SD card. Pi is based on Broadcom SoC (chip system) with a GPU and 26–52 MB and ARM processor. 5 V power supply is provided via USB cord, which displays TV/monitor information via a DVI/HDMI port, HDMI cables or DVI converter HDMI cable, uses USB mouse/keyboard as an input, uses Ethernet to link to the network.

See Fig. 3.

3.4 *UART Protocol*

UART is a hardware interface that provides serial, asynchronous, bidirectional interactions. It needs two data lines one for transmitting, and another for receiving. One device's transmit line is connected to the second device's receiving line, and vice versa for bidirectional transmission.
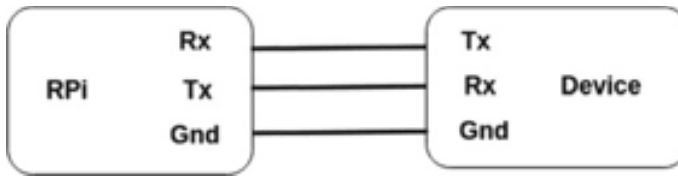
Fig. 4 UART module

See Fig. 4.

3.5 Servometer

Servos are DC engines with optimized gears' circuits and response controls, and motor drivers are also not required. A servomotor is a rotating actuator that allows for accurate circular orientation control. They comprise of the engine and a sensor that will provide location information through a restricted gearbox. They usually need a relatively complex controller, often a device specially constructed for servomotor implementations. For systems like robotics, CNC machines, and automated processing, servomotors are used. The servo motor is attached to the output shaft and contains some control circuits and a potentiometer. This can be used to track the servo motors current angle by the control circuitry's engine switches off when the shaft has a perfect position. If the system feels the inclination is inaccurate, it pushes the motor in the right direction before the trajectory is accurate. The output shaft of the servomotor travels around 180°. It is usually within the limit of 210°, but it largely depends on the producer. Due to mechanical stop built onto the main output gear, a typical servo cannot turn any further mechanically. The power added to the engine is commensurate with the distance to be traveled. Thus, engine must operate at maximum speed if the shaft must turn around great distance. If only a small amount is required, engine is running at a slower speed (Fig. 5).

This is referred to as proportional power. A servomotor is a servo device, as the name implies. More precisely, it is a closed-loop servo mechanism which uses feedback about the position to control its movement and location. The controlling input is a signal which represents the direction regulated for the response wave, which is either analog or digital. The motor, as well as some encoders, are coupled to include information on direction and velocity. In the simplest situation, just the path is measured. The reference location, the controller's external input, is compared with the measured output location. If the location differs from the right position, there is a malfunction in the signal that will turn the shafts in both directions to push the trigger shaft to the proper position. As position approaches, error signal is reduced to zero and the motor stops. The simplest servomotors use position only which sensing through their engine's potentiometer and controller. Such servo motors are often not widely used for the commercial gesture control yet lay the foundation

Fig. 5 Servomotor structure

of cheap and effective servo designs. More advanced motors regulate the location of movement output. They can also adjust engine rpm, instead of running always at full speed. Both improvements are normally combined with PID control algorithm and allow the motor to move faster and more precisely to its controlled location with less overshot. For evaluating servo rotational velocity, the servo turn rate or transit time is applied. It means that servo normally takes 60° to move a fixed distance. For example, suppose we have 0.18 s/60-degree transit servo without load (Fig. 6).

This means the rotation of a full 180° takes about half a second, and further if the servo was mounted. This information is very useful if our robot application needs high servo response speed. If our servo is set to full spinning, it is also useful to calculate the maximum forward speed of robot. Note, when servo is at the lowest rotation angle and is directed to go at the highest turning angle when under load, the worst time to transform. A very high torque servo will take few seconds.
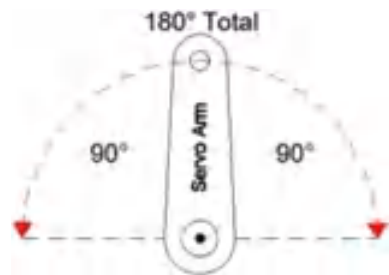
Fig. 6 Servomotor rotation

3.6 Robotic Arm Assembly

The deployment of the robotic arm can be seen in Fig. 7. It is the crucial part of the device in the project that performs angular motions for the role of lift and hold. The robotic arm is fitted with a spindle (to lift and take objects) as well as an arm (to hold and move hazardous waste particles). Servo motors play vital role in hand movement. These servomotors get commands from the central hub, where a human transmits signals by analyzing the video received from camera. When the transmitter and receiver power supply are switched ON, the transmitter will get connected to the receiver's wireless computer framework through UART protocol. As soon as the connection gets established, the camera attached to the robotic arm starts capturing the video and this video can be viewed by the person in central hub. According to the visual obtained, person at the central hub gives commands to the robotic arm. As per the commands, the robotic arm moves to lift and carry heavy intoxicated materials.

4 Conclusion

Robots play an important role in human life where many things can be completed with minimal effort. Robots have a broad range of uses in different sectors. Robots are adjusted well in locations where human entry is difficult. There are two main steps in this procedure, the first one is to guide the robotic arm to the desired place and the second one is to guide the robotic arm to lift heavy objects and clean up the place. The robotic arms based on wireless communication may be used in industries where there is a threat to human life. By implementing this concept of cleaning, our environment will reduce the life risks of many workers and these robotic arms can work more effectively in contributing to greater results.

5 Future Scope

In the future, these robotic arms can be automated to work on their own by using image processing techniques and machine learning algorithms. If automated robotic arms come into existence, then manual tasks can be minimized.

References

1. Kale VR, Kulkarni V (2014) Automation of object sorting system using pick & place robotic arm & image processing. In: Proceedings of 3rd IRAJ international conference, 5 Jan 2014, Mumbai, India
2. Kale VR, Kulkarni V (2013) Object sorting system using robotic arm. Int J Adv Res Electr, Electron Instru Eng 2(7)
3. Sawant YV, Karmuse SM, Kapase NB (2016) Accelerometer based control of robotic arm. Int J Robot Autom
4. Omijeh BO (2014) Design analysis of a remote controlled "pick and place" robotic vehicle. Int J Eng Res Dev
5. Divya Jyothi B, Krishnaiah RV (2013) Haptic technology—a sense of touch. Int J Sci Res
6. Tunwannarux A, Tunwannarux S (2007) Design of a 5-joint mechanical arm with user friendly control program. World Acad Sci, Eng Technol 27
7. Tunwannarux A, Hirunyaphisutthikul S, Design features and characteristics of a rescue robot. In: Proceedings of international symposium
8. Nair SR (2012) Design of a robotic arm for picking and placing an object controlled using LabVIEW. Int J Sci Res Publ
9. Craig JJ (1989) Introduction to robotics: mechanics and control. Addison-Wesley

Computing Workspot Infection Vulnerability for Sandboxing a Business Process

Supriya Vaddi, **Kollapalli Ramesh Babu**, and **Hrushikesha Mohanty**

Abstract   While business continuity plan is for strategic handling of adverse business situations, sandboxing is a proactive strategy to avoid such situations to happen. Particularly, to safeguard a business from pandemic spread due to social contacts, this paper computes the possibility of an infection of a workspot that uses one of the five basic workflow patterns, viz. sequential, concurrent, n-split, n-merge and on-spot-n. A workspot with a high vulnerability of an infection can be put under sandboxing for making a business process resilient.

Keywords   Business process · Sandboxing · Workspot infection · Business continuity plan · COVID-19 · Contact vulnerability

1   Introduction

During pandemic COVID-19, Smithfield meat factory in the USA turned an epicentre in spreading virus across the USA as some workspots in the factory got corona infected. Possibility of a workspot infection is due to the social contacts which can be due to either direct or indirect human contacts. The indirect spread is due to the exchanges of items handled by people at different workspots of a business process. Due to pandemic COVID, world has come to standstill and so also business. A proactive approach, called sandboxing as applied for system security, can be adopted here in safeguarding a business process. This is possible by identifying workspots having

S. Vaddi (✉)
Department of IT, G. Narayanamma Institute of Technology and Science, Hyderabad, India
e-mail: supriya_vaddi@yahoo.com

K. R. Babu
Department of CSE, Vidya Jyoti Institute of Technology, Aziz Nagar, Hyderabad, India
e-mail: krubabu@gmail.com

H. Mohanty
CVR College of Engineering, Hyderabad, India
e-mail: hmcs@cvr.ac.in

low immunity or high possibility to an infection and sandboxing those workspots by applying spread containment protocol like social distancing, hand hygiene and measures proposed by the identified centre for disease control. This is a different kind of BCP: business continuity plan that also needs to be researched by the business and faculty.

This paper takes on the problem in computing science perspective and proposes a computing technique to quantify the vulnerability of a workspot based on its degree of contacts, number of items in average transacted at and the number of people work on the spot. In order to model a unit distance spread, five workflow patterns, viz. sequential, concurrent, n-merge, n-split and on-spot-n, are considered, and the spread due to each is computed. The basic idea proposed here has potential to usher a plethora of research in both theory and application. The paper in the next section has a brief history on BPR: business process engineering and then in the next section talks on need of BCP: business continuity plan. The fourth section presents the computing technique the paper proposes. Then the final section ends with a concluding remark.

2   Business Process Engineering

High-scale industrialization as well as rapid increase in economic activities has given rise to 'business process engineering' (*BPR*) a new discipline for research and development by academia as well as implementation by business houses. The interest in BPR has gone in manifolds further, when there is a large-scale automation. Business process including manufacturing is growing in volume and veracity to a level almost unmanageable without being aided by agents. Now, the trend is even to employ software agents. Automation of a business process is not only to beat the time in managing business activities but also to apply good business practices. It drives business houses to apply process and policy-driven activities, so business remains well defined for engineering as well as analysis. This requirement of well defined business process has given rise to idea of specification, like any software system, the business process needs to be well specified. Specification usually follows two school of thoughts, viz. rigorous and formal, where the former provides a definite method mostly aided by graphic icons, e.g. UML: Unified Modelling Language diagrams, the latter proposes mathematical approaches, e.g. Z and RAISE.

In similar line, business processes are also being specified. In order to give a formal touch to UML, OCL: Object Constraint Language has been designed for annotating design elements with assertions, limitations and case-based actions expressed in FOL: first-order logic for clarity and to some extent for provability of a software system. Similarly, in case of business process specification, BPEL: business process execution language has been proposed by the researchers and the industry also accepted it. BPEL offers a mechanism to model a business process as a composition of varieties of transactions. Now a business process is seen as a network of transactions; though for a big business, its network could be too complex but well structured for tracing and analysis. Having both structural and functional specifications for a

business process, its implementation leads almost hassle-free execution meeting the expectations of business houses. However, a business does not always go through a frozen world following only the specifications defined during its development. As world changes, the need for continuity of business plan has received the concern of the designers. In the next section, we will present some works on business continuity plan.

3   BCP: Business Continuity Plan

Business continuity is a matter of resiliency in execution of a business process. There are some operational hitches like unavailability of raw materials at a source, absence of skilled staff, and fluctuation of market, and many such transient exceptions may crop up which needs to be factored before for continuity of a business. These are like exception handlers in case of programming languages. Business continuity can be seen as a resilient computing that has fault detection as well as system recovery mechanism. The strategy 'checkpointing' is widely used in case of transactional systems for recovery of faulty transactions. Concurrent execution of processes for the same computation, also know as N-version programming, is also another technique to maintain resiliency. Thus, worry for continuity has been there from early stages of computation. And the same concern is increasing when services are catered by software/hardware systems for providing timely and correct services. Responding to defined exigencies is fairly simple, but the complexity arises from unknown exceptions. Some programming paradigm provides default actions for handling undefined cases. The concept is borrowed from the concept of default reasoning in the domain of artificial intelligence. The default actions that are to be taken are decided by default handler. Handling business process cannot afford to shutting down plant or refusing business orders and defaulting a promised delivery and many more such actions which may cost a company very dearly.

In the recent past, many exigencies that disrupt business services come from different unseen directions like environmental problems, terrorism, political upheavals and many more. Such disruption sizes are so huge to handle either by making shift arrangements as interrupt service routine does during computation or by default handling, executing a task foregoing business interest. This has led to think of business continuity as a discipline of research in management science and computing sciences. The issues a business continuity plan needs to have are well described in [1]. The author proposes eight stages of a business continuity cycle. The stages include

   i. Start of a BCP
  ii. Evaluating business threats
 iii. Risk evaluation
  iv. Looking for alternatives for business continuity
   v. Business continuity plan design
  vi. Business continuity process generation
 vii. Testing of business continuity plan

viii. Review of business continuity plan.

BCP goes with business recovery plan that has five components including

- Identify a team to work for business recovery.
- The team needs to design recovery plan in detail.
- A detail recovery plan is to be converted into a process including details required for execution. A recovery process is thus synthesized.
- Then the process is tested before the recovery process is rolled out for execution.
- If still any flaw is found during testing, the recovery process is reviewed.

In nutshell, the paper proposes that following a BCP should be a corporate philosophy so that its readiness to face exigencies with resilience is guaranteed. Another comprehensive plan on BCP is presented in [2]. It identifies the sources from where disruptions are likely to appear. According to it, almost half of the disruptions are internally generated. Further, around one-fifth of the disruptions are from external sources and could be of high proportion like terrorist attack, tsunami, lockdown and something like having international impacts. The researchers hypothesize the need for constant look happening around globe and reviewing the company BCP. He also advocates the use of information technology to foresee the impending disasters to automate a BCP revision, if required. He is also advocating automated triggering of BCP services as and when the disruption is anticipated by processing information emanating from within a corporation and the world external to the company.

BCP planning for BPM: Business process management has strong relation with supply chain management (SCM). While BPM deals with operations inside a company, SCM talks of activities external to a company. For best management of discontinuity, the relation between the two is to be studied [3]. In pandemic COVID-19 spread time, BCP has become very essential particularly of very essential services that need to be managed in lockdown periods. Here, we will quote some of these recent works to comprehend the trend of research in SCM and its impact in management of COVID patient healthcare services. The paper [4] discusses a case study on PPE: Personal Protective Equipment supply management for frontier healthcare workers in the Republic of Ireland. Severe shortage of PPE across the globe has laid to explore the possibility of reuse upon disinfection of used PPEs in hospitals. Technologies like hydrogen peroxide agent-based disinfection and UV-light treatment have come up. Now the companies providing these PPE disinfection services are to be put in proper supply chain, so their services can be used effectively to meet the demands of health workers around the country.

The paper [5] talks of business continuity in different context, i.e. green engineering. Adopting green engineering as the author indicates brings interruptions to business workflow to meet the specified guidelines. A corporate house needs to strategize its works to win over these disruptions. For the purpose, (i) analyse a business workflow to find the trade-off between continuity and greenness, (ii) identify the factors that contribute to business continuity like flexibility, timeliness, robustness, sensibility, etc., and also identify the green factors suppliers adhere to. Thus, a unified

framework is to be worked out that cares gresilience, i.e. greenness and resilience; (iii) thirdly, a quantification of the factors is required for multi-criteria decision-making. This way the work proposes a unified framework for decision taking ensuring both BCP and green engineering.

Co-ordination is an essential dimension to achieve resilience in business paradigms. For a supply chain continuity, suppliers must co-ordinate for the reconstruction of the chain. Similarly, a business process needs to be re-engineered to achieve resilience. In order to understand a resiliency plan, researchers opt to simulate and analyse different scenarios. For simulation, the activities of a business domain is modelled. Here, we will refer to few prominent schemes for modelling. One among them is customer-centric agent-based modelling reported in [6]. The model has three levels of agents, viz. domain, conceptual and operational agents. In order to meet customer targets, the operational strategy is not only to factor operational specifications but also to exigencies that may crop in for disruptions in supply chain. The model proposed takes essence of multi-agent system concepts catering to agent autonomy, co-ordination, reactivity and pro-activity.

Resilience in business process as well as supply chain managements during COVID-19 pandemic is being re-looked primarily aiming for continuity of PPE and other essential equipments [8]. The Sendai framework is being looked in [7]. Post-COVID time will bring forth a great challenge for business continuity. The natural process is to nurture immunity as it happens in human system. The idea of humanity is being extended to corporate sector, so it can sustain upheavals due to disruptions as it happens in COVID-19. A study on corporate immunity in financial sector is reported in [9]. It finds the corporates having sound pre-COVID financial position, less connective to external agencies and particularly to global supply chain and less rigid executives have greater immunity to pandemic COVID-19.

Framework of actions [10] are identified by analysing web-based content of leading corporates. The paper addresses how different organizations have responded to their business during time of crisis by focusing on business continuity and value creation. The business models were changed to enhance operations and value system to improve supply chains, logistic flows and manufacturing continuity and conversion.

With this representative literature survey, it is shown that business continuity plan has been a research interest, and the study is taking larger dimension as supply chain spreads over the globe. The high disruptions with international ramifications are gradually factoring into continuity plan. At the advent of global lockdown in response to pandemic COVID-19, the research focus is changing to immunity to make the system resistant to adverse impacts. In this paper, we visualize a corporate business as a partially ordered workflows. In view of COVID-19, contact vulnerability of a business workflow is found to stress the points that need care for increase in business immunity for DRR: disaster risk reduction [7].

4 Contact Vulnerability

The activities of a business process or a supply chain can be modelled as a network
of workflows each involving an item(s) to handle for the actions scheduled at the
nodes of the flow. In the context of the present discussion on spread of infection due
to social contacts, let us call a node of workflow as a *workspot*. The idea of immunity
to infection as happens to living bodies is considered here. It decreases on the basis
of the body's exposition to infecting agents. Before quantifying the loss of immunity,
let us present the factors that cause the decrease.

- More one is connected there is more chance in loosing immunity.
- Loss of immunity is directly proportional to number of items handled.
- If more hands are on an item, then more is its chance for infection.

At a workspot, the reduction of immunity due to its structural and functional prop-
erties can be computed as

$$I_n = \left(\alpha \times \frac{1}{c}\right) + \left(\beta \times \frac{1}{t}\right) + \left(\gamma \times \frac{1}{h}\right) \qquad (1)$$

where I_n is immunity reduction at a node n, c is the degree of the node, and t and h
are the number of items passed through the node and the number of hands operated
at the node, respectively. The weights given to three factors are α, β and γ each
assuming a value from $(0, 1)$ where $\sum(\alpha, \beta, \gamma) \leq 1$. Then possibility of infection
at a workspot n can be computed as

$$f_n = 1 - I_n \qquad (2)$$

   Now having an idea on computing the infection possibility at a node, we will
discuss on the role of each workflow pattern towards the computation. While the
$eqn - 2$ computes the infection possibility of a node, i.e. the workspot, there is a
need of computing the same at each node considering the workings at the node as
while working there is a chance of contact and spread of an infection. For each
type of work patterns, below we present the resulting infection. A workflow pattern
is a building block in making of a workflow. The patterns considered here include
sequential, $n - split$, $n - merge$, $n - parallel$, and $n - store$.
   A *Sequential* workflow in Fig. 1 shows an activity at node q just after the activity
at p, so the flow is the sequential in nature. This has caused social contacts between
p and q, and the resulting possibilities of infection at both the nodes due to this
transaction are the maximum of their possibilities of infection. Say the infection
possibilities at p and q are f_p and f_q, respectively. Then, the resulting one at q is
increased by half of the difference between their infection possibilities $iff(f_p > f_q)then$

$$f_q = f_q + \frac{1}{2}(f_p - f_q) \qquad (3)$$

Fig. 1 Sequential

Fig. 2 n-split
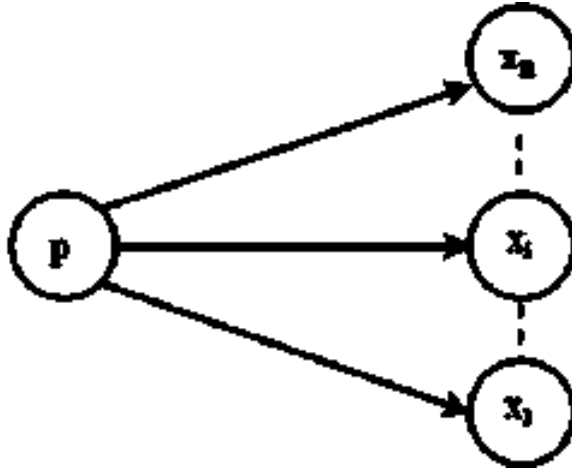
Rather than just half, the spread of infection can be modelled as a spread ratio with varying values depending on spread. The second one is the n-split workflow pattern in Fig. 2 that shows a flow from the node p to n number of nodes. That means n people may come in contact with the workspot p each having chance of getting infection from the node p. Now, the infection at any node say x_i among the n nodes is as below $iff (f_p > f_{x_i})$ then

$$f_{x_i} = \left(f_{x_i} + \frac{1}{2}(f_p - f_{x_i}) \right) \forall x_{i=1}^n \tag{4}$$

The possibility of an infection at a destination node is increased by the contribution of the source of split node. The third one is the n-merge workflow pattern as shown in Fig. 3. This shows at a node p there are n number of activities meet. The contacts due to the activities at n nodes preceding to p, let us say, equally contribute to the infections at the node p. This convergence of workflows at p results to increase in the possibility of infection as $iff (f_{x_i} > f_p)$

$$f_p = f_p + \sum_{i=1}^n \frac{1}{2}(f_{x_i} - f_p) \tag{5}$$

A typical workflow with two concurrent transactions is shown in Fig. 4. The first top workflow moves from workspot p to q where below the second transaction moves from the workspot r to s. Before transactions reach their scheduled workspots,
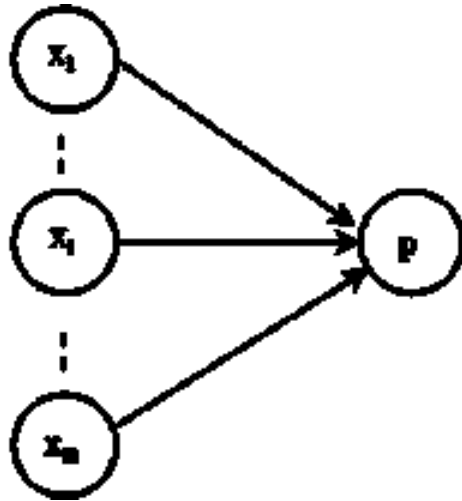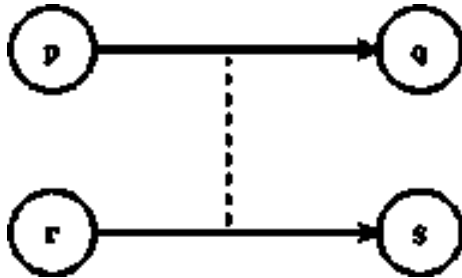
Fig. 3 n-merge

Fig. 4 Concurrent

Fig. 5 On-spot-n

respectively, at q and s, there is a social contact and for that the changed infection possibility counts at their destinations are

$$f_q = f_q + \frac{1}{2}((f_r - f_q) + (f_p - f_q)), \; f_s = f_s + \frac{1}{2}((f_p - f_s) + (f_r - f_s)) \quad (6)$$

when only $f_r > f_q$, $f_p > f_q$, $f_p > f_s$ and $f_r > f_s$.

The fifth workflow pattern, as shown in Fig. 5, is on-spot-n to model the same activities being done at a workspot for n times resulting equal number of social contacts. In

view of the infection possibility computation, the case has similarity with n-merge pattern as there is accumulation of the infection possibility at each contact. In n-merge case, contributing workspots are different, while in case of on-spot-n it is the same workspot. In the later case, the computation is

$$f_q^i = \left(f_q^{i-1} + \alpha^i \left(f_q^{i-1}\right)\right) \forall i = 1...n \tag{7}$$

where α^i is a measure of enabling infection spread at a workspot. It may be noted that $\alpha^i \in (0, 1)$ may take either increasing or decreasing value based on trend of infection at workspot during on-spot-n activities. It can also be noted that the value for f_q^0 can be obtained by applying the equation 1. Thus, a scheme for computation of contact vulnerability for five kinds of workspots is proposed here.

On computation of contact vulnerability, the workspot with the highest risk is identified and sandboxed. The business process in that workspot can be further changed for better continuity of the business.

5 Conclusion

Social contact is inevitable for doing a business. For this, there is always a possibility of infections in a business. For extremely infectious diseases like COVID-19, while breaking a spread chain is essential at the same time for it closing business process could be economically disastrous. Considering the importance of sandboxing of vulnerable workspots, this paper identifies five basic workflow patterns that usually occur in modelling a business process. Assuming a workspot having one of these flow patterns, social contact vulnerability of the spot due to the pattern can be computed as discussed in the paper. It is shown that a workspot assumes contact vulnerability due to the characteristics as discussed in the equation 1. This is based on the degree of connectivity, number of items handled at the workspot and the number of people participate in it. These together lower the immunity of a workspot making it vulnerable to infection. Learning α, β, γ and spread ratio would be our future work. A spread of infection from one workspot to another is modelled due to each workflow pattern. Because of execution of a flow pattern, the resulting possibility of infection is computed. This quantification of possible infection at a workspot gives an idea on a workspot vulnerability to a pandemic spread due to social contracts. Based on this possibility of infection quantification, the sandboxing of a vulnerable workspot can be carried out. The process of sandboxing may include the protocol of social distancing and other measures as suggested by the concerned administration. The proposed idea can be extended to model dynamic behaviour of a business process when it decided to continue with business as well as the fight against pandemic. Further, it is interesting to model such a system for different domains like electronic commerce, municipality services and scientific laboratories.

References

1. Lam W (2002) Ensuring business continuity. IT Prof (May–June):19–25
2. Cerullo V, Cerullo MJ (2004) Business continuity planning: a comprehensive approach. Inf Syst Manag 23(3):70–78
3. Pradabwong J, Braziotis C, Pawar KS, Tannock J (2015) Business process management and supply chain collaboration: a critical comparison. Logist Res 8(6):2015. https://doi.org/10.1007/s12159-015-0123-6
4. Rowan NJ, Laffey JG (2020) Challenges and solutions for addressing critical shortage of supply chain for Personal and Protective Equipment (PPE) arising from Coronavirus disease (COVID-19) pandemic case study from the Republic of Ireland. Sci Total Environ 725:138532
5. Ahmed M (2020) Towards 'gresilient' supply chain management: a quantitative study. Resour, Conserv Recycl 155:104641. https://doi.org/10.1016/j.resconrec.2019.104641
6. Labarthe O, Ferrarini A, Espinasse B (2006) Multi-agent modelling for simulation of customer-centric supply chain. Int J Simul Process Model 2(3/4). https://doi.org/10.1504/IJSPM.2006.012643
7. Aitsi-Selmi A, Egawa S, Sasaki H, Wannous C, Murray V (2015) The Sendai framework for disaster risk reduction: renewing the global commitment to people's resilience, health, and well-being. Int J Disaster Risk Sci 6(2):164–76
8. Djalante R, Shaw R, DeWit A (2020) Building resilience against biological hazards and pandemics: COVID-19 and its implications for the Sendai Framework. Prog Disaster Sci 6:100080. https://doi.org/10.1016/j.pdisas.2020.100080
9. Ding W, Levine R, Lin C, Xie W (2020) Corporate immunity to the COVID-19 pandemic, Posted on 17th April 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3578585
10. Margherita A, Heikkilä M (2021) Business continuity in the COVID-19 emergency: a framework of actions undertaken by world-leading companies. Bus Horiz 64(5):683–695, ISSN 0007-6813, https://doi.org/10.1016/j.bushor.2021.02.020

Electronic Supply Chain: A Bibliographic and Descriptive Literature Review

Susmita Bandyopadhyay and Joy Saha

Abstract There is an ever-increasing awareness and use of electronic supply chain (eSC) as evident from the existing literature. The reason for such growing interest in eSC lies in the increased complexity and continuously increasing demands for more and more quality service and automation. The existing literature shows not-so-significant number of research studies in eSC on the varieties of aspects of it. However, the need for a review of what has been done in this field is a necessity in order to show what are to be done. This paper, thus, reviews the existing literature on the various aspects of eSC in a classified fashion. The paper also shows the gap of research in the existing literature so that the researchers can endeavor to fill the identified research gap.

Keywords Electronic supply chain · Diffusion · Eco-efficiency · Resilience · Risk · Performance

1 Introduction

Discussion on electronic supply chain (eSC) has become a very demanding topic as evident from the increasing number of research studies in the existing literature. The advancement of the latest technologies and the global demand for each of the products starting from the agricultural products to the electronic products are leading to the ever-increasing need to implement eSC. Thus, the researchers all over the world have investigated various aspects of eSC such as the following—eco-efficiency for eSC, diffusion for eSC, risk and resilience of eSC, performance management for eSC, various factors affecting the performance of eSC, impact of eSC, implementation of eSC, technologies related to eSC, integration for eSC, and some more.

S. Bandyopadhyay (✉) · J. Saha
Department of Business Administration, University of Burdwan, Purba Bardhaman, West
Bengal 713104, India
e-mail: bandyopadhyaysusmita2010@gmail.com

Eco-efficiency is basically "a management strategy of doing more with less. It is based on the concept of creating more goods and services while using fewer resources and creating less waste and pollution [1]." Eco-efficiency, therefore, takes the environmental effects into account and based on that finds the value of a product. The purposes are manifolds such as the following—minimization of resources consumed, maximization of the use of renewable energy, maximization of use of some innovative materials, use of technologies to decrease consumption of energy, emphasizing on green suppliers, emphasizing more on green procurement, design for environmental factors. Supply chain integration for eSC is another frequently studied topic as evident from the existing literature. The various factors of eSC integration include the following [2]—collaborative planning, competitive capability, long-term relationship both with the suppliers and the customers, dependence and trust between customer and supplier, eSupply chain planning and trust, eSC relationships, shared IT infrastructure, adoption of different latest technologies, inter-organizational communications, people involvement.

The third frequently discussed topic is eSC diffusion. Diffusion can be defined as "adoption of an internally generated or purchased device, system, policy, program, process, product or service that is new to the adopting organization [3]." eSC management diffusion basically consists of three stages—adoption, implementation, and assimilation. eSCM diffusion can be internal or external. Internal diffusion is between different departments inside an organization and external diffusion is between different trading partners of eSC. The fourth discussed topic is the performance of eSC. The performance of eSC can be measured in varieties of ways. One of the different aspects of performance measurement includes the following. These measuring factors are especially essential for eSC diffusion as well. Customer response time, on-time delivery, extent of sharing of order information, skills of employees, accessibility of different information, identifying market innovation opportunities, transportation tool utilization, profitability, revenue growth, cost structure, product leadership, customer relationship, firm image, management support, delivery management, innovation management, sourcing leadership, collaboration. The risk and resilience against the risk for eSC are another very important topic of discussion. Resilience depends on the extent of vulnerability of the eSC. Both risk and resilience are uncertain in nature. The risk management is related to enterprise risk management, crisis management, and risk measurement procedure as well. There can be many strategies for eSC resilience which include the following [4]—fine-tuning of supply chain design, strengthening of supply base, flexibility of supply contract, dynamic assortment planning, centralizing demand, capacity flexibility, process standardization, agile operations, manufacturing flexibility, risk hedging, visibility enhancement, cross-training of employees, product flexibility via postponement, range of products, inventory flexibility, logistics flexibility, dynamic revenue management, proliferating customer accounts, responsive pricing strategies, and some more.

Besides, the antecedents and consequences of the implementation of eSC have also drawn significant attention from the researchers. The basic advantage of using Internet for eSC include—quick response time, quick access to information,

improved customer services, increased competitiveness, reduction of data and data entry, availability of vertical information at low cost, real-time communication, easier search, ease to locate required information and product, saving time, saving money, easy accessibility, availability of information for taking purchasing and ordering decisions for the customers, ability of the customers to screen and select from among wide range of available options. However, there are many more topics related to eSC which can be discussed.

2 Bibliographic Review

The relevant research papers for the review as presented in this paper have been downloaded from different journal databases such as Elsevier, Springer, Taylor and Francis, IEEE, Emerald, ACM digital library, Inderscience, and IGI Global. The research papers have been searched with the keyword "electronic supply chain." A total of 184 research papers were found from Elsevier; 56 research papers from Springer; 19 research papers from IEEE; 11 research papers from ACM; 253 research papers from Emerald; 60 research papers from Inderscience; 79 research papers from Taylor and Francis; two research papers from IGI Global; and 12 research papers from other journals, from the search by the above-mentioned keyword. This is depicted in gist through Fig. 1. Figure 1 shows that Emerald shows the highest number of papers as obtained out of the search. These downloaded articles sets contain research papers directly related to eSC or eSCM or on some related fields such as e-retailing, e-marketplace, e-business, and so on. Thus, the papers exclusively on electronic supply chain had to be sorted and the results of the papers selected are shown in Fig. 2. Figure 2 shows that the number of articles exclusively on "electronic supply chain" is the highest for Elsevier (= 30) unlike that for Fig. 1. The second highest number is obtained from Emerald (= 20), and the third highest number is obtained for IEEE (= 18). This is consistent with the type of these journal databases since Elsevier is based on Science journals and IEEE is based on Electrical and Electronics Engineering, whereas Emerald is based on both Science and Management although other journals of other disciplines are also present in Elsevier and Emerald in comparatively smaller in number. However, both the research papers and the journals are generally evaluated based on the impact of citation, and thus, analysis of citations of these articles has also been performed in order to identify more demanding articles among all these downloaded articles. Figure 3 shows a graph indicating the citation extent.

Figure 3 shows that the number of citations is the highest for Elsevier, followed by Emerald which is followed by IEEE. This is consistent with the fact that the keyword "electronic supply chain" is basically a scientific keyword and concept. The final analysis has been done for the topic of the articles which is the main focus of any review paper. A large number of topics have been identified as shown in Table 1. The letters within brackets in Table 1 represent the tag for the topics so that the graph to be shown in Fig. 4 becomes easier to understand and explain. Thus, Fig. 4 shows that the highest number of articles is on the impact of eSC, represented by the tag "h."

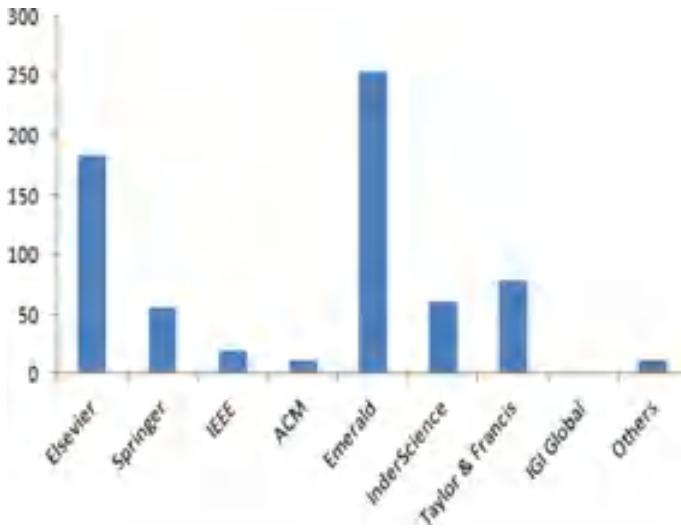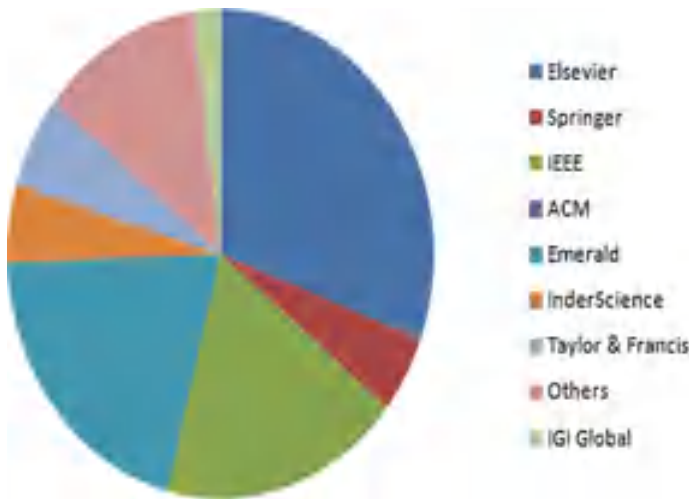**Fig. 1** Total number of journal articles from different databases

Fig. 2 Number of selected journal articles from different databases

Two others with the second highest number of articles are "Factors affecting eSC (a)" and "eSC performance (i)." Such result clearly shows the current main focus of the researchers—impact of eSC, factors affecting eSC, and eSC performance. Some others also have larger areas as compared to the others, such as eSC risk (x) and supplier selection (y). However, since the eSC risk and eSC resilience are closely interrelated topics, they could also be put together which would result in higher share
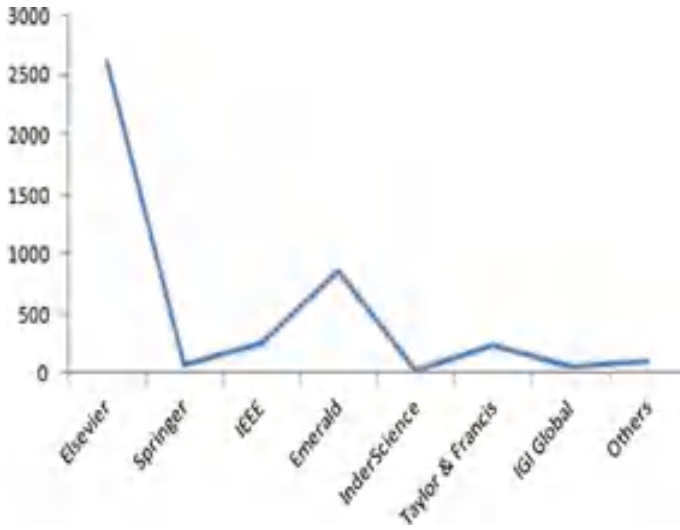
Fig. 3 Citations of the selected downloaded articles

of this combined topic. However, the basic target of any literature review is to identify the gap of research. Therefore, the following section analyzes the topics in order to find the research gap in the most frequently discussed topics.

The above-described papers include some of the research studies like the research studies of Lin [1] (on eSC diffusion), Siddiqui et al. [2] (review of eSC for the period

Table 1 Topics as identified

|	
• Factors affecting eSC (a)	• Impact of eSC (h)
• Design of eSC (b)	• eSC performance (i)
• Decision support for eSC (c)	• eSC diffusion (j)
• Eco-efficiency (d)	• Information system selection for eSC (k)
• Design of eSC (e)	• Selecting eSC software (l)
• SOA adoption for eSC (f)	• Scheduling for eSC (m)
• Benefits of eSC (g)	• Waste discharge for eSC (n)
• Simulation for eSC (o)	• eSC implementation (v)
• eSC technology (p)	• eSC resilience (w)
• Optimization for eSC (q)	• eSC risk (x)
• Application of eSC (r)	• Supplier selection for eSC (y)
• eSC partnership (s)	• eSC integration (z)
• Framework for eSCM (t)	• Analysis of eSC (aa)
• eSC replenishment (u)	
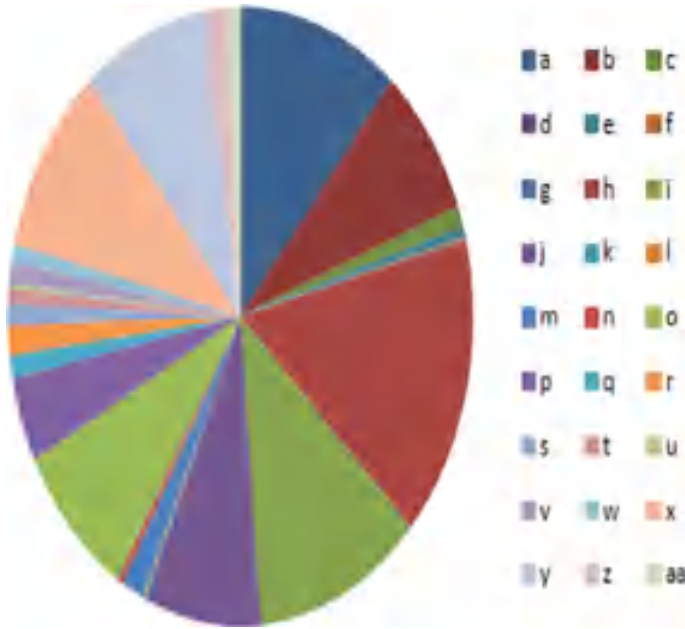
Fig. 4 Number of research papers under each topic

2000–2015), Shamout and Emeagwali [3] (review on impact of eSC on customer satisfaction), Nguyen and Harrison [4] (on developing a framework for integrated electronic supply chain management), Khan et al. [5] (a survey on the adoption of eSC in Pakistan), Fendt [6] (on eSC collaboration in China), Rajesh et al. [7] (risk mitigation strategy for eSC), Bolhari [8] (on information security for eSC), Tang et al. [9] (on development of a simulator for eSC), Chowdhury et al. [10] (on challenges related to the use of RFID for eSC), Sugara and Aziz [11] (on application of eSC on retail industry), Rana and Sharma [12] (on performance measure of eSC), Li and Zhou [13] (on the construction of service eSC), Chakravorti and Erin [14] (on payment method for eSC), Leonard and Cronan [15] (on online retailing for eSC), Al-Hakim [16] (on interdependencies for eSCM), Chen and Kong [17] (on implementation of eSC), Ahmi et al. [18] (on a bibliometric analysis of eSCM), Vlachos et al. [19] (on eSCM tools), Devaraja [20] (on model development of eSCM), Diprose et al. [21] (on challenges of voluntary supply chain governance), Pulevska-Ivanovska and Kaleshovska [22] (on the implementation of eSCM), and some more.

Besides, some of the research studies which have especially required for the bibliographic review include the following—work of Barutçu and Tunca [5] on analysis of the effect of eSC on e-retailing based on different factors; works of Williams et al. [6] (investigated the effects of electronic supply chain management (eSCM) on the partnership and strategic alliances), Almajali et al. [7] (investigated the effect of eSCM on the performance of manufacturing activities in Jordon), Rezayat et al. [8] (investigated the effect of Government intervention on electronic closed-loop supply

chain), Kim and Im [9] (developed a model that had shown the effects of electronic supply chain design on the customization capability for the companies), Korea. Yeh [10] (investigated the effect of the factors on the cooperative eSC). Liu et al. [11] (investigated the effects of power and trust on the willingness to adopt eSCM in China), Ke et al. [12] (analyzed the effects of organizational culture on the adoption of eSC), Tyagi et al. [13] (analyzed the performance of eSCM. The purpose of this research was to improve the performance of eSCM related to Indian automobile industry), Caputo et al. [14] (analyzed the factors for measuring the performance of eSCM) and some more.

3 Conclusion

This paper has presented a brief bibliographic and descriptive literature review on electronic supply chain. Various aspects for bibliographic review have been presented such as total number of journal articles from various databases, number of citations. Descriptive review has also been presented along with the bibliographic review. The reader is expected to get a good glimpse of the overview research studies that have been conducted in this area of study.

References

1. Lin HF (2017) Antecedents and consequences of electronic supply chain management diffusion: the moderating effect of knowledge sharing. Int J Logist Manag 28(2):699–718
2. Siddiqui AW, Raza SA (2015) Electronic supply chains: status & perspective. Comput Ind Eng 88:536–556
3. Shamout MD, Emeagwali OL (2016) Examining the impact of electronic supply chain management processes on customer satisfaction: a literature review. Bus Econ Horiz (BEH) 12:141–163, (1232-2017-2395)
4. Nguyen HM, Harrison NJ (2004) Electronic supply-chain orientation and its competitive dimensions. Prod Plan Control 15(6):596–607
5. Khan SA, Liang Y, Shahzad S (2014) Adoption of electronic supply chain management and e-commerce by small and medium enterprises and their performance: a survey of SMEs in Pakistan. Am J Ind Bus Manag 4(9):433–444
6. Fendt TC (2010) Introducing electronic supply chain collaboration in China: evidence from manufacturing industries, vol 15, Univerlagtuberlin
7. Rajesh R, Ravi V, Venkata Rao R (2015) Selection of risk mitigation strategy in electronic supply chains using grey theory and digraph-matrix approaches. Int J Prod Res 53(1):238–257
8. Bolhari A (2009) Electronic-supply chain information security: a framework for information. Australian Information Security Management Conference, Security Research Center, Edith Cowan University, Western Australia, Wellington, New Zealand, 20–21 Jan
9. Tang NK, Benton H, Love D, Albores P, Ball P, MacBryde J, Boughton N, Drake P (2004) Developing an enterprise simulator to support electronic supply-chain management for B2B electronic business. Prod Plan Control 15(6):572–583
10. Chowdhury B, Chowdhury MU, D'Souza C (2008) Challenges relating to RFID implementation within the electronic supply chain management—a practical approach. In: Software

engineering, artificial intelligence, networking and parallel/distributed computing, 6–8 Aug. Springer, Berlin, Heidelberg, pp 49–59

11. Sugara AA, Azis AM (2020) Electronic supply chain management application analysis in retail industry. Int J Bus Technol Manag 2(2):45–51
12. Rana K, Sharma SK (2022) Supply chain performance in the era of digitisation: the role of electronic supply chain management. Int J Logist Syst Manag 41(4):503–536
13. Li X, Zhou Y (2001) Construction of elderly service electronic supply chain. In: The 2021 12th International conference on E-business, management and economics, 17–19 July, Beijing, China, pp 139–143
14. Chakravorti S, Erin D (2004) An electronic supply chain: will payments follow. Chicago Fed Letter, Special Issue 206a
15. Leonard LN, Cronan TP (2003) Website retailing: electronic supply chain replenishment. J Organ End User Comput (JOEUC) 15(3):45–55
16. Al-Hakim L (2005) Modelling interdependencies for electronic supply chain management. In: Proceedings of the information resources management association international conference (IRMA 2005), 15–18 May, USA, IGI Publishing (IGI Global), pp 1002–1006
17. Chen H, Kong L (2007) An analysis on the implementation of electronic supply chain in international trade. Integr Innov Orient E-Society 1:115–125
18. Ahmi A, Rahim SA, Elbardan H (2018) A global trend of the electronic supply chain management (e-SCM) research: a bibliometric analysis. Int J Supply Chain Manag 7(5):535–542
19. Vlachos IP, Skoumpopoulou D, Gutnik S (2014) Electronic supply chain management tools in international business: evidence from Austrian food clusters. E-commerce Platform Acceptance, Springer, Cham, pp 3–21
20. Devaraja TS (2011) Electronic supply chain management and model development–global perspective. Working paper, Ministry of Human Resource Development, Government of India, New Delhi
21. Diprose R, Kumiawan N, Macdonald K, Winanti P (2022) Regulating sustainable minerals in electronics supply chains: local power struggles and the 'hidden costs' of global tin supply chain governance. Rev Int Polit Econ 29(3):792–817
22. Pulevska-Ivanovska L, Kaleshovska N (2013) Implementation of e-supply chain management. TEM J 2(4):314
23. Barutçu S, Tunca MZ (2012) The impacts of E-SCM on the E-tailing industry: an analysis from Porter's Five Force perspectives. Procedia Soc Behav Sci 58:1047–1056
24. Williams LR, Esper TL, Ozment J (2002) The electronic supply chain: Its impact on the current and future structure of strategic alliances, partnerships and logistics leadership. Int J Phys Distrib Logist Manag 32(8):703–719
25. Almajali D, Mansour K, Maqableh M (2016) The impact of electronic supply chain management usage on firm's performance. Int J Commun Netw Syst Sci 9(6):280–293
26. Rezayat MR, Yaghoubi S, Fander A (2021) The impact of government intervention in competitive electronic closed-loop supply chain to support internal industry. Resour Policy 74:102257
27. Kim KC, Im I (2002) The effects of electronic supply chain design (e-SCD) on coordination and knowledge sharing: an empirical investigation. In: Proceedings of the 35th annual Hawaii international conference on system sciences, Hawaii, January 10, IEEE, pp 2149–2158
28. Yeh YP (2005) Identification of factors affecting continuity of cooperative electronic supply chain relationships: empirical case of the Taiwanese motor industry. Supply Chain Manag: An Int J 10(4):327–335
29. Liu H, Ke W, Wei KK, Hua Z (2015) Influence of power and trust on the intention to adopt electronic supply chain management in China. Int J Prod Res 53(1):70–87
30. Ke W, Liu H, Wei KK, Gu J, Chen H (2006) The effects of relational and institutional factors on electronic supply chain management adoption: does organizational culture matter? In: PACIS 2006 Proceedings 83, Malaysia, 6–9 July

31. Tyagi M, Kumar P, Kumar D (2014) A hybrid approach using AHP-TOPSIS for analyzing e-SCM performance. Procedia Eng 97:2195–2203
32. Caputo AC, Cucchiella F, Fratocchi L, Pelagagge PM, Scacchia F (2004) Analysis and evaluation of e-supply chain performances. Ind Manag Data Syst 104(7):546–557

Synthesizing Realistic ARMD Fundus Images Using Generative Adversarial Networks (GANs)

Sesha Sai Aneeswar Kalisapudi, Vavilala Divya Raj ⓘ**, Shubhasri Vanam** ⓘ**, and Jasvith Chand Anne**

Abstract Age-related macular degeneration (ARMD) is an eye disease that can result in blurred or no vision in the central vision. It happens when aging causes damage to the macula and is the main reason for loss of sight for older adults. Deep learning techniques are widely used in ophthalmology, such as diagnosing age-related macular degeneration (ARMD), which requires a huge image dataset. But the existing datasets are unclear and insufficient for building the training models and require more preprocessing time. Also, Indian datasets are not used prominently for the leading causes of blindness and eye diseases. Therefore, this paper emphasizes on synthesizing large new datasets of artificial retinal images from the existing datasets by a deep learning approach generative adversarial networks which are the GANs. Generative adversarial networks (GANs) will be trained with fundus images from the age-related eye disease study (AREDS), producing synthetic fundus images with the ARMD. The performance of ARMD diagnostic DCNNs will be trained on the combination of both real and synthetic datasets. Images obtained by using generative adversarial networks (GANs) appear to be realistic and also increase the precision of the model. The deep learning model's performance which uses the synthesized dataset should be close to the real images, suggesting that the dataset can be utilized for training humans and machines.

Keywords Age-related macular degeneration · Generative adversarial networks · Deep convolutional neural networks

1 Introduction

Vision is considered as most important senses, and losing it can have a significant effect on one's independence and productivity. Millions of people are affected by retinal diseases, which can cause blindness if they are not detected and treated at an

S. S. A. Kalisapudi · V. D. Raj (✉) · S. Vanam · J. C. Anne
Secunderabad, India
e-mail: divyaraj.vavilalai@gnits.ac.in

early stage. The retinal problems like glaucoma, diabetic retinopathy, and age-related macular degeneration may be cured or slowed down in its progression with early treatment choices. ARMD is a painless, permanent, degenerative eye problem related to the deterioration and definitive death of photoreceptors. With the availability of new therapies, early detection of patients who have a risk for ARMD will be increasingly important. Although ARMD doesn't result in total blindness, it does cause central vision loss, which makes it more complex to carry out close-up chores such as home maintenance and cooking [1, 2].

With advances in image analysis and machine learning technologies, it is now possible to automatically identify retinal diseases and recommend patients for additional medical advice. Recent machine learning research focuses on image classification after feature extraction to diagnose conditions such as ARMD [3, 4]. However, due to a small dataset and poor feature extraction, these encounter difficulties in disease prediction. In order to address these issues, the main target of this paper is the development of the system that would expand the dataset through data augmentation, improving the retinal images accuracy [5, 6]. Analyzing method of retinal images and their clinical implications for the 2-D fundus imaging and methods for 3-D optical coherence tomography (OCT) imaging are very effectively defined for the identification of optical diseases [7, 8].

Various traditional data augmentation algorithms and other advanced algorithms like generative adversarial networks (GAN), CNN, and ResNet are used to generate synthetic images given a small dataset. It is possible to create modified images using conventional data augmentation techniques, but not entirely new ones. Hence, there is a need to implement algorithms that generate new images with more accuracy, and therefore, the generative adversarial networks (GANs) are proposed in this study [9].

The proposed methodology uses GAN to create synthetic images from the training dataset of ARMD fundus images [7, 10, 11].

Generative adversarial networks (GANs) consist of a generator and a discriminator, and their training is done iteratively by the means of adversarial learning. As a strong class of generative models, production of new samples is learnt by the GANs that adhere to the same distribution just like the original samples rather than explicitly estimating the distribution of data samples. Being able to produce innumerable new samples out of a potential distribution has big utilization value in various areas for example image and vision computing, speech and language processing [12, 13].

1.1 Generative Adversarial Networks (GANs)

Generative adversarial networks or GANs are a technique of generative modeling that utilizes deep learning techniques such as convolutional neural networks. In the generative modeling, that is a sort of machine learning, regularities or the patterns in input data are identified in an automated way and learned in order that the model could be used for the production of new samples that might have been correctly deduced from the original dataset. Generative adversarial networks (GANs) are utilized to
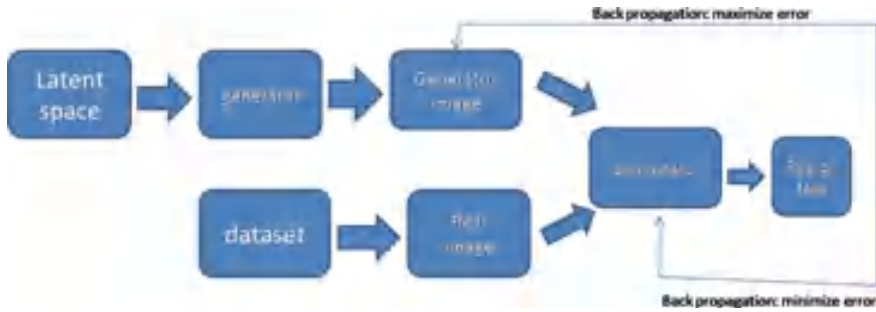
Fig. 1 Architecture of GAN

produce different types of data that resembles what humans make, including text, photographs, music, and speech. By defining the challenge as a supervised learning problem, GANs provide an innovative way to train a generative model along with two sub-models which are the generator model, that is trained for the production of new examples and the second one is the discriminator model, which aims in the classification of samples as either real (originating from the domain) or synthetic (generated). The GAN can be implemented as a game-theory-style problem the minimax problem. The following steps illustrate the creation of this type of problem:

- **Generator goal**: Maximizing the probability which the discriminator mischaracterizes its output as the real one.
- **Discriminator goal**: Optimizing toward an objective of 0.5, in which the discriminator cannot separate between real and synthetic images.

The generative adversarial network's architecture that displays the fundamental operations in block format is represented in Fig. 1. Discriminator is trained using 100% real and 100% fake images. Generator is trained using discriminator loss. After training the discriminator, it classifies both real and artificial or synthetic images. Error of the discriminator is back propagated, and the generator model is updated.

2 Image Synthesis Using Generative Adversarial Networks

Two components make up a generative adversarial network (GAN):

The generator gains the ability to produce images that are plausible. The created instances serve as negative training examples for the discriminator.

The discriminator gains the ability to differentiate real image from synthetic image generated by the generator which is fined by the discriminator when it generates unlikely results.

The discriminator determines if each instance of image that it analyzes is actually a part of the training dataset, whereas the generator, creates new image cases and the discriminator, assesses them for legitimacy.

The two main constituents of a GAN—generator neural network and discriminator neural network are described in Fig. 2.

The generator network requires a random input and then attempts to create a sample of data. In Fig. 2, a generator $G(z)$ is shown which takes an input z from $p(z)$. Z is a sample from probability distribution $p(z)$. It produces data which then enters into a discriminator network $D(x)$. Its responsibility is to receive input from the generator or from the real data and attempt to guess whether the input is synthesized or real. It is responsible for taking an input x from $p_{data}(x)$ where $p_{data}(x)$ is distribution of real data. $D(x)$ then helps in solving an issue of binary classification with the usage of sigmoid function rendering an output in the range 0 to 1. The notations which are utilized to formalize the GAN are

$p_{data}(x)$—the distribution of real data
x—sample from $p_{data}(x)$
$P(z)$—distribution of generator
z—sample from $p(z)$
$G(z)$—generator network
$D(x)$—discriminator network.

Now the GAN is trained as a conflict between generator and discriminator. This can be shown arithmetically as

$$\min_{G} \max_{D} V(D, G)$$

$$V(D, G) = E_{x \approx p_{data}(x)}[\log D(x)] + E_{z \approx p_z(z)}[\log(1 - D(G(z)))] \qquad (1)$$
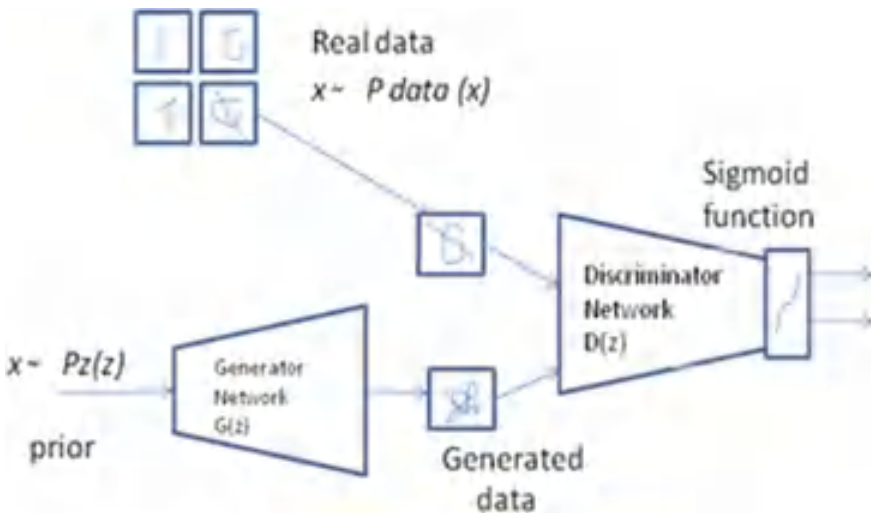

Fig. 2 Components of GAN

In Eq. (1), entropy is the first term in which the data from real distribution ($p_{\text{data}}(x)$) flows through the discriminator which attempts to increase it to 1. Entropy which refers to the second term is the one in which the data from random input ($p(z)$) flows through the generator, which then creates an artificial sample that is then transferred through the discriminator to detect the synthetic image. In entropy, the discriminator attempts in maximizing it to 0 (i.e., the log probability that the data produced is synthetic is equal to 0). So as a whole, the discriminator attempts in maximizing the function V. On the contrary, the generator's task is exactly opposite; i.e., it attempts to minimize the function V in order that the differentiation between real and synthetic data is bare minimum.

2.1 Training of the Generative Adversarial Networks Model

A training phase has two parts which are performed progressively.

Pass 1: Training the discriminator and freezing of the generator as shown in Fig. 3 (freezing refers to the setting training as false. The network is only responsible for forward pass and no backpropagation is used).

Pass 2: Train generator and freeze discriminator as shown in Fig. 4.

The dataset is initially divided into various batches. The GAN model is trained using these batches one after the other. First, the training of discriminator model is performed using purely real and purely fake images. Half or three-fourths of data in a particular batch is used for training the discriminator model. We train the discriminator model with 100% real examples first. Then we generate 100%
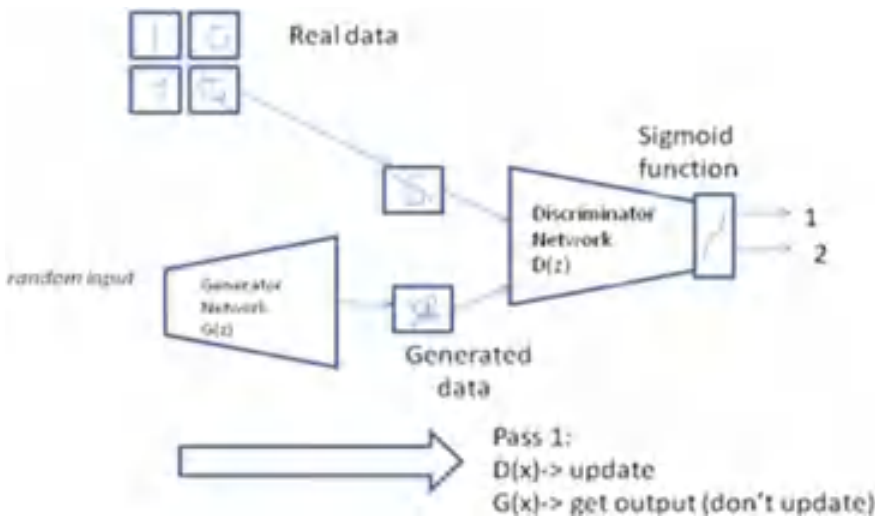

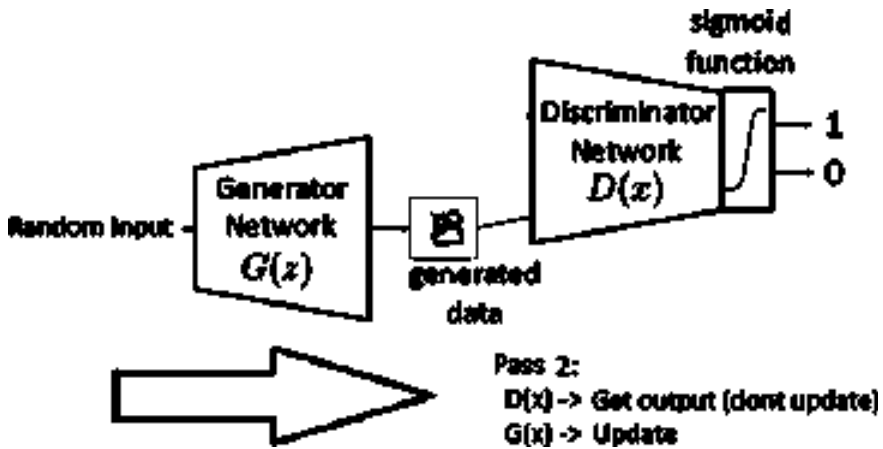
Fig. 3 Training discriminator

Fig. 4 Training generator

synthetic examples using the generator model and train the discriminator using these 100% synthetic samples. Now our discriminator is completely trained with both of the samples (real and synthetic). Next, calculate discriminator loss and update the weights of the generator model by backpropagating the error. We do this by inverting the target value of synthetic samples to 1 and training the generator model with these samples. We do this because the discriminator will separate the generated samples as not real. The backpropagation steps will see a huge error and will update the model weights accordingly. This results in making the generator model better at generating good samples. The methodology is depicted in Fig. 5.

After every ten epochs, a summary of the model is displayed. Both real and synthetic image accuracy is included in the model's summary. Accuracy is analyzed, and the correct number of epochs is decided. The accuracy of both actual and false samples needs to be compromised. We need to come to a point where accuracy of both real and synthetic images is close to 50%. This is the point where error rates of both generator and discriminator are pretty equal. We need to train our model until we reach this point, i.e., to the point where accuracy of both real and fake samples is close to 50%.

2.2 Steps Required for the Training of GAN

Step 1: **Defining the issue**. Do you wish in creating the fake text or fake images? Gather information in accordance with the problem.

Step 2: **Defining the architecture of GAN**. Defining the design of your GAN. Should we use convolutional neural networks or multilayer perceptrons for your generator and discriminator, respectively?
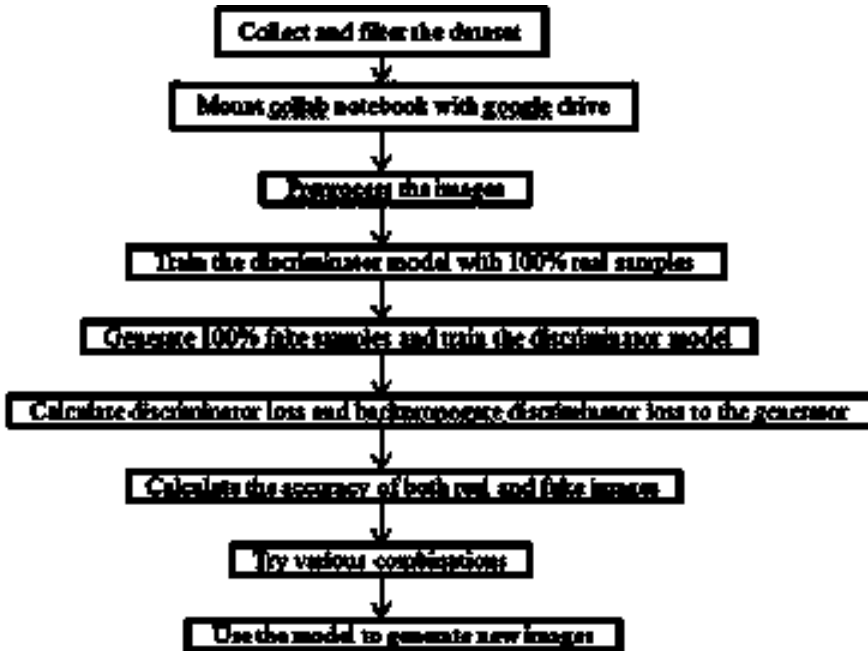
Fig. 5 Methodology

Step 3: **Training the Discriminator on real data for n epochs**. Getting the data for which you need to make fake data, and training the discriminator to determine it as real. Any natural number which is between 1 and infinity could be used here as value *n*.

Step 4: **Generating fake inputs for generator and training of the discriminator on fake data**. Obtaining the generated data and allowing the discriminator to determine them as fake correctly.

Step 5: **Training generator with the output of discriminator**. You can now utilize the predictions of discriminator as a target for training the generator once it has been trained. Creating a generator which can trick the discriminator.

Step 6: **Repeat step 3 to step 5 for a few epochs**.

Step 7: **Check if the fake data manually if it seems legit. If it seems appropriate, stop training, else go to step 3**.

3 Analysis of Results with Respect to Various Parameters

Various results can be generated by varying batch size, epochs, size of the training set. Observations are done and comparisons are made by the varying batch size, increasing epochs, varying size of training data, and considering different

combinations of all the above. Observations and conclusions for each are given below.

3.1   Comparing Images Generated by Varying Batch Size

From Table 1, we can conclude that it is feasible to consider batch size as 2 when compared to 1 as the batch size is increased the correctness of both real and synthetic images is approaching 50%.

Figure 6 is obtained by executing the code for 1000 epochs with batch size 1. Accuracy obtained for real images was 77%. Accuracy obtained for fake images was 10%.

Table 1 Comparison of accuracy of real and synthetic images with batch sizes 1 and 2

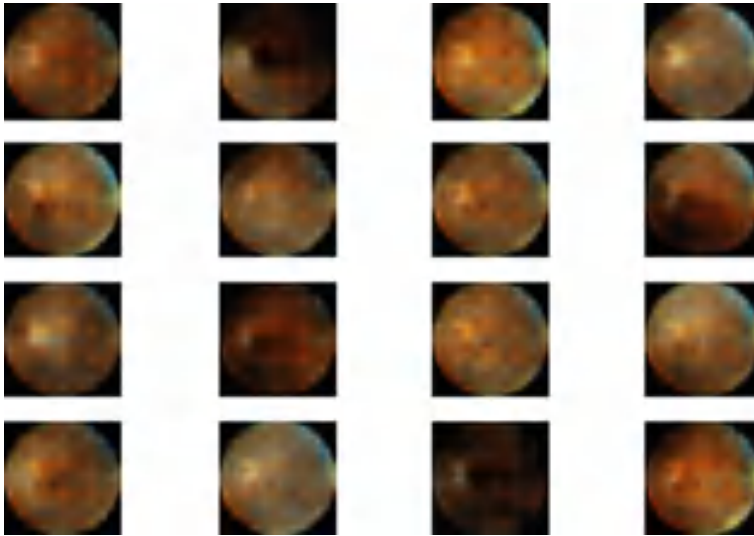| Number of epochs	Batch size	Accuracy of real images (%)	Accuracy of synthetic images (%)
1000	1	77	10
500	2	40	80

Fig. 6 1000 epochs with batch size 1(77–10)

3.2 Comparing Images Generated by Increasing Epochs

From Table 2, it can be concluded that by using more epochs the clarity of images is increasing. It is advisable not to consider too many epochs as this leads to overfitting and degrades the accuracy.

Figure 7 is obtained by executing the code for 100 epochs with batch size 1. Accuracy obtained for real images was 71%. Accuracy obtained for synthetic images was 100.

Table 2 Comparison of accuracy of real and synthetic images with varying number of epochs

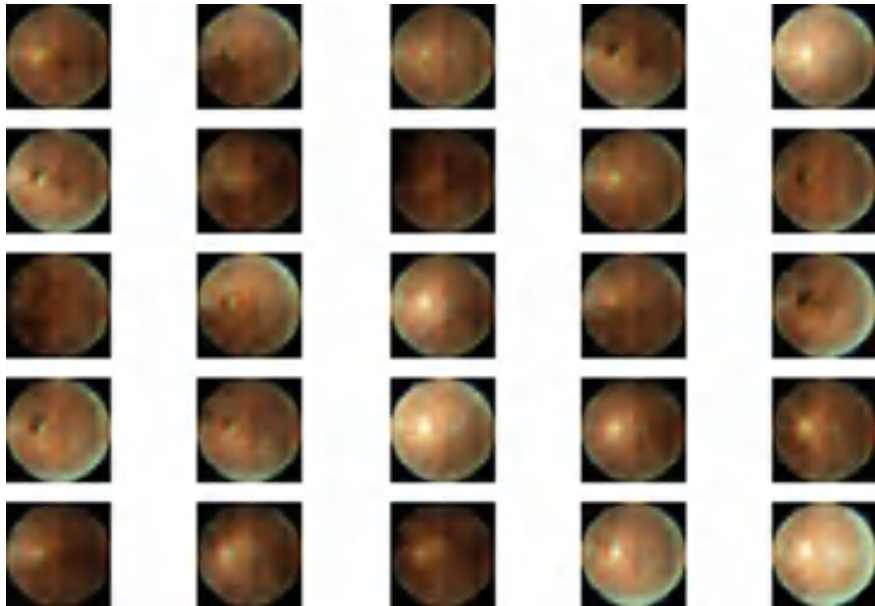| Number of epochs	Batch size	Accuracy of real images (%)	Accuracy of synthetic images (%)
100	1	71	100
1000	1	77	10

Fig. 7 100 epochs with batch size 1(71–100)

3.3 Comparing Images Generated by Varying Size of the Training Set

From Table 3, it might be concluded that it is better to include 50% data for training as images generated using 50% data as training are clearer when compared to that with 75% data.

Figure 8 is obtained by executing the code for 1000 epochs and by using half of the data for training. Accuracy obtained for real images was 77%. Accuracy obtained for synthetic images was 10%.

Table 3 Comparison of accuracy of real and fake images with varying training data

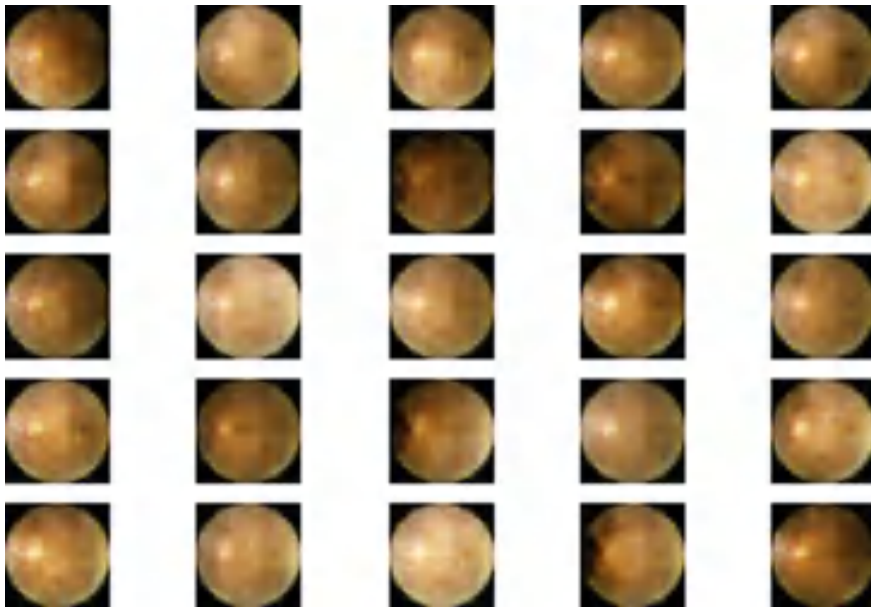| Number of epochs	Size of training set	Accuracy of real images (%)	Accuracy of synthetic images (%)
1000	1/2	77	10
1000	3/4	88	78

Fig. 8. 1000 epochs using 1/2 of data for training (77–10)

3.4   Comparing Images Generated by Various Combinations

From Table 4, it has been observed that the ideal accuracy observed for real to fake images is 50–50.

The images retrieved at epoch 410 have shown clear eyeball structure at 48–75 as shown in Fig. 9. So, it can be concluded that it is better to consider images whose accuracy of both real and fake images is close to 50%.

Table 4 Comparison of accuracy of real and synthetic images for various combinations

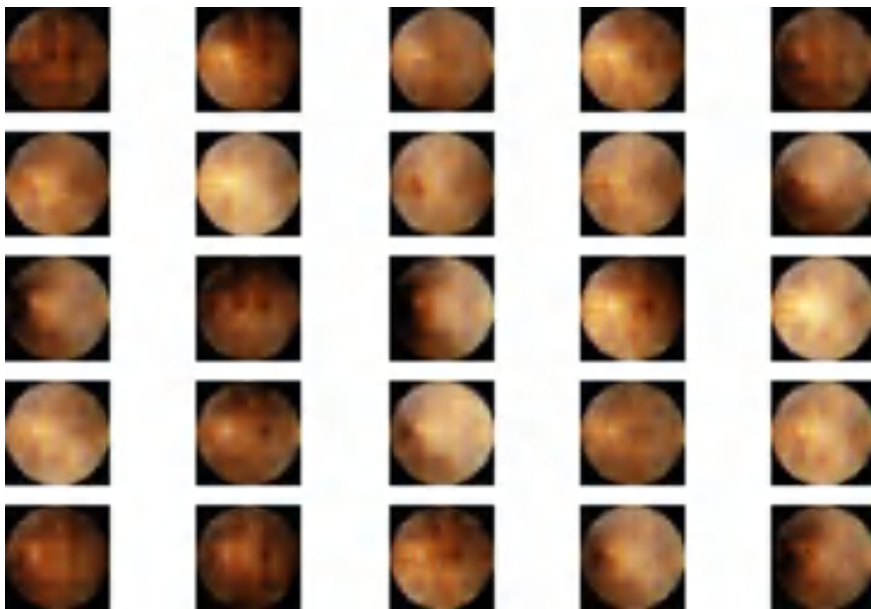| Number of epochs	Batch size	Size of training data	Accuracy of real images (%)	Accuracy of synthetic images (%)
1000	1	1/2	77	10
410	2	1/2	48	75
410	2	3/4	4	100

Fig. 9 Ideal images

4 Conclusions

Generative adversarial networks (GANs) constitute a generator and a discriminator that is trained iteratively in an adversarial learning way. As a strong class of generative models, GANs attempt in learning to produce new samples which adhere to the same distribution as the original samples rather than explicitly estimating the distribution of data samples. Being able to produce limitless new samples from potential distribution has a great utilization value in various fields for example image and vision computing, speech, and language processing.

Initially, the dataset has been examined and preprocessed accordingly. The discriminator and the generator have been trained until optimal results have been produced. At the end, the generative adversarial network produces 100 synthesized fundus images for every execution. Considering the comparisons made above on the categories of batch division, increase in epochs, size of the training set, and various combinations the optimal solution is obtained at 410 epochs using 1/2 of the training data with batch size as 2 with 48% as real and 75% as synthetic. The fundus images generated are with better visibility.

This work can be further extended by building a CNN model like VGG-16, ResNet, and LeNet-5 to test the performance of the generative adversarial networks (GANs) to create better fundus images.

References

1. Grzybowski A, Shtayer C, Schwartz SG et al (2021) The 100 most-cited papers on age-related macular degeneration: a bibliographic perspective BMJ Open Ophthalmol 6:e000823. https://doi.org/10.1136/bmjophth-2021-000823
2. https://www.kaggle.com/meaninglesslives/using-resnet50-pretrained-model-inkera
3. https://www.allaboutvision.com/conditions/blurry-vision.html
4. Sandhya M, Morampudi MK, Grandhe R, Kumari R, Banda C, Gonthina N (2022) Detection of diabetic retinopathy (DR) severity from fundus photographs: an ensemble approach using weighted average. Arab J Sci Eng pp 1–8. https://doi.org/10.1007/s13369-021-06381-1
5. https://pubmed.ncbi.nlm.nih.gov/30629091/
6. Cao Y et al (2019) Recent advances of generative adversarial networks in computer vision. IEEE Access 7:14985–15006. https://doi.org/10.1109/ACCESS.2018.2886814
7. https://analyticsindiamag.com/how-to-build-a-generative-adversarial-network-in-8-simple-steps/
8. Sravani Devi Y, Phani Kumar S (2021) Retinal image synthesis for diabetic retinopathy assessment using dcgan and vae models. Int J Adv Electron Comput Sci 8(10), ISSN (p): 2394-2835
9. Seetha M, Kalyani N, Sravani Devi Y (2022) An ensemble CNN model for identification of diabetic retinopathy eye disease. In: Proceedings of smart intelligent computing and applications, volume 2. smart innovation, systems and technologies, vol 283. Springer, Singapore. https://doi.org/10.1007/978-981-16-9705-0_19
10. https://realpython.com/generative-adversarial-networks/
11. Ramin S, Soheilian M, Habibi G, Ghazavi R, Gharebaghi R, Heidary F (2015) Age-related macular degeneration: a scientometric analysis. Med Hypothesis Discov Innov Ophthalmol 4(2):39–49. PMID: 26060829; PMCID: PMC4458325

12. Karras T, Laine S, Aila T (2019) A style-based generator architecture for gen-erative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2019
13. https://poloclub.github.io/ganlab/
14. Abramoff MD, Garvin MK, Sonka M (2014) Retinal imaging and image analysis. IEEE Rev Biomed Eng 3:169–208
15. https://neptune.ai/blog/generative-adversarial-networks-gan-applications

Interpretation and Assessment of Improved Deep Networks for the Classification of Glaucoma Using Explainable Grad-CAM Approach

Srikarthik Kalisapudi and Rohini Palanisamy

Abstract The human eye is a sensory organ that responds to light and pressure. Many illnesses, ailments, and age-related changes can harm the eyes and their surrounding region. Glaucoma is a chronic age-related eye disorder caused due to increase in intraocular fluid pressure. It results in optic nerve damage and irreversible vision loss if not diagnosed at an early stage. However, this early stage is characterized by asymptomatic changes that are not easily visible to doctors and patients. In this work, an attempt has been made to characterize the changes occurring in the eyes of glaucoma subjects using deep networks, namely, EfficientNet and ResNet-152. For this, 1000 fundus images from the Rotterdam EyePACS AIROGS are considered and given as input to the deep model with a train-test split of 70–30%. The glaucoma classification was studied using various metrics such as train accuracy, validation accuracy, train loss, and validation loss. Further, Grad-CAM visualization is used to investigate on the parts of the fundus image that played an important role in the classification. The results show that the EfficientNet model has a validation accuracy of 0.865, while the ResNet model has a validation accuracy of 0.9199. On analysis of the heatmaps resulting from Grad-CAM visualization, high-intensity activation maps were present near the optic disc and optic cup regions when trained using the EfficientNet model. On the contrary, the activation maps from the trained ResNet model resulted in high-intensity regions spread over the entire fundus image. Thus, it can be inferred that the EfficientNet model classifies the glaucoma images more effectively when compared with the ResNet-152.

Keywords Glaucoma · EfficientNet · ResNet-152 · Grad-CAM visualization · Explainable model

S. Kalisapudi (✉) · R. Palanisamy
Indian Institute of Information Technology Design and Manufacturing Kancheepuram, Chennai, India
e-mail: edm18b023@iiitdm.ac.in

R. Palanisamy
e-mail: rohinip@iiitdm.ac.in

1   Introduction

Glaucoma is an irreversible neurological eye disease that leads to the permanent loss of peripheral or central vision. It is referred to as a category of eye disorders that are attributed to optic nerve injury. Most glaucoma types are asymptomatic, where the disease-associated effects may not manifest until the advanced stages. The delayed diagnosis develops macular degeneration that results in vision loss and blindness [1]. As glaucoma causes irreversible vision loss, early detection is essential to terminate the progression of the disease.

For glaucoma diagnosis, ophthalmologists evaluate the optic disc (also called as optic nerve head) using fundus images. The optic disc is composed of two regions: the optic cup (luminous and center zone) and the neuro-retinal rim. The onset of glaucoma and the disease progression can be characterized by the morphological changes in the optic disc and are specifically associated with neuro-retinal rim thinning. It can be quantified by evaluating cup-to-disc ratio (CDR) [2]. Manually segmented optic disc and optic cup are used for the measurement of CDR. Due to the substantial overlap between the optic cup and neuro-retinal rim areas, obtaining CDR values with high accuracy and reliability remains a significant issue [3].

Deep learning is an emerging tool for automatically extracting features and classifying fundus images for glaucoma diagnosis [4]. Chen et al. [5] have implemented a deep learning model on retinal fundus images to differentiate healthy and glaucoma-infected eye images. Some studies have explored the application of different convolutional neural networks (CNN) to demonstrate the influence on the performance of significant factors such as the dataset size, the architecture, and the use of transfer learning versus newly defined architectures [6, 7]. Bajwa et al. employed two-stage architecture for glaucoma diagnosis. In the first stage, region-based convolutional neural networks (RCNN) are used to segment the optic disc from a retinal fundus image. The second stage is implemented with a deep CNN to assess if the retrieved disc is healthy or glaucomatous [8].

EfficientDet-D0 is a deep learning approach with EfficientNet-B0 as the backbone. Nawaz et al. presented a framework for glaucoma localization and classification that consists of three steps. Initially, the EfficientNet-B0 feature extractor is used to compute the in-depth features from the suspected samples. Then, the EfficientNet bi-directional D0's feature pyramid network (BiFPN) module utilizes the computed features from EfficientNet-B0 and performs top-down and bottom-up key point fusion several times. The final step is to predict the resultant localized area containing glaucoma lesions with associated class [9].

In this study, two deep learning architectures, namely ResNet-152 and EfficientNet, are compared for their efficacy in glaucoma diagnosis using fundus images. Unlike most established detection techniques, this approach does not necessitate feature extraction or precise geometric measurements of optic nerve head structures such as CDR. The explainability of the model is also evaluated using gradient-weighted class activation maps (Grad-CAM). It generates a coarse localization map

highlighting the critical regions in the image that the model utilizes for glaucoma classification [10, 11].

2 Methodology

Figure 1 shows the workflow which has been implemented to achieve the objectives of the study. Classification of glaucoma has been done using deep learning models such as the EfficientNet model and ResNet-152. Results have been validated using metrics like train accuracy and validation accuracy. The model performance is also evaluated using the explainable Grad-CAM approach.

2.1 EfficientNet

EfficientNet was initially constructed as a primary network using neural network model search, a technique for automation neural network creation. It improves accuracy and efficiency, as measured by floating-point operations per second. The mobile inverted bottleneck convolution is used in EfficientNet construction (MBConv). EfficientNet is a fully convolutional design and scaling technique that uniformly scales all depth/width/resolution dimensions using a compounded coefficient. The EfficientNet scaling technique uses a preset set of scaling coefficients to scale network width, depth, and resolution reliably [12].

2.2 ResNet-152

Conventional deep learning networks typically start with convolutional layers, consisting of convolutions with activation and batch normalization, followed by fully connected (FC) layers for classification tasks without skip connections. The problem of vanishing/exploding gradients manifests itself when the plain network is deeper [13]. ResNets try to solve this issue by implementing a skip connection or shortcut connection, adding the input to the output after a few levels [14]. Adding these links allows gradients to flow backward in the network, making training easier. This model's key benefit is that it delivers improved classification accuracy without increasing the model's complexity [15].

Fig. 1 Workflow
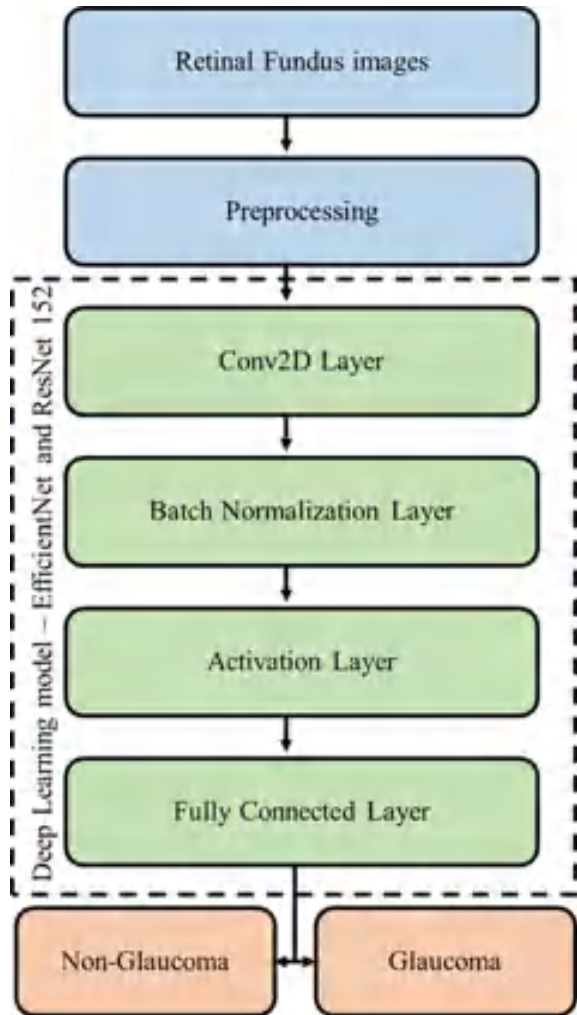
2.3 Model Training and Validation

The models are primarily trained for 20 epochs with a batch size of 16. The learning rate for both models is initially set to 0.0001. The conventional models are modified by adding three additional layers by taking the weights in the fully connected layer. A dropout of 0.3 is added to the first additional layer after the fully connected layer to nullify some of the features and prevent overfitting.

The parameter utilized to calculate the loss function is cross entropy and the Adam optimizer is used in training the model. The loss function is calculated after each training epoch on the entire dataset. The loss generated during the validation phase is used to tune additional hyperparameters. The model is further tuned to store the

weights of the epoch with the lowest validation loss among all epochs up to that point, and it loads the weights back if the validation loss increases in subsequent epochs. If the validation loss increases, the hyperparameters are fine-tuned, and the model is rerun. The model is evaluated using performance metrics, namely training loss, validation loss, training accuracy, and validation accuracy.

2.4 Grad-CAM Visualization

Grad-CAM visualization and analysis are performed to test the model's predictions over the validation data. The model is stopped at the fully connected layer to produce a Grad-CAM heat map, as it will be used for prediction. The output from the layer and loss are obtained from the model after it has been run on that specific image. The gradient of the fully connected layer's output is then calculated regarding the model loss. To overlay the heatmap with the original image, the gradient parts that contribute to the prediction are then reduced, resized, and rescaled [11].

3 Results and Discussions

3.1 Dataset and Training

The study dataset comprising of retinal fundus images is obtained from the publicly available Rotterdam EyePACS AIROGS train set. A total of 1000 images from the dataset were considered, with 500 images corresponding to the referable glaucoma class and 500 images from the non-referable glaucoma class. Human experts labeled the images with referable glaucoma and non-referable glaucoma. The images are split into the train-test split with a ratio of 70–30%. Figure 2 represents the representative fundus images of referable and non-referable glaucoma. On visual examination, no significant differences are observed between the two groups.

3.2 EfficientNet

The ImageNet model's pretrained weights are used to initialize the EfficientNet model. The EfficientNet model accepts both batched and individual image tensor objects. The images are resized using bicubic interpolation to a size of (256, 256). After being rescaled between 0 and 1, the values are finally normalized.

Figure 3a shows the graph between train loss and epochs for the EfficientNet classification model. It is observed that the training loss decreased gradually with the increase in the number of epochs. The graph between validation loss and epochs
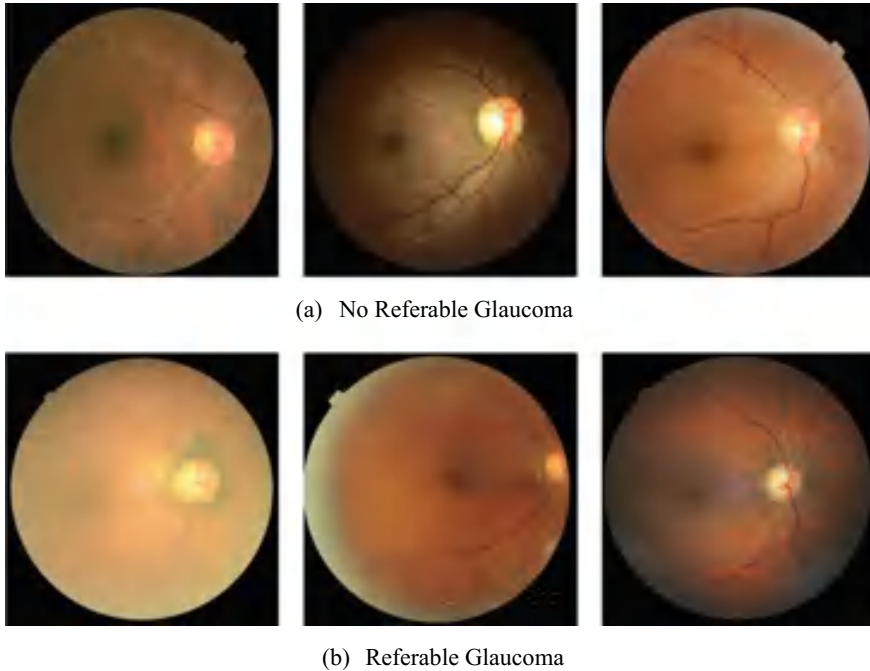
(a) No Referable Glaucoma

(b) Referable Glaucoma

Fig. 2 Representative retinal fundus images

for the EfficientNet model is also shown in Fig. 3b. From the graph, it can be seen that the validation loss reached a minimum value of 0.30 after 20 epochs and train loss reached a minimum value of 0.23 after 19 epochs.

Figure 4a shows the graph between training accuracy and epochs for the EfficientNet classification model. It is observed that the training accuracy increased gradually with the number of epochs. The graph between validation accuracy and epochs for the EfficientNet model is also shown in Fig. 4b. From the graph, the training accuracy improved gradually with the increase in the number of epochs and reached a maximum of 0.908 at the end of 20 epochs, while the validation accuracy reached 0.865.

3.3 ResNet-152

The ResNet-152 model is initiated with pretrained weights of the ImageNet model. The model then accepts both batched and individual image tensor objects. Bilinear interpolation is used to resize the images to (256, 256). Finally, the values are normalized after being rescaled between 0 and 1.
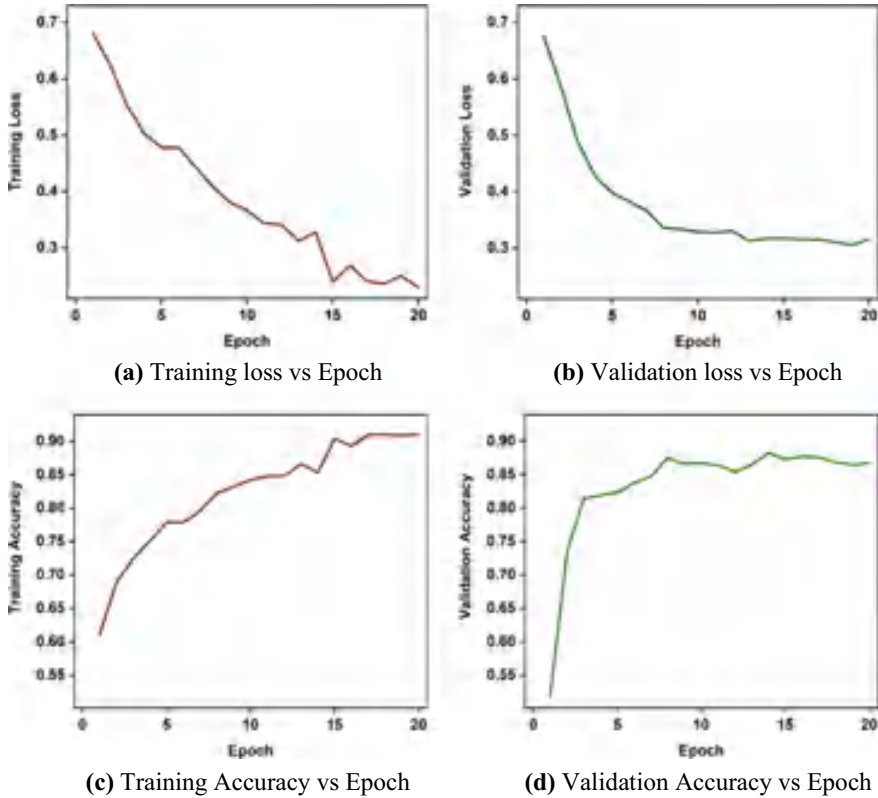
(a) Training loss vs Epoch

(b) Validation loss vs Epoch

(c) Training Accuracy vs Epoch

(d) Validation Accuracy vs Epoch

Fig. 3 Performance metrics of **EfficientNet** model

Figure 4a depicts the relationship between train loss and epochs for the ResNet-152 classification model. The training loss decrease declined significantly in the initial epochs, reaching a value of 0.0054 after 20 epochs. Figure 4b also depicts the validation loss versus epoch graph for the EfficientNet model. The validation loss dropped considerably during the initial epochs and reached a minimum value of 0.134 at the 18th epoch.

Figure 4c depicts the training accuracy versus epochs graph for the ResNet-152 classification model. The validation accuracy increases dramatically in the initial epochs. Figure 4d depicts the relationship between validation accuracy and epochs for the model. The first few epochs observed a steep increase in the training accuracy, after which the curve flattened. As the number of epochs increased, training accuracy reached a maximum of 0.998 at the end of 20 epochs, while the validation accuracy reached 0.9199. However, because of the deep architecture of the ResNet-152 model, the train and test data are seen to be overfitting. Thus, it may not be suitable to recognize accuracy as the actual performance metric.
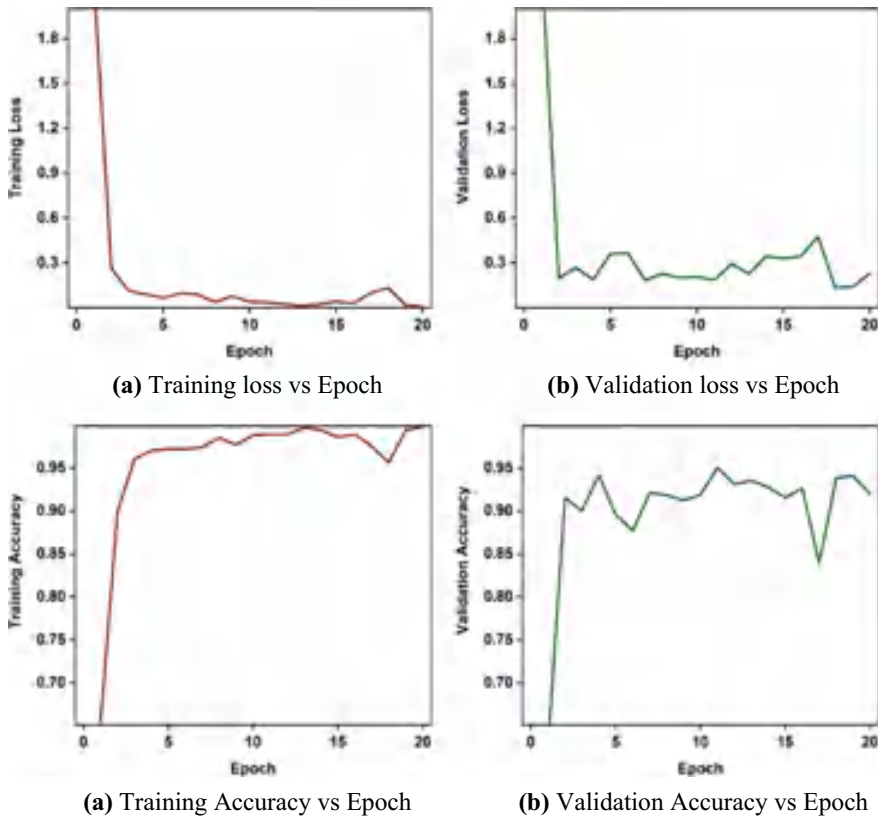
(a) Training loss vs Epoch

(b) Validation loss vs Epoch

(a) Training Accuracy vs Epoch

(b) Validation Accuracy vs Epoch

Fig. 4 Performance metrics of ResNet-152 model

3.4 Grad-CAM Visualization

Grad-CAM uses the gradients in the model's final convolutional layer and generates a heatmap after running the gradient values through the activation function. All the images in the validation data were examined using the algorithm for both models. Figure 5 shows the representative Grad-CAM visualization of the fundus image for the EfficientNet and RestNet-152 models. It is observed that the area where the weights appeared for the EfficientNet model is near the optic disc. However, for the ResNet-152 model, the weights are pointing toward other regions and edges in the image, indicating overfitting issues. Therefore, it is inferred that the EfficientNet model is more suitable for deployment than ResNet-152 because of its greater model explainability on the given dataset.
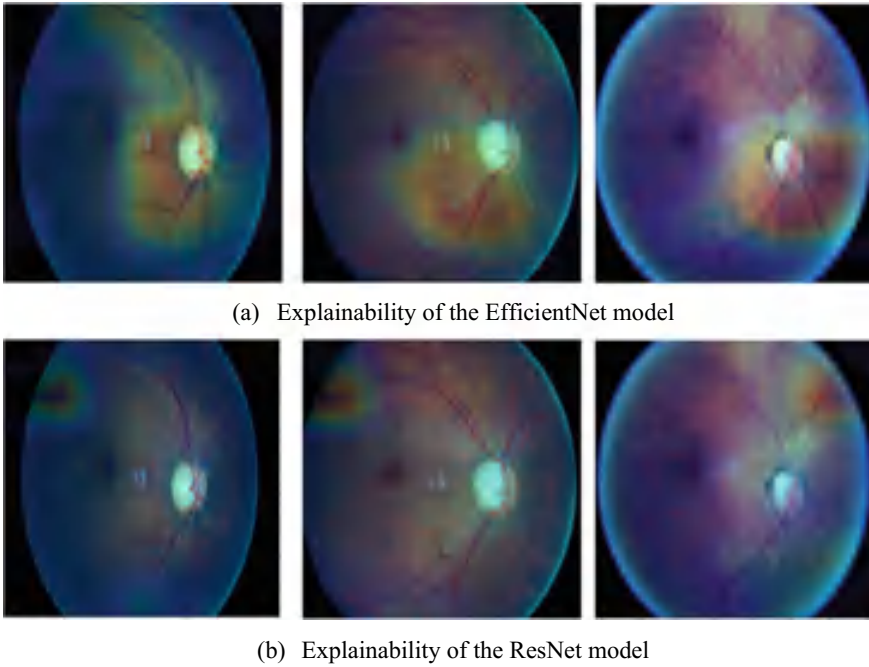
(a) Explainability of the EfficientNet model

(b) Explainability of the ResNet model

Fig. 5 Grad-CAM visualization (high-intensity areas are marked with red)

4 Conclusions

The potential of deep learning methods, namely EfficientNet and ResNet-152, for glaucoma classification is investigated in this work. For this, 1000 fundus images from the Rotterdam EyePACS AIROGS are considered and given as input to the deep model with a train-test split of 70–30%. The two models are contrasted based on training and test accuracy for performance analysis. The EfficientNet model is found to have a validation accuracy of 0.865 and a training accuracy of 0.908. Furthermore, the ResNet-152 model has training and validation accuracy of 0.998 and 0.9199, respectively. Grad-CAM technique is used to visualize the models to assess the explainability and efficiency of both models. On analysis of heatmaps from the EfficientNet model, it is observed that the model is learning the features near the optic disc region that are involved with the disease condition. Based on the model explainability using Grad-CAM, it is possible to conclude that the EfficientNet is more suitable for deployment than ResNet-152.

References

1. Garg H, Gupta N, Agrawal R, Shivani S, Sharma B (2022) A real time cloud-based framework for glaucoma screening using EfficientNet. Multim Tools Appl 81:34737–34758
2. Thakur N, Juneja M (2020) Classification of glaucoma using hybrid features with machine learning approaches. Biomed Signal Process Control 62:102137
3. Zhao R, Chen X, Liu X, Chen Z, Guo F, Li S (2020) Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning. IEEE J Biomed Health Inform 24:1104–1113
4. Zheng C, Johnson TV, Garg A, Boland MV (2019) Artificial intelligence in glaucoma. Curr Opin Ophthalmol 30:97–103
5. Chen X, Xu Y, Yan S, Wong DW, Wong TY, Liu J (2015) Automatic feature learning for glaucoma detection based on Deep Learning. Lecture Notes in Computer Science, 669–677
6. Gómez-Valverde JJ, Antón A, Fatti G, Liefers B, Herranz A, Santos A, Sánchez CI, Ledesma-Carbayo MJ (2019) Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. Biomed Opt Express 10:892
7. Samanta A, Saha A, Satapathy SC, Fernandes SL, Zhang Y-D (2020) Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. Pattern Recogn Lett 135:293–298
8. Bajwa MN, Malik MI, Siddiqui SA, Dengel A, Shafait F, Neumeier W, Ahmed S (2019) Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. BMC Med Inf Decis Mak 19
9. Nawaz M, Nazir T, Javed A, Tariq U, Yong H-S, Khan MA, Cha J (2022) An efficient deep learning approach to automatic glaucoma detection using optic disc and optic cup localization. Sensors 22:434
10. Nguyen LD, Lin D, Lin Z, Cao J (2018) Deep CNNS for microscopic image classification by exploiting transfer learning and feature concatenation. In: 2018 IEEE International symposium on circuits and systems (ISCAS)
11. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International conference on computer vision (ICCV), vol 128, pp 336–359
12. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning, pp 6105–6114
13. Brigato L, Iocchi L (2021) A close look at deep learning with small data. In: 2020 25th International conference on pattern recognition (ICPR)
14. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR)
15. Dawod RG, Dobre C (2022) Resnet interpretation methods applied to the classification of Foliar Diseases in sunflower. J of Agric Food Res 9:100323

Deep Learning Model with Progressive GAN for Diabetic Retinopathy

Vinay Mathukumalli, Vaishnavi Yada, Shreya Bashetty, Seetha Maddala, and Kalyani Nara

Abstract When eye illnesses are found early on, they can often be treated more successfully in the medical sector. One of the main causes of blindness globally is diabetic retinopathy, an eye condition that is a consequence of diabetes. Regular check-ups via diabetic retinopathy (DR) screening programs are essential for detecting the disease as early as possible and for determining the adequate treatment. If this is not detected early, it leads to blindness and hence to be prevented from causing further damage or reducing the risk of vision loss. A significant obstacle to deep learning algorithms' efficacy has been a lack of relevant class-specific data, particularly in the field of medical imaging. So, the main challenge for developing deep learning models is the limited amount of data. Although datasets are available from different resources, they are not sufficient for developing deep learning models. To address this issue, the dataset is artificially expanded by creating various iterations of a real dataset using generative adversarial technique (GAN), a data augmentation technique. The outcomes demonstrate that the GAN technique (i.e., data augmentation) aids in the generation of new samples from the existing dataset and improves the deep learning models' accuracy to provide the best-fit model for the dataset.

Keywords Diabetic retinopathy · Generative adversarial networks · Deep learning

V. Mathukumalli
University of New Haven, West Haven, USA
e-mail: Vmath3@unh.newhaven.edu

V. Yada · S. Bashetty · S. Maddala · K. Nara (✉)
G. Narayanamma Institute of Technology and Science, Shaikpet, Hyderabad 500104, India
e-mail: nara.kalyani@gnits.ac.in

S. Maddala
e-mail: maddala.seetha@gnits.ac.in

1 Introduction

Diabetic retinopathy is a disease which can cause to the eye and may get the vision loss and sometimes blindness for the people who are having diabetes. It is affected to the blood vessels at the back of eye, and the initial stages of diabetic retinopathy usually do not have any symptoms [1]. People observe some changes in their sight, like trouble reading or seeing faraway objects, and overtime, it can damage the retina. Diabetic retinopathy may not have any symptoms at initial times, but detecting it early can help you follow some procedure to protect your vision [2]. As prevention is better than cure, early detection could improve the chances of cure and also be able to prevent blindness.

In recent times, machine learning research is going on for diagnosing diseases like diabetic retinopathy, which can be made by extracting features and then classifying the image. Researchers are trying to come up with the optimal ways of classifying retinal problems from those of the healthy ones with feature extraction. But most importantly, in order to classify retinal fundus images accurately, the dataset size matters as a deep learning model needs a large dataset. When the size of the dataset does not match with the user requirements, generation of more images is needed. Hence, the usage of GAN that is generative adversarial networks became necessary. We can also increase the dataset size by using traditional data augmentation which was already done by many other researchers with the goal that we cannot recursively check if an image is generated in the proper way or not. Using generative adversarial networks data augmentation technique, the dataset that is currently being used could be increased as it synthesizes fake images which are hard to separate apart from the actual images.

Automatic segmentation of retinal lesions has a long standing and challenging task for learning-based models, mostly due to the deficiency of available and accurate lesion segmentation datasets. Here, this paper, a two-step process, is proposed for generating photorealistic fundus images conditioned on synthetic "ground truth" semantic labels and shows its potential for further downstream tasks, such as not limited to automated grading of diabetic retinopathy, dataset balancing, and creating image examples for trainee ophthalmologists, etc. [3].

To be able to design a model which can accurately predict the patient's eye conditions, a decent dataset size is necessary. The main problem is to increase the dataset size for enhancing the performance of the classification model.

A technique that was intended for solving the problem of having inadequate fundus images was data augmentation technique. It is a technique used for increasing the amount of data in the dataset by adding some images which are slightly modified copies of already existing data. Data augmentation uses the techniques of flipping and folding of the current images which were not able to improve the performance to the expected level [4].

Some of the image augmentation techniques discussed in this review include geometric transformations, color space augmentations, kernel filters, random erasure,

feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning [5].

By creating new, synthetic, but plausible instances from the input issue domain on which the model is trained, data augmentation works. The processes are simple and entail cropping, flipping, zooming, and other uncomplicated changes of the pre-existing images in the training dataset in the case of picture data. The need for quality evaluation of the output will increase as data augmentation techniques are used more frequently.

Convolutional neural networks and other deep learning methods are used in generative adversarial networks, a subset of generative modeling [6]. Convolutional neural networks have achieved great success for many vision tasks, including classification of images [7], detection of objects [8], semantic segmentation [9, 10], and synthesis of image [11].

The goal of this generative modeling is an example of unsupervised learning job in machine learning, which is to automatically identify and learn the regularities or patterns in incoming data in which the model may be used to produce and generate new samples that can be taken from the original dataset.

With the two sub-models—the generator model, which is trained to create new instances, and the discriminator model, which attempts to categorize examples as either real (from the domain) or fake—GANs behaves intelligently when training a generative model.

The discriminator model is tricked around half the time during training of the two models, indicating that the generator model is producing believable examples. Generative adversarial networks (GANs) are fulfilling the usage of generative models in an exciting and rapidly growing field by generating realistic examples across a range of problem domains, most notably in image-to-image translation tasks such as translating photos of summer to winter or day to night and in creating photorealistic images of objects, scenes, and people that even humans cannot tell are generated image (fake). The dataset that is currently being used could be expanded with the aid of generative adversarial networks because they create fake images from real images that are exceedingly difficult to distinguish from the real ones.

The results show that generative adversarial networks data augmentation technique has the potential to improve or enhance classification performance over the baseline [12].

Generative adversarial network (GAN) is a technique to generate modeling using deep learning methods. Generative adversarial networks (GANs) will be trained with fundus images generating synthetic fundus images with DR. The performance of the classification CNN model will be trained on the combination of both real and synthetic datasets. Images obtained by using generative adversarial networks (GANs) should appear to be realistic, and also increase the performance of the model. The performance of the deep learning CNN model which uses the synthesized dataset should be close to the real images, suggesting that the dataset could be used for training humans and machines.

Joint adversarial losses are used to optimize the entire network from end to end using multi-scale discriminators that operate from large to small receptive fields

[13]. Retina generators are suggested on vascular and lesion masks, and they may be used to adjust the synthesized grading severity by adopting adaptive grading vectors collected from the latent grading space.

Various researchers worked on deep learning models, but still, there is a dearth in the dataset size-increasing techniques for improving the accuracy of the model and validating the model on global dataset rather than local dataset. Hence, this paper emphasizes on the development of data augmentation techniques like GAN for increasing the size of the dataset for real-time dataset and improving the accuracy of the model. Numerous tests are conducted on the EyePACS dataset of Kaggle to verify their efficacy in producing highly realistic (1280 × 1280) controllable fundus images as well as aiding in the DR grading task [14].

The challenge is not only in the design of portable devices but even with the design of suitable models that can classify the image. Simi Sanya et al. [15] used an ensemble CNN model to classify the diabetic retinopathy and achieved 74% accuracy. The review by Sravani et al. [16, 17] reported that the detection of retinal abnormalities using techniques in deep learning has good potential, but this research requires huge data. Comparison of results obtained by methods like CNN and DCGAN and observed loss on general images is less and can be recommended for classification of eye diseases. The ensemble method is observed to be more accurate compared to basic CNN models for the hyperparameters chosen [18]. The accuracy of the model is 75% achieved in the identification of diabetic retinopathy.

2 Generative Adversarial Networks for Diabetic Retinopathy Detection

Firstly, a discriminator model is lined up using 100% real samples and 100% fake samples, respectively. The generator model inputs a random vector from latent space and generates new images. The discriminator model receives these images together with the real photos, and it categorizes them as either fake or real images. The accuracy of real and fake images is calculated along with the discriminator loss. Discriminator loss is backpropagated and the generator model is updated.

The above procedure continues until the specified epochs are reached or until the discriminator is fooled to the required extent. A point should be reached where the accuracy of both real and fake images is close to 50%. This is the point where equilibrium is established, i.e., the point where the discriminator cannot distinguish an image sampled by a generator and an image sampled from the real data distribution. Many parameters can be changed, and different images can be generated. The point where the accuracy of both real and fake images is close to 50% which is considered the most prominent one. Until the requisite accuracy is attained, all of the aforementioned procedures are repeated. The organization of the project and a clear overview are both essential for the successful application of this methodology.

Deep learning-based generative models are known as generative adversarial networks (GANs). A generative model is a type of unsupervised learning method used in machine learning that involves discovering patterns or regularities in incoming data automatically so that the model can be used to generate new examples. In order to assess, capture, and duplicate the variations within a dataset, generative adversarial networks (GAN) generally consist of a system of two competing neural network models that compete with one another. A generative adversarial network has two parts: generator and discriminator. The generator gains the ability to produce plausible data. The instances that are generated serve as negative training examples for the discriminator. The discriminator learns to discriminate between actual and fake data produced by the generator. The generator is penalized by the discriminator for creating implausible results. In order to prevent easy misclassification, generative adversarial networks are being used.

Generative adversarial networks (GAN) models should be trained using realistic images. As we know earlier generative adversarial networks (GAN) constitute two models: generator and discriminator, both generator and discriminator models should be trained using training data.

The discriminator is trained using 100% real and 100% fake images. The generator is trained using discriminator loss. After training the discriminator, it is used to classify both real and fake images. The error of the discriminator is backpropagated, and the generator model is updated.

Figure 1 depicts the GAN implementation process in more detail. The discriminator is initially fed real images, and once it determines that these are real images, the generator is turned off. Later, the generator would also be fed with real photos, and by combining its understanding of the input images and the feedback from the discriminator, the generator would learn to produce fake images. The generator gains the ability to convince the discriminator to mark its output as real. In the training phase, the generator and the discriminator, both of which are neural networks, compete with one another. The generator and discriminator get better after each iteration of the stages, which are performed numerous times.

The discriminator loss is the main factor that enables the generator and discriminator to perform better over time. Discriminator loss is the loss generator at the discriminator after classifying the input images given to the discriminator as real or fake. As a result of the discriminator loss being backpropagated to the generator, the generator is projected to produce more realistic pictures or images in the subsequent interaction than it did in the prior iteration.

Generative adversarial networks (GANs) are divided into three categories:

- Generative: This describes how data is generated with respect to a probabilistic model.
- Adversarial: A model is refined in a competitive environment.
- Networks: Deep neural networks can be used as artificial intelligence (AI) algorithms for training.

GANs consist of a generator and a discriminator. The generator creates fake data samples (images, audio, etc.) in order to trick the discriminator. The discriminator,
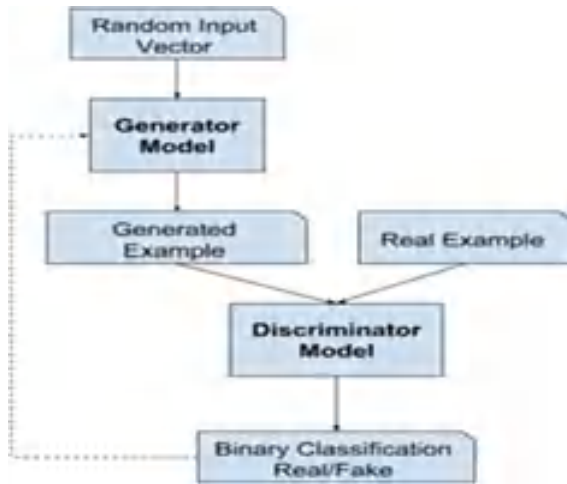
Fig. 1 Flowchart of GAN

on the other hand, attempts to differentiate between actual and fake samples. In the training phase, the generator and the discriminator, both of which are neural networks, compete with one another. The procedures are done repeatedly, and the generator and discriminator get better and better at their jobs with each iteration. The tight connection between the generator and the discriminator as depicted in Fig. 2 is the foundation upon which the constructed system is entirely built. The GANs are designed as a minimax game, with the discriminator attempting to reduce its reward V(D, G) and the generator attempting to decrease the discriminator's reward or maximize its loss. Generative adversarial networks (GANs) have proven to be an effective foundation for creating synthetic databases of anatomically consistent retinal fundus scans or images. In ophthalmology, GANs in particular have shown increased interest [19].

The sole reason why the error at the generator is minimum and ideally considered to be zero is that the error rate in this phase could be calculated between the given
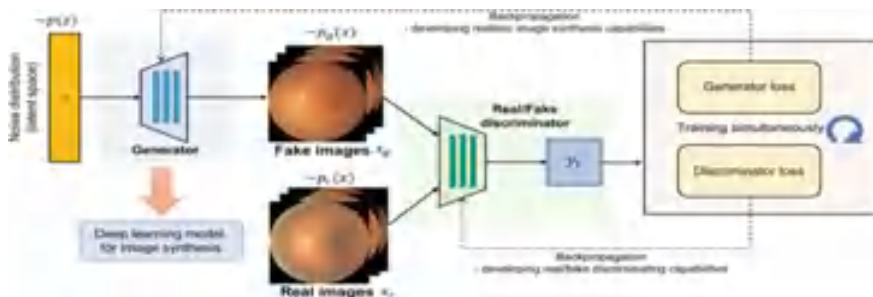

Fig. 2 Architecture of GAN

input images and the generated output images. The main intention of the generator in the model is to create images that look precisely like real images, i.e., the input images and the output images look very close to one another, and hence, the error rate is close to zero.

On the same page, the error rate at the discriminator would have to be maximum, ideally close to 100% because the discriminator would have to come to a point where it should not be able to correctly distinguish whether the image that it has been inputted with is real or not. So, when a fake image is given to the discriminator, it should classify that image to be real. Hence, as the discriminator is classifying a fake image as a real one, the error rate would be maximum.

2.1   Generative Adversarial Networks Algorithm

The algorithm of generative adversarial networks (GAN) is as described below.

The algorithm of Generative Adversarial Networks (GAN) is as described below.
Algorithm GAN {
1. Import the dataset into google drive
2. define the standalone discriminator model
   - def define_discriminator(in_shape=(32,32,3))
     ▪ normal
     ▪ downsample
       o For focusing just on the image's primary elements.
       o Having an image with large size might force the model to even consider the miniscule details with greater importance so as to avoid, downsampling is done.
     ▪ Downsample
     ▪ classifier
     ▪ compile model
3. define the standalone generator model
   - def define_generator(latent_dim)
     ▪ foundation for 4x4 image - Sequential()
     ▪ upsample to 8x8
     ▪ upsample to 16x16
     ▪ upsample to 32x32
     ▪ output layer

4. define the combined generator and discriminator model, for updating the generator
   - def define_gan(g_model, d_model)
     ▪ make weights in the discriminator not trainable
     ▪ connect them
       o   add generator

          ○   add the discriminator
        ▪ compile model

5. load and prepare training images
   - def load_real_samples()
6. Train the model
   - Divide number of batches
   - Divide the batch into half
   - In every batch take half or three fourth of the images randomly and train the discriminator model with the real images.
   - Generate 100% fake samples randomly.
   - Train the discriminator model with fake samples. Now the discriminator model is trained with both real and fake samples.
   - Generate a few random latent points, invert y label and train the generator model i.e., calculate discriminator loss and update the generator model.
   - Display the model's executive summary every ten epochs.
   - NOTE: After every batch, weights are updated and the model is built based upon the previous weights.
7. Find accuracy using accuracy_score from sklearn.
   - Images present in train, test and validation are converted to numpy arrays.
   - Normalize the data.
   - Generate the data using ImageDataGenerator(preprocessing).
   - To avoid overfitting, use early stopping.
   - Fit the model.
   - Analyzation of the train_accuracy and test_accuracy.
   - Analyzation of train_loss and test_loss

The GAN algorithm is implemented, and the accuracies of both real and fake data are compared and analyzed.

Used the CNN classification model for finding the performance of model.

3 Dataset Description

Dataset-related diabetic retinopathy is obtained from L. V. Prasad Eye Institute, Hyderabad. Diabetic retinopathy is the medical term, where blood vessels present at the back of the eyes are damaged as a result of diabetes. This may be with no symptoms initially but eventually notice vision problems.

This dataset is an image dataset consisting of 250 images of different levels in each separated folders; levels [20] of disease are Level 0—No DR, Level 1—Mild, Level 2—Moderate, Level 3—Severe and Level 4—Proliferative DR. Each level of disease has 50 images of different patients. Dataset represents a real-life set of patient information collected by LVPEI from different patients.

Table 1 Accuracy of real and the generated fake images for various number of epochs

Epochs	Discriminator loss	Generator loss	Accuracy for real images (%)	Accuracy for fake images (%)
25	0.155	3.696	52	65
50	0.468	3.426	52	66
75	0.441	3.057	53	66
100	1.817	4.522	54	67
150	2.271	3.482	55	67
200	1.920	2.192	56	68
500	1.072	1.809	60	70
1000	0.743	2.372	62	71

4 Results and Discussions

GAN comprises two models—generator model and discriminator model. Generator model generates new images. Discriminator makes an effort to distinguish between actual and fake images. There is a tough competition between generator and discriminator models. The generator and discriminator models should be balanced. Images, whether actual or fake, should be about 50% accurate. At this time, there is no way for the discriminator model to distinguish between actual images and those produced by the generator model.

Various combinations should be checked, and comparisons need to be made to examine the output where accuracy of both real and generated fake images is close to 50%. **The comparison for the data existing and data generated after data augmentation is made, and results obtained are tabulated as exhibited in Table 1**.

According to Fig. 3, the training and testing accuracy is seen to have stabilized after 600 epochs.

As shown in Fig. 4, the training and validation loss is seen to decrease as the number of epochs increases.

The contribution of this work demonstrates how the GAN data augmentation technique aids in producing new samples from the existing dataset and how deep learning models are seen to have an accuracy of 28% for actual real images, 60% for false images, and 50% for a combination of fake and real images.

5 Conclusions and Future Enhancements

A generator and a discriminator are trained iteratively using an adversarial learning technique of generative adversarial networks (GAN). Generative adversarial networks (GANs), a potent class of generative models, learn to produce new
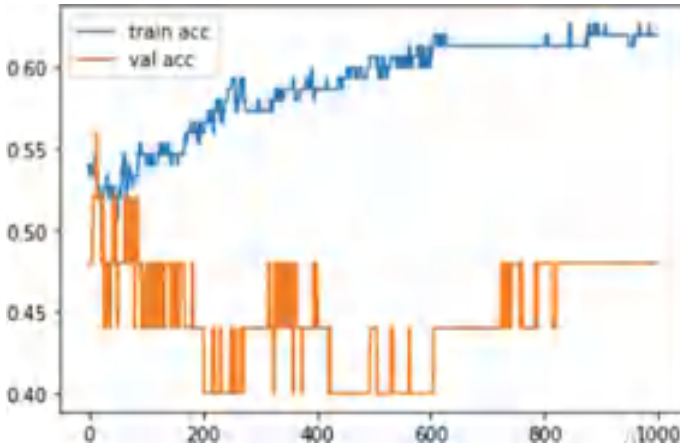
Fig. 3 Graph plot of accuracy for training and testing
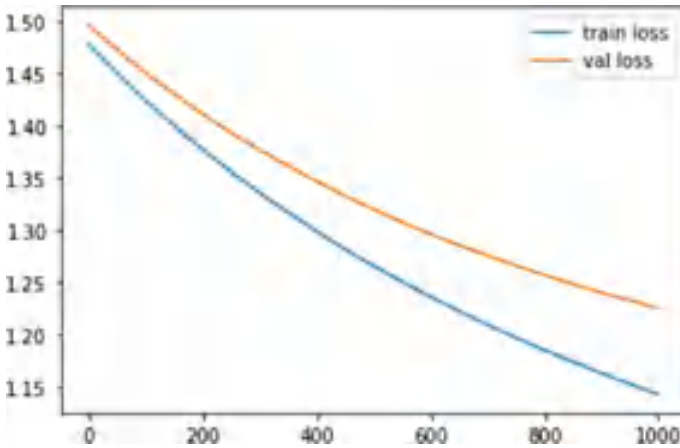

Fig. 4 Graph plot of accuracy for training and validation loss

samples that follow the same distribution as the actual samples without than explicitly estimating the distribution of the samples. Speech and language processing, image and vision computing, and other fields have key application value for the ability to produce "infinite" new samples from the prospective distribution.

Initially, the dataset has been examined and preprocessed accordingly. Building a CNN model to measure the effectiveness of the generative adversarial networks (GANs) required training the generator and the discriminator until the desired results were obtained.

References

1. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410
2. Sarki R, Ahmed K, Wang H, Zhang Y (2020) Automatic detection of diabetic eye disease through deep learning using fundus images: a survey. IEEE Access 8:151133–151149
3. Joon-Ho A, Rueckert D, Kainz B, Hou B, Synthesis of diabetic retina fundus images using semantic label generation
4. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6, Article number: 60
5. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. Math Comput Simul 6:60. https://doi.org/10.1186/s40537-019-0197-0
6. Wang B, Zhou Y, He X, Cui S, Shao L, DR-GAN: conditional generative adversarialnetwork for fine-grained lesion synthesis on diabetic retinopathy images. https://drive.google.com/file/d/1mFh6AzV_8jYxDoLL_uRTLNYfNxN59z79/view?usp=sharing
7. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. CVPR, pp 770–778
8. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikainen M (2020) Deep learning for generic object detection: a survey. Int J Comput Vis 128(2):261–318
9. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Artinez-Gonzalez P, Garcia-Rodriguez J (2018) A survey on deep learning techniques for image and video semantic segmentation. Appl Soft Comput 70:41–65
10. Fan D-P, Zhou T, Ji G-P, Zhou Y, Chen G, Fu H, Shen J, Shao L (2020) Inf-net: automatic Covid-19 lung infection segmentation from CT images. IEEE Trans Med Imag
11. Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. Med Image Anal 58:101552
12. Lim G, Thombre P, Lee ML, Hsu W, Generative data augmentation for diabetic retinopathy classification. https://doi.org/10.1109/ICTAI50040.2020.00167
13. Zhou Y, He X, Cui S, Zhu F, Liu L, Shao L (2019) High-resolution diabetic retinopathy image synthesis manipulated by grading and lesions. In: MICCAI. Springer, Berlin
14. Kaggle diabetic retinopathy detection competition. https://www.kaggle.co/c/diabetic-retinopathy-detection
15. Sanya S, Maddala S (2021) An ensemble of convolutional neural network for diabetic retinopathy detection. Int J Eng Res Technol (IJERT) 9(2). ISSN: 2278-0181
16. Sravani Devi Y, Phani Kumar S, A Scoping review of diabetic retinopathy detection techniques using deep learning: taxonomy, methods, and recent developments. High Technol Lett. ISSN: 1006-6748
17. Sravani Devi Y, Phani Kumar S (2021) Retinal image synthesis for diabetic retinopathy assessment using dcgan and vae models. Int J Adv Electron Comput Sci 8(10). ISSN: 2394-2835
18. Seetha M, Kalyani N, Sravani Devi Y (2021) An ensemble CNN model for identification of diabetic retinopathy eye disease. In: 5th international conference on smart computing and informatics, SCI
19. Bellemo V, Burlina P, Yong L, Wong T, Ting D, Generative adversarial networks (GANs) for retinal fundus image synthesis
20. American Academy of Ophthalmology (2012) International clinical diabetic retinopathy disease severity scale

Channel Estimation in Massive MIMO Using BS Identification Code

Bhagya Bharathi Padavala, G. Krishna Reddy, and G. Merlin Sheeba

Abstract Massive Multi-input Multi-output (MIMO) systems plays an important role in 5G communications. In Massive MIMO, a base station with large antennas simultaneously uses the same frequency and time slots by multiple users, due to this pilot contamination may occur. In order to avoid the pilot contamination in LTE-TDD Massive MIMO systems, a technique is to be considered which uses an orthogonality property of Hadamard code. The pilot signal needs to be modulated with the Hadamard code in order to identify the base station in the cellular network. The identification of base stations can be done by applying the low computational complexity suboptimal linear LS criterion and also the mitigation of the effect of pilot contamination in case of full pilot reuse in multi cell scenarios can also be performed.

Keywords Massive MIMO · LS · LTE · Channel estimation · OFDM · TDD

1 Introduction

To meet the objective or prerequisites representing things to come arranged and correspondences society with steady admittance to and sharing of data everybody and all over the place. The remote association network should upgrade limit and information rate, further developed assistance quality and diminished dormancy. Boycott of channels different shaft division multiplexing (FBMC) it is proposed to take on fifth age (5G) various access (BDMA) innovation. They help for the cell business' developing client request expanding limit and information rate for networks, low finish to lessening network idleness and laying out associations with tremendous number of equipment [1, 2].

B. B. Padavala (✉) · G. Krishna Reddy
G. Narayanamma Institute of Technology and Science, Hyderabad, India
e-mail: bhagyabharathi98@gmail.com

G. Krishna Reddy · G. Merlin Sheeba
Sathyabama Institute of Science and Technology, Chennai, India

For now, symmetrical recurrence division various OFDMA access advances are viable with the fourth-age (4G) remote norm and will be viable with the accompanying age It isn't important to remotely modify the 4G organization arrangement. There should be upgrades and consolidate a 4G fundamental organization application fulfill the requests of the client. The outcome is utilization, the specialist co-op moved to the 5G organization business 4G organization arrangement. In any case, to achieve the need of meeting client standards and achievement the trouble that the 5G cell engineering presents organizations should adjust their plan draws near.

In the 5G organization plan, the base station (BS) side conveys a Gigantic Various Information Different Result (Enormous MIMO) exhibit of ten to one hundred receiving wires [3] to help framework limit [3, 4]. In any case, to appraise Channel State Data (CSI) of the MIMO divert in both the uplink and downlink sides, direct assessment is important to profit from taking on Monstrous MIMO innovation.

In an ordinary MIMO downlink framework, the base station (BS) communicates a pilot sign to the versatile station (MS) to gauge the channel, and the MS in this way sends CSI criticism to the BS. In the event that the channel makes some fast memories differing blurring, Huge MIMO can't be utilized. Since with the increase of number of array antennas, there is an increase in the ideal time and recurrence slots of the pilot signal, the orthogonality should be maintained by the pilot signal emitted from BS antennas. The high computational complexity of channel estimation is thus reduced by the BS to MS data transfer rate, and CSI feedback takes longer. Therefore, Massive MIMO's uplink channel estimation is less complicated and more suitable than its downlink.

Every cohesive time the resource block is used, the channel estimations must be updated (RB). By performing comparison between the number of active MS users in cells with the number of orthogonal frequencies allotted to pilot signals, the number of frequencies has been constrained. As a result, pilot is reused by the BSs nearer to the targeted cell in the cellular network and pilot signals are transmitted at the same time and frequency, which is known as pilot contamination [5]. The exact channel's channel estimation is affected by the pilot signals from other cells that transmit at high power. Cellular networks' overall capacity and data transfer rate are thus decreased. This method uses the Hadamard coding to recognize the base station. Estimation the channel and allay the impact of pilot contamination, orthogonal property of the code is being used by the cost function of the LS algorithm. The recommended technique can magnify the SINR performance and the possible downlink rate of the LTE-TDD Massive MIMO downlink system, according to numerical results.

2   System Model

As claimed by the LTE-TDD Massive MIMO technology, this segment presents both uplink and downlink systems MU-MIMO baseband model. LTE-TDD and Massive MIMO baseband model is shown in Fig. 1.
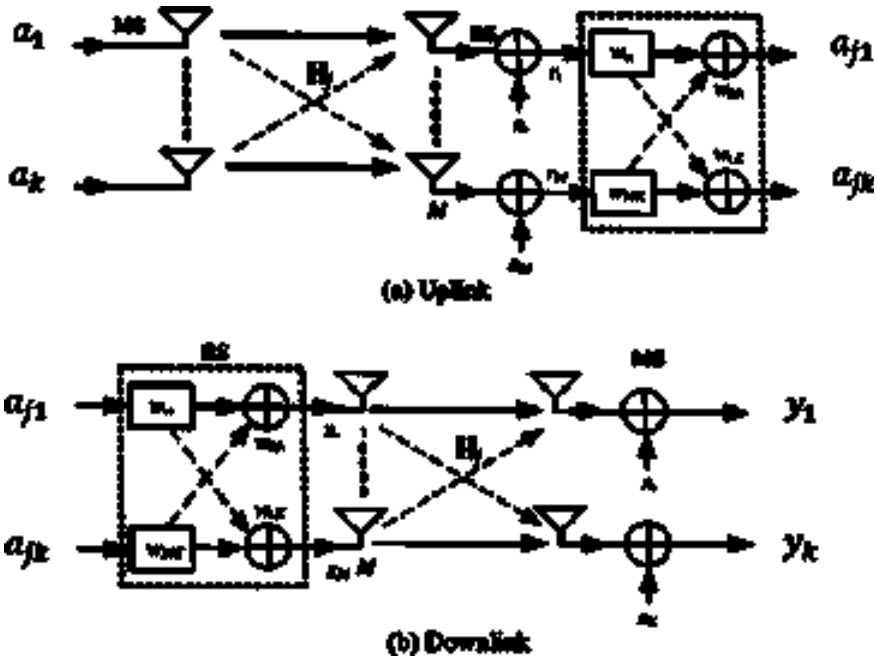
Fig. 1 MIMO base band model

Signal a_j from K MS users in cell j with a size of K_1 is modulated with K orthogonal frequencies and broadcast across time-varying fading channels in the uplink while considering the TDD time slot ($H_j = [h_j 1\ h_j 1\ ...\ h_j K]$). Prior to reaching base station (BS) antennas, this signal is also tainted by nearby cells ($l \neq j$) and Additive White Gaussian Noise (AWGN) with a variance of δ_n^2. As a result, the signal that was received at M antennas in cell j can be expressed as

$$r_j = \sqrt{P_T} H_j a_j + \sqrt{\zeta P_T} \sum_{l=1, l \neq j}^{L} H_l a_l + n_j \tag{1}$$

where P_T is the transmitted power from MS users, and ζ is the cross gain and j is the interested cell with M being the number of BS antennas. And in sequence to get the estimation channel that is most alike to the practical channel, the BS processor estimates channel H_j. The Zero-Forcing (ZF) criterion is used in this section to lessen the impact of interfering signals.

The following is the calculated weight vector for ZF:

$$\hat{H}_j = \left(H_j^H H_j\right)^{-1} H_j^H \tag{2}$$

where -1 denotes inverse matrix, H Hermitian transpose.

Prior to being received by each MS receiver in the downlink, data $a_j(K_1)$ is damaged by interference from other cells $(l \neq j)$ and AWGN before being transmitted at BS antennas over fading channel H_j with a size of MK. Fading channels and noise cause the signal to appear at MS receivers with distortion due to a greater Bit Error Rate (BER), which lowers network capacity.

As a result, the processors at BS must function as weighted vector coefficients for each MS to offset the effects of fading channels and reduce pilot contamination. The following is the definition of the weighted coefficient matrix with dimension $M \times K$:

$$\hat{W}_j = \left| \hat{W}_{j1} \hat{W}_{j2} \ldots \hat{W}_{jK} \right|$$

$$\hat{W}_{jK} = \frac{\hat{h}_{jk}}{\left\| \hat{H}_j \right\|} \tag{3}$$

The received signal at MS user can be written in a mathematical model as

$$y_j = H_j^T X_j + \sum_{l=1, l \neq j}^{L} H_l^T X_l + n_j$$

$$X_j = \sqrt{\frac{P_{BS}}{MK}} \hat{W}_j \left[a_{j1} a_{jk2} \ldots a_{jk} \right] \tag{4}$$

where is the total BS transmitted power and L is the total number of active cells in the cellular system.

3 Pilot Based Hadamard Code for Channel Estimation

The whip hand of Hadamard code having the shared symmetry property is examined in this article. To recognize BS and diminish the effect of pilot pollution, the symmetrical code is balanced with every BS pilot signal. To appraise time-differing blurring channels, the low computational intricacy sub-standard direct LS strategy is utilized.

Hadamard code (c_j) with the size of $1 \times \tau$ that is modulated with pilot signal $(JP_{jkPilot})$ to recognize BS.

The modulated signal at cell j is defined as

$$a_{jkPilot} = S_{jkPilot} C_{jkPilot} \tag{5}$$

Prior to reaching the BS receiver with M array antennas, the transmission of modulated pilot signal from K users in cell j across a fading channel is done, contaminated by nearby cell pilot signals $(l \neq j)$, and distorted with AWGN.

4   Numerical Results and Discussion

For framework reproduction, a LTE-TDD MU-MIMO framework with M exhibit radio wires introduced at the base station and a solitary receiving wire introduced at the MS client is set up. The framework's boundaries stick to the Asset Block matrix of OFDM. As per the LTE standard, it comprises of seven OFDM images, every one of which involves twelve symmetrical subcarriers. Each subcarrier's tone is given a 15 kHz transfer speed; subsequently 180 kHz of lucid data transmission is required. Because of the RB pair (14 OFDM images) prerequisite that the time-recurrence component be equivalent to $12 \times 14 \times 168$ components, the OFDM image term is 71.4 s with watch time or Cyclic Prefix (CP) of 4.76 s. We consider the situation where gigantic MIMO is put on BS with $M = 100$ radio wires and $K = 10$ MS clients dynamic in every cell. The eight nearby cells that broadcast pilot signals simultaneously and frequency or full pilot reuse disrupt the fascinating cell j [4].

Every cell has a functioning pilot signal from MS clients that is tweaked with a Hadamard code of 16 images for BS ID. Prior to showing up at M exhibit receiving wires of the BS recipient, the regulated pilot signals are sent by time-shifting Person on foot blurring channels. Intelligible time and recurrence asset 416 is hence required. We accept that every dynamic tone in the phone has a symmetrical recurrence and that the sign to clamor proportion (SNR) of the got signal at the BS beneficiary is set to 30 dB. There is hence no intracellular impedance. Expecting that the diverts in both uplink and downlink frameworks capability in a reasonable way, their measurable qualities ought to stay consistent. Pilot signal transmission power from user k in cell j is = 1, while pilot signal transmission power from each adjacent cell is PT. By changing the cross gain (ζ) from 0 (no contamination from adjacent cell) to 7, the simulation explores the consequences of pilot contamination.

The proposed technique's downlink rate and SINR performance are compared in the simulation [5]. To distinguish between BS and low complexity, a pilot signal is modulated into a Hadamard code. By contrasting the ideal ZF algorithm (2) with the standard LS algorithm and assuming perfect channel knowledge, the LS algorithm is employed to accomplish channel estimation. Assuming that the uplink and downlink systems' channels are symmetric at RB time. As a result, the downlink side can use the channel coefficient that was acquired from the uplink side's channel estimation.

The simulation results are depicted in Figs. 2 and 3. When power contamination from surrounding cells increases, the procured SINR performance from each of the three algorithms falls. Furthermore, the possible downlink rate (b/s/Hz) derived from three techniques diminishes as the cross gain rises.

The suggested algorithm's SINR performance and feasible downlink rate (b/s/Hz) are both better and farther up than those of the conventional LS algorithm, even if the ideal ZF algorithm's SINR performance has been enhanced and is superior to both. The resulting SINR from the suggested technique is 0.2 dB better than that from the traditional LS algorithm. Additionally, the suggested approach achieves a downlink rate of 0.3 b/s/Hz as opposed to the traditional LS algorithm's 0.2 b/s/Hz.
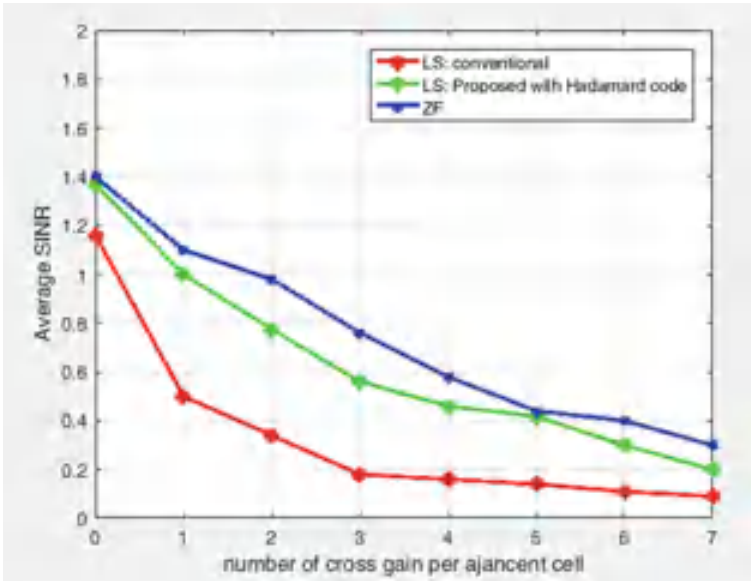
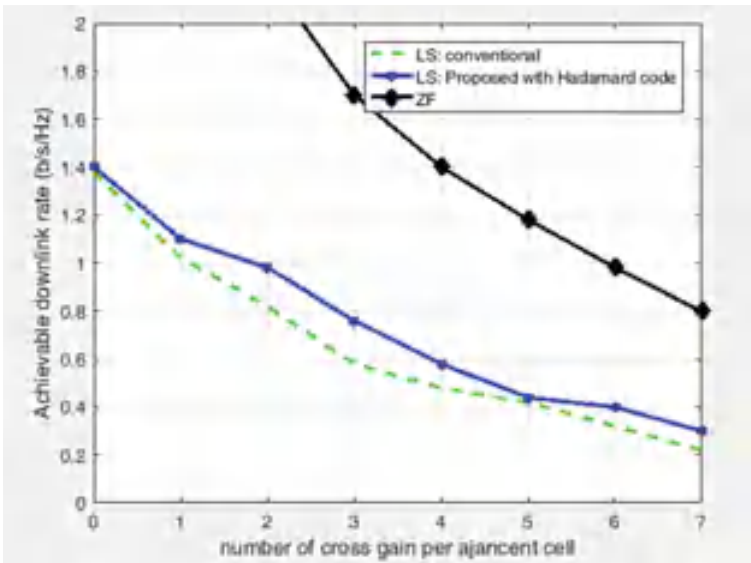**Fig. 2** Averaged SINR versus cross gain (ζ)

Fig. 3 An achievable downlink rate versus cross gain (ζ)

As a result, it is practically difficult to turn over data from BS to MS using the traditional LS technique.

5 Conclusion

Because of the developing number of remote gadgets utilizing the cell organization and doing as such for longer timeframes. There is a restricted measure of RB transmission capacity accessible for LTE-TDD Monstrous MIMO frameworks. In this manner, the impacts of pilot pollution can't be forestalled when the pilot signal from cells in a phone network is reused. To recognize the BS, this examination gives a clever technique that utilizes symmetrical code regulated with pilot signal. For channel assessment at BS, the low computational intricacy LS procedure is utilized. To decrease the effect of pilot tainting, the expense capability of the LS calculation is constrained to utilize the advantage given by the symmetry property of the code. The proposed strategy can lessen the impacts of full pilot reuse, as per the mathematical outcomes. The proposed strategy outflanks the conventional LS calculation regarding SINR execution and downlink rate that might be accomplished.

References

1. Fallgren M et al (2013) Scenarios, requirements and KPIs for 5G mobile and wireless system, document ICT-317669-METIS/D1.1
2. Industry Proposal for a Public Private Partnership (PPP) in Horizon 2020 (Draft Version 2.1), Horizon 2020 advanced 5G network infrastructure for the future internet PPP [Online]. Available: http://www.networks-etp-eu/fileadmin/userupload/Home/draft-PPP-proposal.pdf
3. Hsieh M-H, We C-H (1998) Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels. IEEE Trans Consum Electron 44(1):217–225
4. Khlifi A, Bouallegue R (2011) Performance analysis of LS and LMMSE channel estimation techniques for LTE downlink systems. Int J Wirel Mob Netw (IJWMN) 03(05):141–149
5. Xu P, Wang J, Wang J (2013) Effect of pilot contamination on channel estimation in massive MIMO systems. In: Proceedings of international conference on wireless communication signal processing (ICWCSP), pp 1–6

Improvement of Modified Social Group Optimization (MSGO) Algorithm for Solving Optimization Problems

Sai Shaktimayee Sahu and Suresh Chandra Satapathy

Abstract Social Group Optimization (SGO) is a class of evolutionary optimization technique which is having many applications recently. There have been few modifications to SGO so far. Modified SGO (MSGO) is a new variant suggested by original developer of SGO. In both SGO and MSGO the individual is chosen based on the better fitness between its past value and current value. However, the change in dimensional values is not been considered when both past and current fitness remains same. In this work, the check is done dimension-wise and the individual having different dimensions is retained even though the fitness remains same between past and current values. This improvement is incorporated in MSGO and simulation experiments are carried out with benchmark functions and comparisons are done with MSGO.

Keywords Function optimization · SGO · MSGO

1 Introduction

Social Group Optimization (SGO) [1] is developed in the year 2016 for solving optimization problems. SGO belongs to class of evolutionary optimization techniques. Several researchers have adopted SGO for various applications in healthcare domains and claimed to find better results compared to other evolutionary algorithms [2–4]. Further to these, original developers of SGO have done few modifications of standard SGO algorithms and Modified Social Group Optimization (MSGO) is relatively recent [5]. In this MSGO algorithm an interesting concept has been introduced. The concept is based on the Self-Awareness Probability (SAP) of an individual that

S. S. Sahu (✉) · S. C. Satapathy
School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India
e-mail: saishakti1997@gmail.com

S. C. Satapathy
e-mail: suresh.satapathyfcs@kiit.ac.in

enable him/her to acquire knowledge from better peer member. However, an exhaustive study has revealed that while choosing the best person or more improved person based on knowledge the standard SGO or even MSGO fail to investigate the differences of individual dimensions between the past and current competencies. In our suggested algorithm we have incorporated a check of dimensions and even though a person might have same competency between his/her past and present selection shall be done based on the differences of dimension values. This slight insight has resulted in improvement of the performance of MSGO algorithm considerably.

The rest of the paper is divided into several sections. In Sect. 2 a brief discussion of MSGO algorithm is presented. Section 3 discussed our proposed method. In Sect. 4 simulation results on few benchmark functions are done and comparisons are shown with MSGO with concluding remarks.

2   Modified Social Group Optimization (MSGO)

MSGO is almost like SGO. The two stages of SGO are Improving phase and Acquiring phase. During the improving phase, an individual learns from the best individual of the group. In acquiring phase, an individual not only an individual learns from best peer but also learns from the more competent peer which is randomly picked up. The main objective of both these phases is to finally see that an individual is more competent from its previous state and able to solve complex problem at hand more efficiently and effectively. However, in MSGO there is a small interjection in acquiring phase. And this interjection is based on the philosophy that an individual must have some pre-defined self-awareness probability to acquire the knowledge from other better individual. This philosophy matches with the learning mentality of any individual in the societal metaphor. Unless one is having a self-awareness to learn/acquire knowledge from others, no amount of pressure in learning can be helpful. In MSGO [] a new term SAP is introduced and its value is empirically fixed. The two phases of the MSGO algorithm is reproduced below.

The social group consists of N individuals, so let's call them p_i, $i = 1, 2, 3, …, N$. Each individual p_i is defined by $p_i = (p_{i1}, p_{i2}, p_{i2}, …, p_{iD})$, where D is the number of traits allocated to an individual that ascertain both dimensions additionally f_i, $i = 1, 2, …, N$ is their corresponding fitness value.

Improving Phase

Each social group's best individual (best_P) attempts to spread knowledge among all individuals so that others in the group might become more knowledgeable.

$$[\text{minvalue, index}] = \min\{f(P_i), i = 1, 2, 3, …, N\}$$
$$\text{best}_P = P(\text{index},:) \tag{1}$$

to address significant minimization issue.

Each individual receives knowledge from the best (best$_P$) individual in the group during the improving phase. Every individual's updates could be calculated below:

Algorithm 1: The Improving phase

For i = 1 : N
    For j=1:D
        $Pnew_{ij} = c * P_{ij} + rand * (best_P(j) - P_{ij})$
    End for
End for
 Recognise *Pnew* if that leads to improved fitness *P*

Such that *rand* is just a random number, $rand \sim U(0, 1)$, and c is referred to as self-introspection parameter is located among both 0 and 1.

Acquiring Phase

As is well known, in acquiring phase, a individual of a social group communicates with both the best member (best$_P$) of that group and at random with several other individuals of the group. If the other individual is more knowledgeable then some individual, then they can learn something new from other knowledgeable individual. Since the (best$_P$) is constantly superior to the other individuals, one to mostly learns by the (best$_P$). If someone else has greater information and a higher self-awareness probability (SAP) to obtain there knowledge, the individual will learn something new from them. Self-awareness probability (SAP) refers to the capacity to learn a significant amount from others. Consequently, the modified acquiring phase is stated along with:

$$[\text{value, index\_num}] = \min\{f(P_i), i = 1, 2, 3, \ldots, N\}$$
$$best_P = P(\text{index\_num, :}) \tag{2}$$

If P_i's is the updated value just at completion of improving phase, is used to solve minimization issues.

Algorithm 2: The Acquiring phase

For i = 1 : N
   Randomly choose an individual P_r, such that $i \neq r$
     If f (P_i) < f (P_r)
       If rand>SAP
         For j=1:D
           $Pnew_{i,j} = P_{i,j} + rand_1 * \left(P_{i,j} - P_{r,j}\right) + rand_2 * \left(best_P(j) - P_{i,j}\right)$
         End for
       Else
         For j=1:D
           $Pnew_{i,:} = lb + rand * (ub - lb)$
         End for
       end if
     Else
       For j=1:D
         $Pnew_{i,j} = P_{i,j} + rand_1 * \left(P_{r,j} - P_{i,j}\right) + rand_2 * (best_P(j) - P_{i,j})$
       End for
     End If
  End for
Recognise $Pnew$ if that leads to improved fitness P

Such that $rand_1$ and $rand_2$ both are independent random numbers, $rand_1 \sim U(0, 1)$, and $rand_2 \sim U(0, 1)$. The lower and upper bounds of the respective design variable, lb and ub, with SAP's range of 0.6–0.9 are employed as random numbers that influence the algorithm's stochastic nature.

3   Improvement on MSGO

In this section we introduce the ideas of the improvement done on the MSGO. In standard SGO and MSGO whenever the fitness of an individual is compared with its previous fitness the absolute fitness value is only checked without checking the values of the respective dimensions. There may be a situation that the fitness remains same for both previous and current values but the dimensions may have been differed. In this case it may be beneficial to consider the individual giving same fitness values but with different dimensions. This will provide a new individual to participate in the process. This idea is demonstrated by a simple illustration below.

Let X_i be the current individual having D dimensions. And X_{inew} is the same individual after obtaining values using Algorithm 1 or 2.

f(X$_i$) and f(X$_{inew}$) are corresponding fitness values for Xi and X$_{inew,}$ respectively.
If f(X$_{inew}$) > f(Xi) then Xi is chosen for next step (for minimization problem)
Else if
f(X$_{inew}$) ≤ f(Xi) then X$_{inew}$ is chosen for next step.

If the fitness value of both X$_i$ and X$_{inew}$ is same then X$_{inew}$ is chosen irrespective of its dimensions.

Let X_i = [20 40 60 80] and fitness is sum of maximum of individual = 200.

X_{inew} is now found to be [40 20 80 60] after running Algorithm 1 or 2. In this case the fitness also 200. However, for our improvement version we shall chose X_{inew} to next step for further processing and ignore X$_i$. In the same line of idea, the gbest is also selected in our improvement proposal.

4   Simulation Results and Conclusion

We have taken up few benchmark functions as given in Table 1 for our experiment and comparisons.

Both MSGO and our improvement on MSGO are simulated on four benchmark function as a trial basis and results are reported in Table 2.

The outcomes are reported along the mean fitness and standard deviation (std) over an average run of 10 simulations. This is clearly observed from the table that our proposed improvement on MSGO gives better performance compared to simple MSGO. This is due to the differed dimensions have larger influence on arriving at

Table 1 Benchmark functions

S. no.	Function	D	Range	Formulation	Value
1	Step	30	[−100, 100]	$f(x) = \sum_{i=1}^{D}([x_i + 0.5])^2$	f_{min}=0
2	Sphere	30	[−100, 100]	$f(x) = \sum_{i=1}^{D}x_i{}^2$	f_{min}=0
3	Sum Squares	30	[−10, 10]	$f(x) = \sum_{i=1}^{D}ix_i{}^2$	f_{min}=0
4	Quartic	30	[−1.28, 1.28]	$f(x) = \sum_{i=1}^{D}ix_i{}^4 + random(0, 1)$	f_{min}=0

Table 2 Fitness comparisons of MSGO and our proposed MSGO

Function name	MSGO	Our proposed MSGO
Step	Mean 0 Std 0	Mean 0 Std 0
Sphere	Mean 3.20e−03 Std 2.50e−03	*Mean 7.82e−035* *Std 1.20e−034*
Sumsquares	Mean 1.20e−03 Std 1.50e−03	*Mean 1.22e−31* *Std 1.245e−33*
Quatric	Mean 1.46e−02 Std 8.90e−03	*Mean 2.58e−05* *Std 5.67e−05*

optimal values. This finding of ours can be helpful in further studying similar type of evolutionary algorithms. Though we have found good performance on very few benchmark functions, as a further study this can be extended to many benchmark suits and exhaustive comparisons can be done.

References

1. Satapathy SC, Naik A (2016) Social group optimization (SGO): a new population evolutionary optimization technique. Complex Intel Syst 2(3):173–203
2. Naik A, Satapathy SC (2020) A comparative study of social group optimization with a few recent optimization algorithms. Complex Intell Syst. https://doi.org/10.1007/s40747-020-00189-6
3. Dey N, Rajinikanth V, Ashour AS, Tavares JMRS (2018) Social group optimization supported segmentation and evaluation of skin melanoma images. Symmetry 10:51. https://doi.org/10.3390/sym10020051
4. Dey N, Rajinikanth V, Fong SJ et al (2020) Social group optimization-assisted Kapur's entropy and morphological segmentation for automated detection of COVID-19 infection from computed tomography images. Cogn Comput 12:1011–1023. https://doi.org/10.1007/s12559-020-09751-3
5. Naik A, Satapathy SC, Abraham A (2020) Modified social group optimization—a metaheuristic algorithm to solve short-term hydrothermal scheduling. Appl Soft Comput. https://doi.org/10.1016/j.asoc.2020.106524

Facial Recognition System with Secured Dynamic Implementation and Time Restriction

Abhilash Budharapu, Mohammed Amaan, Poojith Ramagiri, and K. Krishnaveni

Abstract In today's era, there is an increasing need for improved security measures as privacy and security are significant issues in information systems. Security plays an important role in everyday life examples in offices, institutions, laboratories, houses, banks, ATMs, and others. With the increase in prominence of automation and security, facial recognition systems are widely being used at entrances for human identification and access particularly for security functions. The already existing facial recognition system can be improved by making it dynamically implementable and time-bound by adding additional security and operative measures which has been focused in this paper. Facial recognition system captures human images to compare with stored database images to give authorization to the user to pass through the door. However, in the majority of systems, it is necessary to create the database for the photos of persons in advance before using them. The proposed system has secured dynamic face storing, where the images of a new user can be stored on-site without any security complications, which is secured using WhatsApp and OTP request to the authorized person(s). Additionally, a system enhancement that can be used in a variety of situations is time restriction.

Keywords Face recognition · Secured · Dynamic implementation · Time-bound

1   Introduction

Over the past few years, there have been several developments in the convectional technology and biometric technology to meet security needs for household and offices [1]. Because of the reliability issues with traditional security measures like keys, passcodes, and RFID cards, there is now a demand for biometric systems [2]. These biometric systems are developing rapidly because they can be used for identification

A. Budharapu · M. Amaan · P. Ramagiri · K. Krishnaveni (✉)
Department of Electrical and Electronics Engineering, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, India
e-mail: krishnaveni_eee@cbit.ac.in

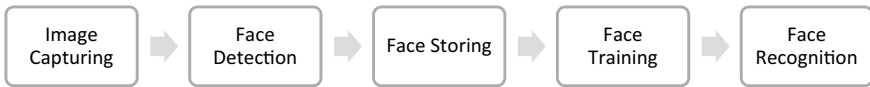| Image Capturing | → | Face Detection | → | Face Storing | → | Face Training | → | Face Recognition |

Fig. 1 Development of a facial recognition system

and verification; biometrics have characteristics that cannot be faked as they are inherently different from human to human so that their uniqueness is guaranteed [3]. One such biometric system which is prominent is facial recognition. Facial recognition is a way of recognizing human faces through technology. It maps facial features from photographs or videos to the information present in the database of known faces to find a match. If the captured image matches with admin's image, then door is unlocked [2]. The study of Python and OpenCV and their in-built library functions can be used to code an effective facial recognition system. The development of a facial recognition system can be classified into steps (Fig. 1).

To begin the process of faces recognition, there needs to be a dataset of images created and trained features of this dataset to perform the face matching. OpenCV utilizes the device's webcam to record live footage of the person.

This is followed by face detection which presently plays a critical part as the primary step [4] in numerous key applications such as face tracking, face analysis and face recognition. The face detection application utilizes viola-jones algorithm and machine learning (ML) to discover human faces inside bigger picture, which frequently consolidate other non-face objects such as landscapes, building, and other human body parts. The algorithm regularly begins to look for human eyes—one of the most effortless highlights to identify, the mouth, nose, nostrils, and iris in a cascaded approach. Once the algorithm concludes that it has found facial locale, face is identified.

Face storing involves storing of multiple images of the person's face [2]. These typical face pictures are converted into gray scale pictures and put away in specific framework folder called dataset, and afterward they are utilized for assistance in feature extraction and training.

The data is then trained using machine learning algorithms to extract critical facial traits of people which are unique for recognition process [4]. The model does this by making records, which contains all information features within the frame of networks and values of each feature are further calculated and utilized.

Face recognition is the final step where the camera is used to detect the images, and the images are coordinated with the database of facial traits using a distinguishing proof, after a coordinate is found the door is unlocked, with their names shown on the screen. Faces are recognized based on the confidence/accuracy from 0 to 100%. The accuracy increments with an "Optimal Dataset," if no individual is found within the dataset it'll be regard it as "Unknown," and the entry is restricted. Local binary pattern histogram (LBPH) algorithm is used over other algorithms such as Fisher Confront, Eigen Faces due to its superiority and accuracy.

The dynamic face storing feature ensures that it can be deployed on-site without any preplanning. Additionally, the system creates a secured-room environment where

a person can stay inside for a limited amount of time specified by the admin, where the time starts ticking as soon as the person enters. The person needs to exit through the same door before the timer runs out, if failed to do so the person gets trapped inside until further actions by the authority. This system can be used in ATM's, bank lockers, restricted entrances, and secured office rooms, where a person would typically stay inside for short durations of time. Implementations of this system ensure maximum safety, reduce thefts, and prevents unnecessary strolling of people inside secured rooms. It uses Python software to carry out the entire process. The proposed system ensures implementation of secured restricted environments rapidly and robustly.

1.1 Disadvantages/Drawbacks

Facial recognition system being such effective still has some notable disadvantages.

For the first time, a facial recognition system addresses a face to register it and associate it with an identity, in such a way that it is recorded into the system. The great difficulty is ensuring that this process is carried out in real time in a secured manner. Systems need to be fed with the dataset of all its clients before being deployed on-site. Not many institutions have the resources and time to set up the entire process of storing images which hinders them of using facial recognition systems.

Facial recognition can be made even more secured and advantageous by limiting people's moments inside secured environments by making the system time-bound this ensures safety and minimizes thefts. This time-bound facial recognition system can be used in secured biometric attendance; automatic attendance can be granted by verifying in–out time of employee if they meet the company hourly time limit. Attendance manipulations, where clients enter, give their biometrics and leave which can be effectively prevented since their attendance is based on their premises time.

There can occur a situation where a client's face it not accurately matched and needs emergency access which cannot be achieved by traditional facial recognition systems.

2 Proposed Facial Recognition Model

In this paper, the authors propose a facial recognition system, which is dynamically implementable and time-bound that can overcome the above-mentioned drawbacks, makes the system secured, robust, and to have improved applications (Fig. 2).

The entirety of facial recognition process remains the same except for these additional features to face storing and face recognition steps.

Fig. 2 Flow diagram of the facial recognition system

3 Secured Dynamic On-Site Implementation

Dynamic on-site implementation results in deployment of facial recognition system without any prior planning and resources, which is quick, effective and completely automated without the need of human expertise.

For the system to be completely automated, it is necessary to have a friendly interface for the users/clients this is achieved using Python's Tkinter module. Tkinter is a standard Graphical User Interface (GUI) library for Python which can provide simple and quick interface solutions. By using Tkinter GUI application fundamental window with various widgets according to user requirements are created (Fig. 3).

For the system to have secured dynamic implementation, we make use of WhatsApp and One-time password (OTP) request. To register a new user/client, all their details along with their unique verification ID (Employment ID) is sent through the client's WhatsApp to admin. Selenium is a web scraping package in Python which supports several browsers/drivers, this is used for accessing and sending information through WhatsApp. The registering process after taking the details of the user/client opens a WhatsApp web QR code, which when scanned sends the details using the message box. At the same time an OTP generation code is used to generate a random four-digit code which is delivered to admin. Text messaging is considered as secured and ubiquitous channel of communication for OTP; sending of text message is done using Twilio service. Twilio is an American cloud communication platform which

(a)   (b)

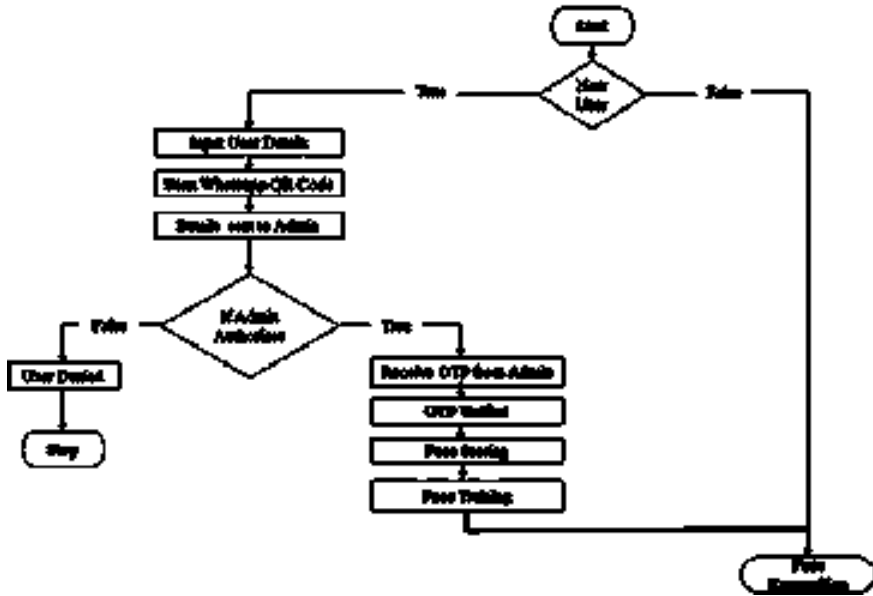**Fig. 3 a** Tkinter module user form and **b** admin form

Fig. 4 Flowchart for secured dynamic on-site implementation

allows software developers to programmatically make and receive phone calls, send, and receive text messages. Its built-in Python modules and freely available resources make it effective in sending the OTP to admin's mobile number. Verifying the authenticity of the client, the admin shares the OTP which grants access to store images. An optimal dataset of 20–30 images of face is mapped with the client's unique employee ID and is stored into the drive for further facial recognition process (Fig. 4).

4 Time-Restricted System Implementation

Limiting the facial recognition system with timer ensures maximum safety and applicability. It creates a secured-room environment, implementation of strict visitor policy by continuous time monitoring prevents unnecessary public wandering thereby reducing thefts.

Threading module in Python helps us make the system time-bound. This module is used to add concurrency to programs, multiple threads/tasks can be simultaneous executed.

A timer program is threaded with facial recognition system which establishes a countdown of specified time set by the admin as soon as the recognition starts working timer starts getting displayed on the screen. The person inside is notified couple of times using an alarm as the time depletes. When the time runs out, the timer program
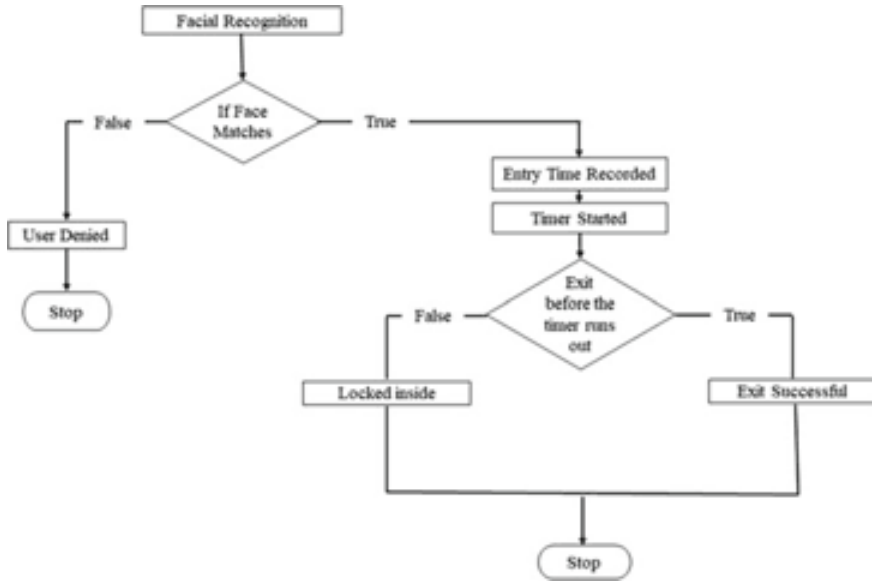
Fig. 5 Flowchart of time-restricted system implementation

sends command to completely shut down the facial recognition process, leaving the person trapped inside until further actions by the authority (Fig. 5).

5 Results

Selenium sends user details by accessing WhatsApp; after being verified and authenticated through OTP by admin, face images of the user are stored into the disk (Figs. 6 and 7).

Timer programs gets executed simultaneously with face recognition and shutdowns the system when the time depletes (Fig. 8).
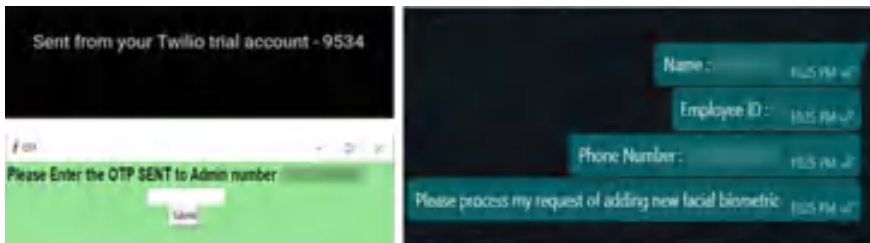
Fig. 6 OTP authentication

Fig. 7 Storing of facial dataset

Fig. 8 Timer
implementation


```
You have 15 seconds time to scan your face
Sorry scanning time limit reached
Entry time 2021-05-17 23:04:45.049860
EXIT ROOM BEFORE 00:01:02
```

6 Conclusion

The proposed system ensures implementation of already existing facial recognition system without any preplanning and need of human expertise. It can be quickly and effectively implemented on-site without the need of any resources. Adding a two-factor authentication using WhatsApp and One-time Password (OTP) maximizes the security. Door access audit trail can track people entering and leaving by recording in–out time. By restricting the system with time, helps in preventing thefts and ensures maximum safety. The working hours of the worker can be easily registered into log file and attendance based on this can be provided. Numerous applications of time restricted secured rooms can be implemented in ATM's, bank lockers, premises entrances, offices, and institution.

References

1. Umm-e-Laila FA, Uddin SS, Sadiq MS, Siddiqui MS, Ibrar ul Haque M (2020) Real time door security system with three point authentication. J Mechan Continua Math Sci 53–64

2. Krishna Vamsi T, Charan Sai K, VijayalakshmiM (2019) Face recognition-based door unlocking system using Raspberry Pi. Int J Adv Res Ideas Innov Technol
3. Irjanto NS, Surantha N (2020) Home security system with facial recognition based on convolution neural network. Int J Adv Comput Sci Appl
4. Syafeeza R, Mohd Fitri Alif MK, Nursyifaa Athirah Y, Jaafar AS, Norihan AH, Saleha MS (2020) IoT based facial door access control home security system using raspberry pi. Int J Power Electron Drive Syst 11:417–424

Modelling and Simulation of a Hybrid Electric Vehicle with the Electric Power Train

K. V. Dhanalakshmi, A. N. Venkateswarlu, Mahaboob Shareef Syed, and S. S. Tulasi Ram

Abstract Hybrid Electric Vehicles (HEVs) are greater in demand in the recent instances because the fuel expenses are excessive and also they cause environmental pollutants which has unfavourable results at the dwelling matters. So, it has end up wide spread to analyse the running of the thorough electric powered train to achieve the specified presentation using HEV. The E-motor which has been decided on has capabilities suitable for HEV utility and its modelling in electric energy educate of the HEV is being described in this paper. The E-motor in this example, ever-lasting Permanent Magnet Synchronous Motor (PMSM) is controlled by a 3 phase 2 level Voltage Source Inverter (VSI) whose switching is managed via SVPWM method. A fault sign has been created and passed thru the PI controller by means of which the speed and torque are managed by making use of the rate and modern day responses. The significance and reference perspective is estimated from the twin voltage mentions, that are the results of the PI controllers and it is given to the SVPWM module. The specified voltage as predicted is done via suitable pulses produced by using SVPWM module to the VSI. The modelling and simulation were carried out the use of MATLAB Simulink and the planned electric power educate is unpolluted and effective for possible utility in hybrid electrical vehicles.

Keywords Electric powered strength-train · PMSM · 3 phase 2 level VSI · SVPWM manipulate

K. V. Dhanalakshmi (✉) · S. S. Tulasi Ram
GNITS, Hyderabad, India
e-mail: kv.dhanalakshmi@giet.edu

A. N. Venkateswarlu · M. S. Syed
VLITS, Guntur, India

645

1   Introduction

Traditional mobility systems (automobiles) are driven via inner combustion engines (ICE) and as a consequence they burn fuel, petrol into carbon dioxide there with the aid of affecting the surroundings [1]. Hybrid electric vehicle era is the speedy growing technology which targets at low consumption of gasoline at the side of extended performance and low poisonous waste. The main modules of the electric energy train of a HEV are the traction motor, inverter for monitoring the motor and the manipulate unit to manipulate the output of the inverter that is based at the cutting-edge motor working factor and the demand of the purchaser. There are several styles of traction motors while thinking about HEV makes use of inclusive of DC motor, induction motor, PMSM, and switched reluctance motor. In a trendy EV or series HEV device, there is only one propulsion unit, that's the electrical motor [2]. PMSM is the high-quality traction motor for HEV utility as it has quicker response, larger electricity density, and higher effectiveness because of decreased rotor losses. The main drawback of PMSM is its charge, because the rotor coils are substituted with the rare earth magnets which can be too costly. Induction Motor is a really well-known AC motor [3]. Additionally, for the manipulate of the inverter, SVPWM technique is used. On this paper an effort is prepared to ideal the whole electricity train for a HEV in MATLAB—Simulink which includes exhibiting of PMSM, VSI, and SVPWM control. PM synchronous machines (PMSMs) with magnets on the stator are of latest starting place and they're protected underneath hybrid machines. The precept of operation of synchronous machines with PM rotors is presented within the direction of which machine fundamental associations consisting of the magneto cause force (mmf), brought on emf (also called back emf), and torque are derived for PMSMs with sinusoidal-triggered emf however they can be prolonged alongside the identical lines to brushless dc machines with trapezoidal-caused emf [4].

Standard HEVs encompass an internal combustion engine (ICE), electric motor (EM), single or couple of electricity storage structures (ESS), power electronic converters, and controllers [5]. The paintings additionally offer the primary investigational outcomes to help theoretical derivations and simulations effects. Consistent with a forecast by way of international electricity business enterprise, using electric powered motors will grow from three million to one hundred twenty-five million by the year 2030. This is nearly forty-one times of what it's far nowadays, with the increasing call for of fossil gas and problems with pollutants it seems most possibly to appear. Owing to that, all main IC Engine car producers like Ford and GM are slowly turning their interest in the direction of the electric vehicles. The marketplace and customers are in want for a cheaper personal transportation or even on top of that, the government has begun assisting electric powered motors via its policies (Fig. 1).

The Engine of a conventional IC Engine vehicle is replaced by using an electrical Motor and the gas tank is changed by the Battery p.c. Of all the components most effective the Battery p.c. and Motor on my own contributes to about more than 50% of the full vehicles weight and the fee. As you may see the Battery percent, Battery
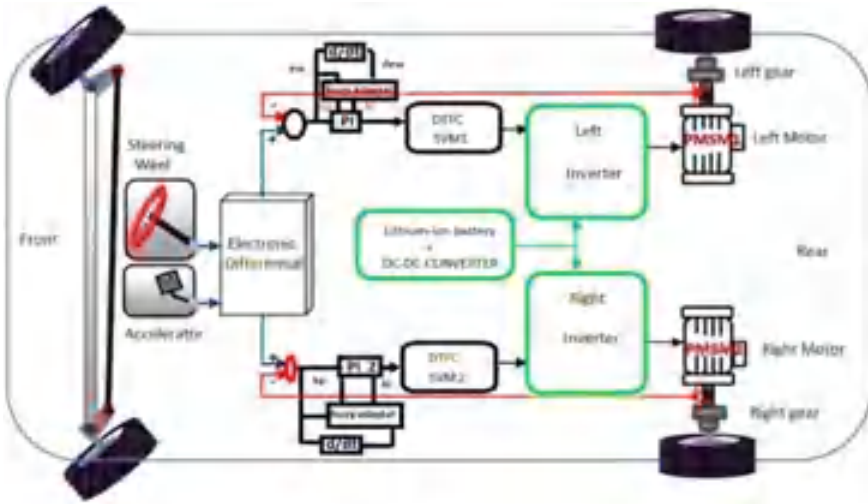
Fig. 1　Block diagram of EV

management device (BMS) Controller, Motor and the Transmission unit bureaucracy the important additives in an EV.

2　Proposed System and Design

A new kind hybrid energy educate system become established for HEV, that work as electricity coupler and transmission machinery through 3 input shafts, one output shaft and stuck transmission ratios. The HEV implementing an innovative device can comprehend 3 action modes: hybrid strength using mode, pure electric powered propulsion mode, and power generation method at car forestall circumstance. It could make engine work at most effective gas intake area, make use of motor to deliver great begin torque and recognize methods conversion change at feasible rotation velocity. The complete shape of this gadget is easily produced, that could encounter vehicle driving necessities at complex conditions.

Narration of entire EV scheme is brought; and the concert of this one strength is being evaluated, now which took a look at seat that is planned for exhibiting functions of minor EVs furthermore, the energy educates of EV and HEV systems had remained supplied by means of dual rotor everlasting magnet synchronous device (PMSM), in which a whole measurement strategy technique and electricity control system were evolved. Thinking about the profitable EVs and car providers, induction system (IM) and PMSM are the greatest typically cast-off in EVs. The judgement among commercial and electric vehicle motor drives may be discovered. Floor straddling PMSM is cast-off for light EV. Additionally, field orientated manipulate (FOC) of

PMSM energies is carried out for EV/HEV software. The accomplishments of the examine stand with appreciation to the incorporation of a whole powertrain of an EV. This electrical powered power gadget consists of a hard and fast of 2 SM-PMSMs and their drive and regulator structures.

Speeding up test of the PMSM and car A examination of the speeding up and most speediness of the P.M.S. MS is carried out at great speediness short of load. This take a look at is completed in the lab the use of the twisting force speeding up pedal. It is observed that the motor rapidity is elevated from dead end to about 4510 rpm in 4.3 s. Formerly, in keeping with the automobile's velocity turns into 31.2 m/s (in 4 s). In a speeding up examination of the P.M.S. MS whilst the automobile is raised is shown (i.e. PMSM is inner of car). It is able to be observed that the motor pace is grasped approximately one thousand rotations per minute in 5 s. In keeping with, the automobile's speed is 5.73 m/s in 4 s. The usage of the acceleration of the vehicle is 1.396. It could be observed that the two speeding up standards are greater than 1.17 in since the PMSM turns at no-load then the automobile is lifted.

Settling Examination of the Battery for the intentional EV, the electrical energy scheme is determined from 100.2 Vdc battery platform. Hence, the expulsion current rate for this battery-operated is verified in the research laboratory to discover its ability of supreme expulsion rate. By an estimated fixed D.C. voltage, the E.S.S. is verified at dissimilar loads to confirm its power for EV at diverse driving conditions. Refereeing to the "battery current control" in Fig. 2, the battery current is directed at 40.1, 60.1, 80.1, and 100.2 A. Throughout these trials, the D.C. voltage is roughly continued constant at 100 Vdc. The Driving Test Using the IM240 Dynamometer in this unit, the whole EV organization is verified underneath diverse load conditions.

The photos of E.V. on the Framework Dynamometer gadget (CHS). Display of CHS is cast-off to match the extraordinary street burden riding cycles and to expose the slow pace, power, length, and so forth. An automobile is examined at beginning at
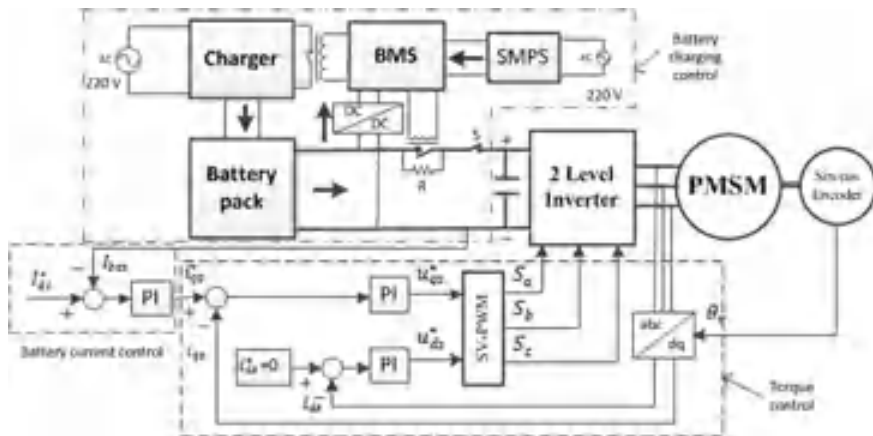

Fig. 2 Charging and current management of battery pack

Fig. 3 Experimental driving test using Im240 dynamometer

the same time as the battery is about complete fee. Whilst the vehicle runs at unique pace ranges, the SOC is regularly reduced. It is able to be observed now there are slowing times, the SOC is not multiplied and it's miles simply flat. This occurrence is befalling due to the fact the reformative braking approach is nowhere implemented but, as well, the rate is about to reach the graded fee, i.e. 60/five), as portrayed, but, the acceleration isn't always verified. The reason is whilst the PMSM is strolling at great speeds, a horrific presentation of the PI controllers is located and therefore the modern apparatuses be afflicted by exact excessive ripples. The scarcity is devoted inside the prolonged effort (Fig. 3).

The possibility of the q-axis current of the PMSM and extreme advanced energy from the two PMSMs. The current module is around 100.1 A, while the energy is static at 21.52 kW. It is observed that these standards are nearby the stipulations of P.M.S. MS that are listed in Table 1.

3 Hybrid Battery/Ultra Capacitor E.S.S.

In the lab, a UC element is incorporated in equivalent to the B.P. via DC/DC converter to stake a still load, the performance of drive drift of charge and liberation methods is practically verified. By the trial consequences while best battery-operated is cast-off, it's miles initiate that the general efficacy is zero.95 and zero.8 at little and precise excessive currents, respectively. The efficacy of battery shrinkages at very excessive currents because of its inner resistance impact. The UC can be stimulating at some point of two methods, regenerative braking and charging from the battery at mild tonnes. Whilst the UC is amplified to the battery, the general efficacy of indicting from the battery and discharging to the load being 2 * (0.ninety six * zero. Ninety five) = 0.83. Charging of the UC throughout reformative method has no longer been

Table 1 PMSM limits

Varying quantities	Sign	Rate
D.C.—Buses volt [V]	V_{dc}	102
No. of poles	P	7.9
Graded voltage [V]	V	72.3
Rated current A.C./D.C. [A]	I	125.3/180.2
Max. current A.C./D.C. [A]	1.2 min	420.1/600.1
Stator resistance value [Ω]	R_s	0.012567
Stator inductance value [mH]	L_s	0.102567
Stable flux relation	λ_{PM}	0.02576
Inertia constant [kg m^2]	J	0.004567
Graded/max motor speediness [rpm]	ωN	3000.2/5000.3
Torque standard [Nm/A]	K_t	0.156
Graded/highest stall torque [Nm]	T	22.3/90.3
Voltage standard [V/Krpm]	K_v	16.75
Graded motor frequency [Hz]	f_N	60.2
Graded/max power [kW]	P_N	12.2/30.2
Sample frequency [kHz]	f_s	10.3
Efficacy at 24.1–96.2 V.D.C.	η	92%

verified for this reason, this one is able to be decided that at excessive present day float, the UC complements the power float. Future paintings can be committed to enhance performance at together charging and is charging methods the usage of the UC/B.P.-E.S.S. presently, indicting of the UC at the reformative decelerating method is taken into consideration.

4 Simulation Diagrams and Results

In this chapter, for the persistence of verifying the usefulness of the mathematical model developed and algorithms adopted in the thesis, numerous simulation results are presented. The steady-state features of a PMSM are analysed. Moreover, the variable-speed drive analysis is carried out. Simulation models are implemented in MATLAB/SIMULINK, operating at a sample frequency of 10 kHz. The motor limits used in the imitation are tabulated below in Table 2.

This section defines diverse tools accessible for electrical and electronic systems imitation and then explanation is set for choosing imitation for the PMSM system [12]. Lump by lump a clarification is specified for Simulink imitation of the drive system [12].

Table 2 Motor parameters

Parameter	Unit	Value	Description
P_n	W	30,000	Rated power
P_{max}	W	75,000	Peak power
V_n	V	215	Rated voltage (AC, phase-to-phase)
I_n	A	150	Rated current (RMS)
I_{max}	A	430	Peak current (RMS)
ω_n	RPM	4775	Rated speed
ω_{max}	RPM	13,000	Maximum speed
T_n	Nm	60	Rated torque
T_{max}	Nm	175	Peak torque

Line voltage
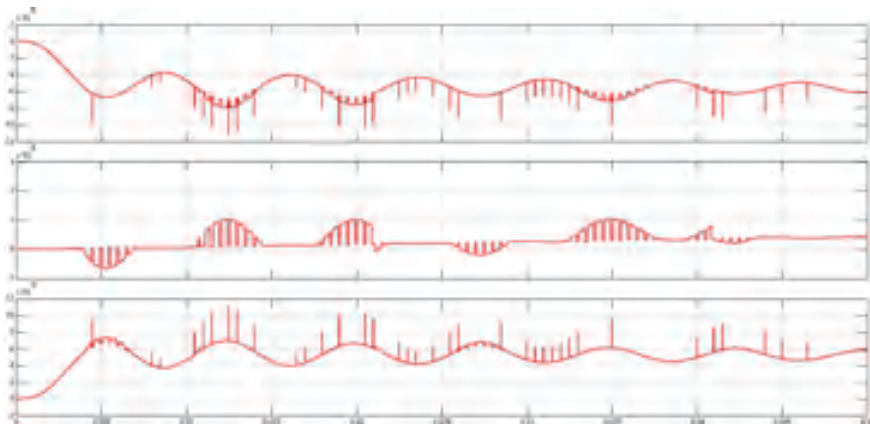
Scale—*X*-Axis—Time in seconds

Y-Axis—Line Voltage

The power and energy outline as step idea is used as input. Originally the input power drained from the battery is great that is owing to early great load condition, mostly owing to the hastening required. Motor has to deliver power to overwhelmed force and similarly relative power is strained from the battery. As the vehicle rapidity spreads steady state, hurrying becomes zero and input, output power become theoretically equal. Similar reply is detected in instance of Dynamism input and output.

Test input for inverter model

Scale—*X*-Axis Time in Seconds

Y

-
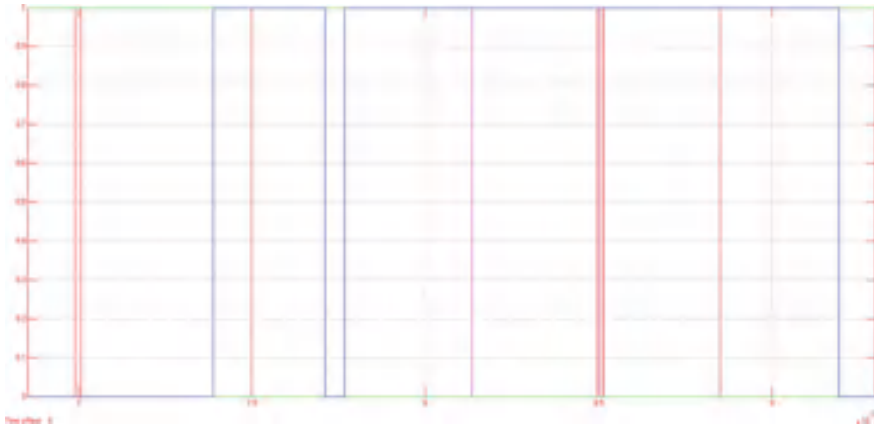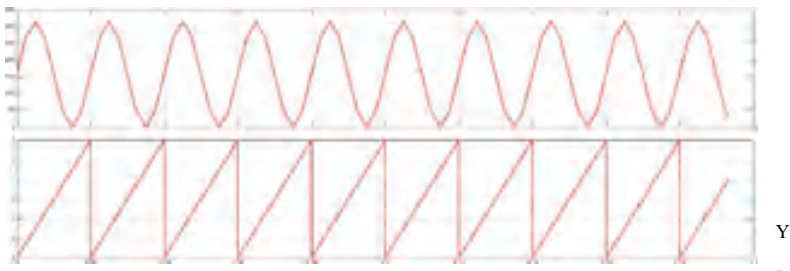
Y-Axis Inverter input Pulse

Speed

Scale—*X*-Axis Time in seconds

Y-Axis Speed in Rpm—1 p.u

The Voltage on load side is shown which is seen not to be pretentious by the speed or load changes. This shows that the controller and battery are capable of maintaining voltage constant. The load voltage of the scheme halts throughout the operation.
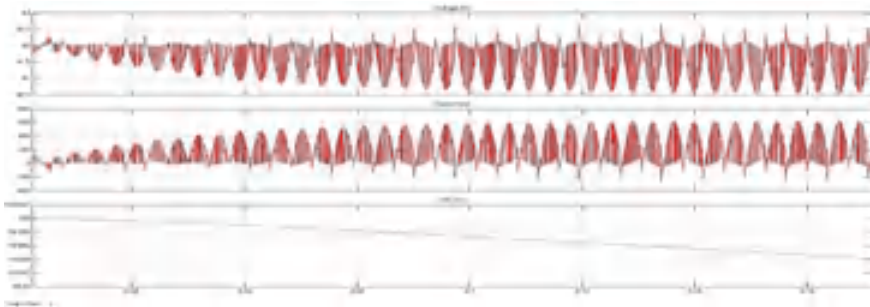
Torque
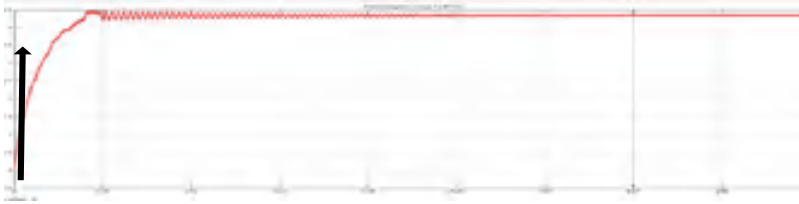
Scale—*X*-Axis Time in Seconds—0.09 sec

Y-Axis Torque 4.5 (N-M)

The torque production of a motor is the quantity of rotating force which the motor grows. The torque of a tiny electric motor is usually measured in Newton's Meters or other directly transformed units of quantity.

Battery outputs

Scale-*X*-Axis—Time in Seconds 0.17 s

Y-Axis—Y1-Battery Voltage 42-Nominal Voltage Y2-Battery Current 600 amps

Y3-Battery SOC 10

5 Conclusion

The proposed work deliberated the demonstrating and imitation of entire electric energy teach of a HEV. MATLABSIMULINK has been used to run the simulation at diverse points along with the demonstrating. The suitable motor is chosen built totally on HEV utility, for this the PMSM was designed replicated and verified. Then the usage of the voltage equation for the 3 segment 2° VSI, along with S.V.P.W.M. control algorithm, become applied in Simulink and examined one by one, ultimately, the PMSM and the inverter unit became blended and specific control

became attained the usage of speed and present day feedbacks. The centre for this combined module turned into sensibly precise to display the rate and torque response for a HEV in order that the predicted torque and speed is produced by means of the simulation. The projected mixed unit has been verified to be operative for feasible utility in HEV. Effectiveness of the device can nonetheless be more desirable by way of selecting a hybrid PMSM and additionally thinking about the damage version within the imitation.

References

1. Zarma TA, Galadima AA, Aminu MA (2019) Review of motors for electrical vehicles. J Sci Res Rep 24(6):1–6
2. Rahman et al (2000) An investigation of electric motor drive characteristics for EV and HEV propulsion system. SAE technical paper series, Paper #2000-01-3062
3. Bhatt P, Mehar H, Sahajwani M (2019) Electrical motors for electric vehicle–a comparative study. In: Proceedings of recent advances in interdisciplinary trends in engineering & applications (RAITEA)
4. Krishnan R, Permanent magnet synchronous and brushless dc motor drives. PHI, Ch 9, pp 518–555
5. Fitzgerald AE, Umans SD, Electrical machinery, 6th edn. TATA McGraw, pp 245–293, 578
6. Pillay P, Krishnan R (2008) Modelling of permanent magnet motor drives. IEEE Trans Ind Electr
7. Hong W, Chol C, Lee W, Hong JP, Kum D (2009) Advanced permanent magnet motor drive modelling for automotive application. J Autom Technol 10(4):489–495
8. Ranganathn VT (1997) Space vector pulse-width modulation, a status review. Sadhana 22(6):675–688
9. Cao Z, Mahmoudi A, Kahourzade S, Soong WL (2021) An overview of electric motors for electric vehicles. In: 2021 31st Australasian Universities power engineering conference (AUPEC), pp 1–6. IEEE
10. Bose BK, Modern power electronics and AC drives, pp 224–240
11. IEEE 2014 annual international conference on emerging research areas: magnetics, machines and drives. 24–26 July 2014
12. Khamesra K (2013) Performance investigation of intelligent power module based PMSM drive. Published by MPUAT, Udaipur
13. Karki A, Phuyal S, Tuladhar D, Basnet S, Shrestha BP (2020) Status of pure electric vehicle power train technology and future prospects. Appl Syst Innov 3:35. www.mdpi.com/journal/a

A Simple Policy Update Method for the Sharing of Privatized Personal Health Information

Anwar Basha Shaik, Raj Anand Sundaramoorthy, and Nirupama Panabakam

Abstract A few medical services suppliers utilize electronic individual wellbeing records (PHRs) to empower individual patients to deal with their own wellbeing information in a vigorous and versatile climate, especially with the high adaptability and openness of information re-appropriating conditions like distributed computing. On the other hand, protected health records (PHRs) contain extremely private and private information. In addition, PHR owners ought to be able to customize a secure and adaptable access policy for their outsourced data. In addition to basic authentication, existing commercial cloud systems frequently include symmetric or public key encryption as an optional feature to protect users' data confidentiality. Nonetheless, due of the great key administration above of symmetric encryption and the high upkeep cost of keeping up with various duplicates of ciphertext for public key encryption arrangements, such old-style encryption calculations are not great for information reevaluating environments. A lightweight access policy update and a secure, fine-grained access control system for outsourced PHRs are the focus of this study. The ciphertext strategy quality-based encryption (CP-ABE) and intermediary re-encryption are the underpinnings of our proposed technique (PRE). We also offer a policy versioning strategy that makes it possible to trace every policy change. Finally, in order to demonstrate that the suggested strategy works, we carried out a performance evaluation.

Keywords Encryption · CP-ABE · Proxy · Policy updating method · Control and access methods

A. B. Shaik · N. Panabakam
Department of CSE, Vemu Institute of Technology, Pakala, Chittoor, Andhra Pradesh 517112, India
e-mail: sb.anwarbasha@gmail.com

N. Panabakam
e-mail: nirupamacse@vemu.org

R. A. Sundaramoorthy (✉)
School of Computing, Sastra Deemed University, Thanjavur, Tamil Nadu 613401, India
e-mail: rajanand@cse.sastra.ac.in

1 Introduction

Due to the advancements in the technologies like Internet of things (IoT), Artificial Intelligence, Machine Learning, Deep Learning, and Data Science, the industries have vast applications [1, 2]. The IoT is the enabling technology for the technologies such as Data Science, Bigdata, machine learning, deep learning, and artificial intelligence. Before industry 4.0 the IoT is anywhere, anyplace, any time, and anything, after industry it becomes as everywhere, everyplace, every time, and everything computing. The major outcome of this technology is to get volume amount of data from everything and to automate things. As everything is connected to the Internet, huge amount of the data from that connected things will be collected and by adding the intelligence to that data, the works can be automatized. The IoT is combination of both the hardware and software components, the major hardware components are, the sensors, microcontrollers or microprocessors, and actuators. To handle with the hardware components, the user need to have the knowledge on the specific hardware components, for example the pin description of the sensors and microcontrollers and microprocessors, the power that need to be supplied and connections that to be given. User must have knowledge on capacitors, resistors, transistors, relays and power converters, etc. In major, need to know, how to use the multimeter. So that it will be very useful to connect the hardware components and to test the developed hardware kit and to trace the wrong connections. The software components such as firmware in c or python programming languages, cloud computing and the user interface need the website or mobile applications. The open-source platform, Arduino IDE can be used to write the code and upload the compiled program into the microcontrollers. Make sure to select the port number and the board name, to upload the code successfully. In hardware components, the sensors are used to collect data from the things, the sensors are like thermistor to get temperature, moisture sensors that which is used in agricultural fields for equal distribution of water, gas sensors to know the air and sound pollutions and camera sensors to capture the things, the major role of the sensors is to get data from the particular thing. The sensors are input devices in IoT technology. To processors the input data, the processors such as microcontrollers and microprocessors will be used. The microcontrollers are specific and microprocessors or generic. The names of few microcontrollers are Arduino and its family boards, Intel Galileo and its family boards, spark fun and nodeMCU and its family boards. The best microprocessor for Internet of things is the raspberry pi [1, 2]. The python and embedded c programming languages can be used to program to this boards. For raspberry pi, has an operating system such as Raspbian PI and windows. It a open source and can be easily used for doing the real time projects. The actuators are such as AC, alarm/buzzer, fan, light, and gate valve. Based on the different applications in the different fields the sensors, microcontrollers or microprocessors and actuators can be a choice. For example, for the application in the agricultural fields, the moisture level sensor for knowing the water percentages in the fields and DHT11 sensor for knowing the temperature and humidity in the field and the water level sensors to know the level status of the water can be used, based on these sensors input values

the tap/gate valve sensor can be switch on/off for the equal distribution of the water and saving of the water can be done. Similarly for the application in the health sector, the ECG, BP, sugar, and heart beat sensor will be used for continuous monitoring the health condition of the patients, specifically for chronical diseases patients. Like this based on the application in various fields the components will be selected. The IoT is a technology, used to connect the things around us with the unique IP address to the Internet. This is used to connect everything to the Internet and can be remote control and monitor from every corner of the globe. With this technology the data will be collected from the things and will be stored in the data storage device. The fog/ edge computing will play a crucial role to store the data. It is a server which will be placed near to the sensing devices, to do the immediate actions. On top of that globally the cloud will be used. The fog will be the local computing and the cloud will be the global computing and storing devices. On the data, that which collected by the IoT nodes that which is stored in the cloud the deep learning and machine learning algorithms were applied to do the classification/prediction of the data, to extract the interesting patterns. The artificial intelligent is the super set of all the technologies which is used to enable the human behavior with the machines. Machine learning is the subset of the artificial intelligence, used to train the machine. The training is done by the past data, based on the past data the algorithms will classify or predict the data as per the class labels. The past data may be in the form of the datasets, the datasets may get from the sources such as UCI repository, Kaggle, and data.gov. the learning techniques may be a supervised, un supervised, and reinforcements methods. In the supervised learning methods, the class labels is known and we need to classify the data based on the labels. In the unsupervised learning the class labels are unknown and need to classify or predict the data, this is done by identifying the similar and dissimilar clusters and the reinforcement learning is a reward and punishment-based learning methods. With these methods the machine learning will be done. The deep learning is also a subset of the artificial intelligence, the major difference between the machine learning and deep learning is the feature extraction and classification/ prediction will be done in 2 steps in ML and one step in the DL. This all together has vast applications in all disciplines. Almost all the fields like agriculture, healthcare, food industries, construction, and finance has very resulted applications. The agriculture has applications like, remote monitoring the fields, equal distribution of the water, automatic detection of the intruders in the fields and remote controlling and monitoring the crops and early prediction of the dieses. Similarly, healthcare also has applications such as remote assisting the patients, remote and continuous monitoring of the chronical diseased patients. Due to these advanced applications, the security also a most concern with this kind of applications. The major problem of this kind of advanced applications is data theft and the impersonate attack. For this issues lot of research is undergoing and the current work is focused on this work. With the IoT technology the data can be collected from the various sources and that should be stored in the centralized servers. Particularly, in the healthcare applications, the sensors will be embedded in the or near to the human body to collect the data related to the diseased person, the data such as body temperature, blood pressure, and sugar levels. The data will be generated in un-structured and huge. To store this data, it

should be stored in the central place, for the storing of the data the cloud technology is used in the current work. As the patient's data is sensitive, it should be disclosed to others, so need to provide security for the patient data, here the data storage security, while communication to others as well the data should be secured, the current work is focused on this area. Many researchers have done work on the data security, here the security is particularly done in the field of the healthcare for storing the patient's data.

The outsourced server in a cloud storage system or other outsourced data sharing environment must always be accessible for unrestricted access to shared services and data. Due to the cost savings and effective resource management offered by cloud providers, a lot of people and businesses now prefer to store sensitive data on outsourced servers like cloud storage. For reasons of privacy and security, the majority of data owners encrypt their data before sending it to a cloud server [3, 4]. The best way to stop accidental access is to encrypt sensitive data. Encryption, on the other hand, is not enough to guarantee strict security management. Another type of security barrier that is frequently required is an access control system. To address this issue, attribute-based encryption was developed.

ABE has been used in many works [3]. One-to-many encryption and fine-grained access control are available in ABE. It can also control who has access to data and encrypt it. Two types of ABE (KP-ABE) are crypt text-policy attribute-based encryption (CP-ABE) [4] and key-policy attribute-based encryption (KPABE). Characteristics are used to create the user's decryption key, and access policy is used to encrypt the data in CPABE. The user key is linked to the access policy by KPABE, and the data is encrypted by a set of characteristics. CP-ABE is leaned toward as far as information encryption requirement since the information proprietor can lay out their own pol-frigid.

The capacity to manage group keys is one of the advantages of using CP-ABE [5]. One of them is the distinction between concrete attributes and abstract properties. Additionally, rather than one-to-one encryption, it offers flexible one-to-many encryption; it is thought to be a possibility for dealing with the problem of safe and fine-grained data sharing and decentralized access control. CP-ABE adds costly overheads such as ciphertext re-encryption, key re-generation, and key re-distribution whenever an attribute is removed or a policy is updated. These revocation and policy update procedures must be carried out with caution due to the substantial propagation effect on the ciphertext and the user decryption key. When there are a lot of users, the costs of calculating and sending the updated key are especially expensive.

The amount of ciphertexts that must be downloaded and re-uploaded from and to the data outsourcing environment determines the critical sector, while policy updates and data re-encryption are the responsibility of the data owner. For a genuine information trade situation, such overheads bring about wasteful execution. When the access policy needs to be updated, encryptors may also not be available. PHRs are taken into consideration. The data owner, such as a patient, can selectively disclose their information to anyone they wish. In this study, we demonstrate how to update CP-ABE access controls without having to re-encrypt the data to provide effective encryption, improved data access, and improved policy performance. We use symmetric

encryption data to encrypt the data because it has better encryption performance, and the CP-ABE method is used to encrypt the symmetric key [5]. Since we have been together, the CP-ABE method should be used to encrypt the symmetric key. Policy updates only have an effect on the encrypted metric key. Therefore, no additional encryption is required for any cipher texts. On the proxy side, this significantly lowers the cost of calculations. We developed a proxy re-encryption (PRE) protocol to manage the ciphertext re-encryption, which is the main cost of updating policies.

The following is a brief summary of our contributions:

1. We present a paradigm for access control and policy updates for PHRs in the context of multi-authority data outsourcing. Due to our cryptographic construction and the PRE method that we developed, the re-encryption operation is offloaded to the proxy when the policy is modified, while the data owner is responsible for small computations. The expense for both the information proprietor and the intermediary is decreased in light of the fact that to two-step encryption.
2. We provide a policy version control system that enables complete documentation of all update activities and the reconstruction of previous strategies at any time for comprehensive analysis.
3. The PRE system uses parallel programming to parallelize all cryptographic operations. Any ciphertexts that are affected by the new policy in our model will be effectively re-encrypted when the policy is changed.
4. We carry out a performance and security study to show that the method we propose is actually both secure and effective.

The remainder of this work is arranged in the following manner. The background and related work are discussed in Sect. 2. The proposed methodology is discussed in Sect. 3, in Sects. 4 results were presented along with the discussions and finally conclusion is discussed in Sect. 5.

2 Related Works

Policy updating, which reduces scalability and efficiency, is one of CP-ABE's biggest overheads. The data owner must retrieve the data in this instance [3] in order to update the policy by adding, modifying, and deleting records or logical gates, revoking attributes (AND, OR, or M of N) in terms of the policy. All ciphertexts will be updated once the policy is revised, and polices were typically kept on the premises of the data owner. The encrypted files of the affected policy need to be re-encrypted. After that, the ciphertext will be re-encrypted and stored in the cloud. IoT applications for healthcare are also discussed, and security is a big deal [1, 2]. While communication from layer to layer in the IoT-based networks, data security is major concern. While routing the data also most of the data should be secured. Here the secured architecture for the healthcare systems were discussed and proposed [2]. Sharing the personal healthcare information is not a novel method, here with care hiding the sensitive information has been done and achieved the good results in terms of securing the

data [6]. To securing the data, recent technology blockchain as a play a role, while transferring the medical data of exchanging from one source to other source, the data can be secured digitally, based on the block chain technology [7]. The data collected from the wearable devices will be communicated to the cloud, in the cloud with the parameters like cost, safety the security methods were discussed here [8].

Owners of data frequently engage in such activities, which incur overhead for processing and communication. In a PHRs management scenario, patients may need to modify the access policy that encrypts their medical data in order for other doctors at different hospitals to access their treatments.

There are two methods for implementing policy update in the CP-ABE configuration: re-encryption via proxy (PRE) and updating of the ciphertext.

A. **Ciphertext Update**

In this method, data owners are required to generate update keys or tokens and send them to the external server that stores the cryptic texts. Update keys are generated through the calculation of impacted characteristics and will be used to update the corresponding ciphertext components. The linear secret sharing scheme (LSSS), which effectively alters the access matrix and platform when the policy is changed [5, 9–11], is the foundation of the majority of techniques that employ this strategy.

Li et al. presented a speedy policy and file update in the CP-ABE environment in [9]. The data owner's key update modifies the components of the ciphertext. The client's storage and transmission costs are reduced. Additionally, the suggested method is demonstrated to be secure under the presumption of a decision q-parallel bilinear DiffieHellman exponent. However, the data owner must preserve the encrypted parts of the existing ciphertext in addition to the LSSS-based update key generation. Yang et al. recommended a ciphertext updating way to deal with handle cloud server strategy changes in [11–13]. They developed linear secret sharing schemes for adding and removing characteristics in the AND, OR, and threshold gates of ABE policies (LSSS) and examined the cost of policy updates. In this method, data owners must construct update keys using generic order groups. The complexity increases linearly with the number of policy characteristics. Bilinear operations are therefore impossible for devices with limited resources.

In [10], Yuan developed a policy update method based on the matrix update algorithm and integrated it with the encryption algorithm for the fundamental CP-ABE scheme. The data owner must use a ciphertext update method to compare the number of characteristics between the old and new policies as part of the method. The size of the policy determines the cost of calculation and transmission. Varri and others [8] recently suggested using CP-ABE for a multi-keyword search with dynamic policy updates. Instead of selecting a new secret value, this method brings the policy up to date by making use of the encryption data that is already in place. The UpdateKeyGen algorithm is then created in order to update the access policy. Even though this system is effective for ciphertext updates brought on by updates, the cost of calculating an upgrade key—which includes changing the LSSS structure

and mapping function and comparing an existing plan to a policy change at the data owner's side—is significant if the policy has a lot of features.

B. **Proxy Re-encryption (PRE)**

Intermediary re-encryption was at first introduced by Mambo and Okamoto [10, 14] (PRE). A delegator notion is utilized in the suggested method for re-encrypting the originator's ciphertext. Under this strategy, neither the original data nor the decryption keys are learned by the delegator. This method has been adopted by several works [10, 14–16] due to the fact that it outsources substantial cryptographic operation costs to the proxy. In other designs, the PRE server is isolated and connects to a cloud server to compute the secret configuration for user revocation support [14].

In [13], the authors suggested the PRE method to make key updating easier. To send the values of encrypted ciphertext and the symmetric cryptography to the proxy, the user must interact with the proxies. This enables the proxy to retrieve a portion of the encrypted ciphertext. The user is then sent another piece of the ciphertext to decrypt by the proxy. Consumers still have to deal with two inconvenient procedures, and the communication cost between the client and the proxies is a significant burden when there are multiple decryption events, even if the proxy handles some of the calculation. To facilitate efficient and computationally cost-effective policy updates in the cloud, we presented a method we referred to as VL-PRE in [12]. A key generation optimization method for PRE and re-encryption forms the foundation of the fundamental strategy. Despite the fact that this method does not require data owners to deal with significant cryptographic operations, the cloud-side ciphertext re-encryption operation continues to be based on the CP-ABE. If there are a lot of re-encryption jobs, there may be performance issues for users who need to access the revised ciphertext.

However, as evidenced by three flaws, all of the aforementioned efforts that rely on proxy re-encryption and ciphertext updating have failed to take practicality into account. First, the new guidelines' cost of re-encryption is determined by a number of factors that are not compatible with PHR owners' mobile devices, particularly when the policy is extensive and frequently updated. Second, the reliability of the delegated system, like an outsourcing server if a dedicated proxy server hasn't been explicitly mentioned in any of their works. Existing works disregard the trust of the data owner and delegate in key updates or any other secret component. At last, no procedures that emphasis on the interaction that occurs of strategy refreshes exist.

This article focuses on the practical and effective use of secure outsourced CP-ABE policy updating with complete traceability. We combine CP-ABE with a proxy re-encryption method to make lightweight policy updating simpler. The proposed system makes use of parallel processing in order to enable effective data re-encryption. Under our proposed system, the owner of the data can change the policies that are in the outsourced data storage at any time. Users are also able to see the CP-ABE's cryptographic details, making it possible to use the tool. We offer the arrangement forming approach as one more fundamental component of our proposed admittance control methodology to give solid responsibility of strategy alteration history.

3 Methodology

Data transfer is a key task, and to secure data, we employ data encryption to disguise the content. In this study, the value of information is examined, and encryption methods are used to safeguard information. In this web application there is some modules called Data transfer is a key task, and to secure data, we employ data encryption to disguise the content. In this study, the value of information is examined, and encryption methods are used to safeguard information. In this web application there is some modules called Patient, Doctor, Hospital Management, Authority, Proxy server. Patient and Doctor are the major roles and rest modules are playing vital roles. Patient is the first person who register and login into the system. After successful login he will raise an appointment for treatment with their symptoms. With successful submission of appointment patient can view his/her medical report. The medical report will be generated if doctor accepts is appointment only. Doctor plays a major role like accepting the patient's appointments and generating medical report of patients and sends it to the proxy. Doctor also login with the credentials which are used in registration.

Hospital Management also plays major role. Management has the responsibility to maintain all the appointments and manage doctor requests and patient's request passed to the doctor. Management don't have any registration they just login with valid credentials and performs operations like appointments requests, doctor requests, and sending information to the doctor. Authority is also similar to management but their operations are view requests from proxy and generates keys to the authorized patients.

Proxy server plays a vital role to perform to view all the request from the doctor and pass those requests to the authority and authority also login with the default credentials.

4 Results

Patient Appointment Form

After patient successful login patient will raise an appointment by giving medical details (Figs. 1 and 2).

Doctor

Doctor will accept the request from the patient and views patients information (Fig. 3).

Medical Report

Doctor will upload patients report and view all the reports which are uploaded by the doctor and performs an action to send file to the authorized patient (Fig. 4).
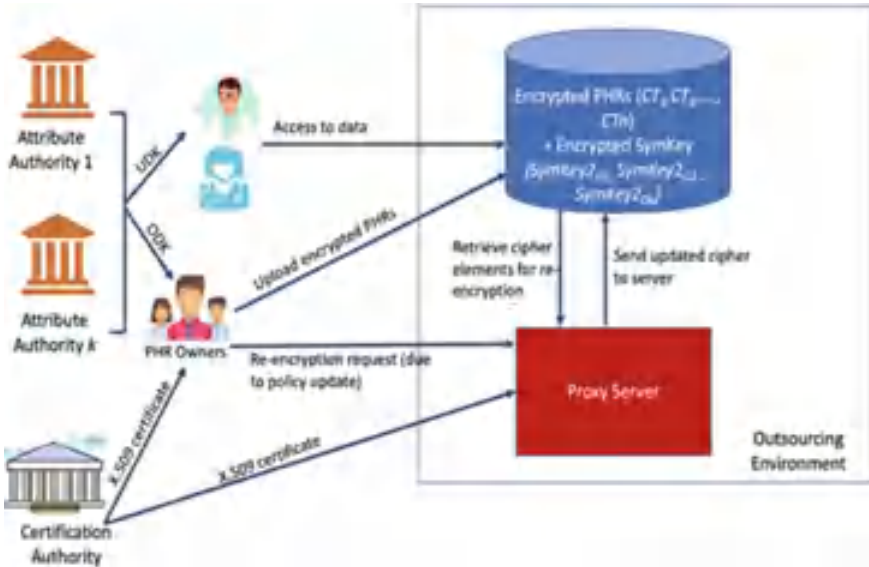
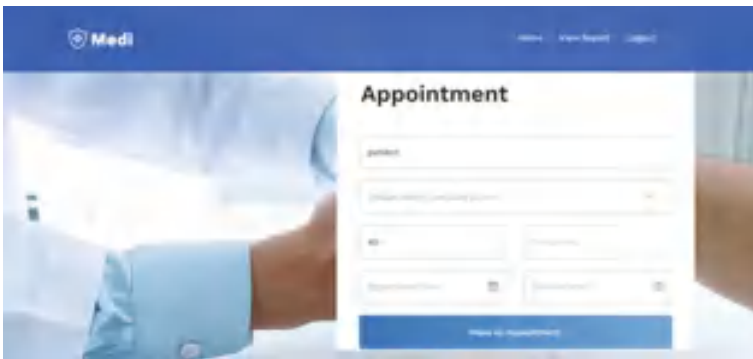**Fig. 1** Overall architecture of hospital management system

Fig. 2 Doctor appointment form

Patient Report

Patient will view his report after successful treatment (Fig. 5).

Management

Management will view the doctor request and patient requests and accepts the request based on patient requests (Fig. 6).

Proxy

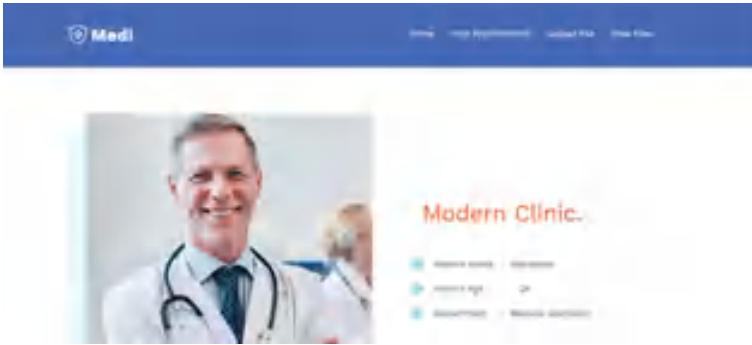Proxy views the request of a patient with patient report (Fig. 7).
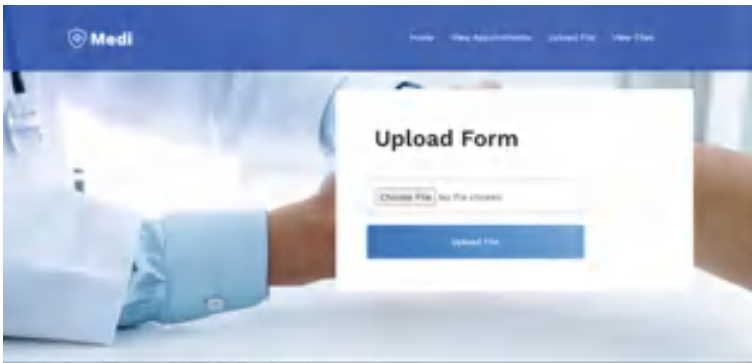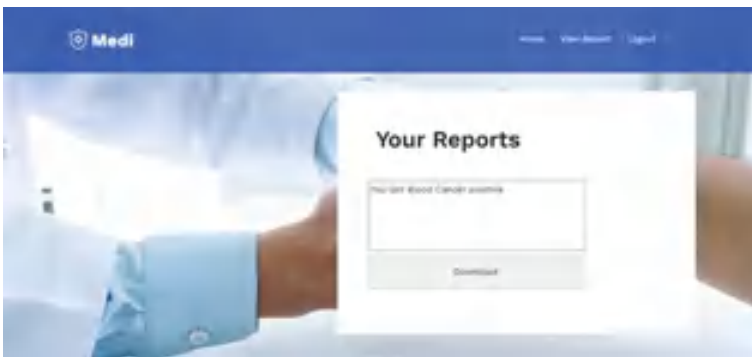
Fig. 3 Patient details

Fig. 4 Medical report

Fig. 5 Patient report

Fig. 6 Hospital management information

Fig. 7 Patient information

Authority

Authority will generate key and send that key to the user with help of that key user can view their report.

5 Conclusion

We have developed a policy update strategy based on proxy re-encryption and policy outsourcing. The outsourced server under our system bears the entire cost of policy updates. Multi-thread processing is also encapsulated in the re-encryption procedure, enhancing system performance, and scalability. For the experiment, we developed a graphical user interface (GUI) tool for putting CP-ABE policy updates into action. Owners of data can upload encrypted files and policies to outsourced storage using our technology. Re-encrypting data does not necessitate obtaining policies from a local database or communicating with an external server. Our web-based application allows for the modification of policies at any time and from any location. As a

result, the policy update management system and the file storage system provide transparent access control. We also talked about the policy versioning method, which lets you quickly recreate previous policies for a thorough audit. Last but not least, we demonstrated the effectiveness of file re-encryption. The outcomes demonstrated that a re-encryption method with multiple threads performed better than one with only one thread. To evaluate the cloud-based proxy with greater data volumes and broader access rules, we will in the future carry out extensive experiments in the actual cloud environment. In future, this can be extended to provide the security by using the technology blockchain.

References

1. Neerugatti V, Rama Mohan Reddy A, Rama A (2018) Detection and prevention of black hole attack in RPL protocol based on the threshold value of nodes in the internet of things networks. Int J Innov Technol Explor Eng 8(9)
2. Neerugatti V, Rama Mohan Reddy A (2019) Secured architecture for Internet of Things-enabled personalized healthcare systems. In: Internet of things and personalized healthcare systems. Springer, Singapore, pp 75–80
3. Wang Y et al (2021) Security-aware and privacy-preserving personal health record sharing using consortium blockchain. IEEE Internet Things J
4. Deng F et al (2018) Ciphertext-policy attribute-based signcryption with verifiable out sourced designcryption for sharing personal health records. IEEE Access 6:39473–39486
5. Klein EL (2014) The visualization of privacy filters for sharing sensor-based health data. Diss. University of Massachusetts Lowell
6. Gajanayake R, Iannella R, Sahama T (2011) Sharing with care: an information accountability perspective. IEEE Internet Comput 15(4):31–38
7. Do Hoang H et al (2021) A blockchain-based secured and privacy-preserved personal healthcare record exchange system. In: 2021 IEEE international conference on machine learning and applied network technologies (ICMLANT). IEEE
8. Xu C et al (2019) Achieving searchable and privacy-preserving data sharing for cloud-assisted E-healthcare system. IEEE Internet Things J 6(5):8345–8356
9. Sahai A, Waters B (2015) Fuzzy identity-based encryption. In: Proceedings of 24th annual international conference on the theory and applications of cryptographic techniques (Eurocrypt). Lecture Notes in Computer Science. Springer, Berlin, pp 457–473
10. Zhang Y et al (2020) Efficient identity-based distributed decryption scheme for electronic personal health record sharing system. IEEE J Select Areas Commun 39(2):384–395
11. Zhang et al (2021) SIP: an efficient and secure information propagation scheme in e-health networks. IEEE Trans Netw Sci Eng 8(2):1502–1516
12. Indhumathi V, Prakasham V (2015) On demand security for Personal Health Record in cloud computing. In: 2015 international conference on innovations in information, embedded and communication systems (ICIIECS). IEEE
13. Li et al (2012) Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. IEEE Trans Parall Distrib Syst 24(1):131–143
14. Deng F et al (2018) Ciphertext-policy attribute-based signcryption with verifiable outsourced designcryption for sharing personal health records. IEEE Access 6:39473–39486
15. Wang Y et al (2021) Security-aware and privacy-preserving personal health record sharing using consortium blockchain. IEEE Internet Things J
16. Li et al (2018) Unified fine-grained access control for personal health records in cloud computing. IEEE J Biomed Health Inform 23(3):1278–1289

Performance Evaluation of ML-Based AWS Security Evaluation Model for Cloud Computing

B. L. Malleswari, Rakshita Kolachalama, and Voruganty Sesha Srivallii

Abstract As there is a great increase in demand for cloud services, the number of companies that provide cloud services have been increased in number. Cloud-based web services wholly depends on the service provides by the companies. Hence, it is very essential to choose the cloud service provider for obtaining best services. API consistently gives access to different cloud services. In this paper a quantifiable security evaluation system for various clouds is analyzed. A set of evaluation elements are present in the security evaluation model which compresses of various sectors via storage, computing, maintenance, application security, etc. The major aim of this paper is to deal with the issues related and exhibits the comparison of Amazon Work Space (AWS) and International Business Machines (IBM) cloud. Either of the cloud service platforms are compared with each other under some situations for respective instances. A machine learning-based CNN method is used for developing the proposed cloud computing. This mechanism calculates the score for each field and gives the total score as result in order to create one secured pattern. Dynamic security scanning is done for single or multiple clouds by using visual graphs and cloud users are guided for improvement of cloud operations, changing the configuration and to provide protection against vulnerabilities. This ultimately results in the enhancement of security in cloud resources.

Keywords Cloud computing · Security evaluation · Quantifiable evaluation · AWS · Cloud security · Security visualization

B. L. Malleswari (✉)
Sridevi Women's Engineering College, Hyderabad, Telangana, India
e-mail: blmalleswari@gmail.com

R. Kolachalama
CSE, Sri Chandrasekharendra Saraswati Viswa Mahavidyalaya University, Kanchipuram, Tamil Nadu, India

V. S. Srivallii
CSM, JNTU College of Engineering, Wanaparthy, Telangana, India

667

1   Introduction

Cloud computing (CC) is a trending technology in today's world which is useful in obtaining simple and flexible interaction with the resources. The main advantage of this cloud storage is that it is not dependent on physical resources. Hence, fast allocation is enabled and resources are reallocated depending on the user demands. The virtual resources are allocated and they can be abstracted with the improvement of cloud computing technology continuously. In present day technology, the information infrastructure of various users can be stored and operated with the help of cloud which is a common method now-a-day. Hyper scale data centers for cloud services [1] are present such as AMAZON AWS, Microsoft AZURE, and Ali Cloud which are emerging very quickly at present.

The huge number of users makes use of different kinds of clouds such as public, private, community cloud, and hybrid clouds. Even though there is a great demand for this machinery, the services and operational conveniences are available at a lowest price. The users of cloud may not be aware of the security standards provided by the service providers and the data they store in clouds is which type of cloud is also not known to the user.

The cloud users cannot know whether their cloud services are safe, and whether their data can be safely placed in different clouds and security related issues are handled in an efficient procedure [2, 3]. A network shared computer systems are present in a data center where the business and other issues relating to management of organization are stored and processed in this network for huge amounts of data.

Cloud services are well known for the expandable storage capacity as it is resizable basing on the needs of the cloud user. It also depends on the unlimited storage capacity offered by the vendors. As third party manage could functionalities, the user may not have much control over it [4]. While considering administrable resources and accessibility it is difficult for the users, visitors and administrators to maintain and grasp the entire data security in single cloud. Providing security becomes complicated when we opt for different type of cloud at once. So that Cloud Service Providers (CSPs) are introduced to enhance the security in cloud environment. But facing different CSPs and its own private clouds, the CSPs cannot set up one universal security management mechanism to solve the security means across different clouds [5].

Infrastructure of cloud computing is designed in secured evaluation system in a quantifiable manner which includes scanning engine, quantifiable security evaluation system and visual display module, and so on [6]. For obtaining total score, one field score count mechanism adopts a process known as "One vote vetoed" mechanism. Through security visual UI of the resource library containing information infrastructure is provided with security for the users of cloud and the score decides security standards.

2 Literature Review

Cloud computing requires components and sub elements in a CC architecture model. This CC architecture consists of two attributes. The front-end deals with the cloud clients which resembles the characteristics of infrastructure. Here the client has authority to control the applications quickly and efficiently. This result has low maintenance with better services that are enabled by the companies to achieve the desired results by obtaining the expected requirements. Functionality of the cloud system [7] is accessed through the central server which manages the system, monitoring traffic, and giving permission to the request id by the client.

The Information Technology field that deals with service module and the fundamental technology is nothing but cloud computing. To fix the bugs and debug the errors relating to web-based services, test cases, and a smaller number of research papers identifies tools for testing of web composition services for a formal specification and test case implementation [8].

FengDengguo, Zhang Min et al. explained cloud computing requirements such as security key technology, regulation of security standards generates framework for cloud computing security [9]. Rong et al. stated even lot of technological methods results in best performance of cloud computing services yet, there is no complete remedy for many drawbacks and for various challenges such as Service Level Agreements (SLA) of security that are to be solved in future days [10].

Zunnurhain and Vrbsky explained about identification and description of the major security attacks that occur on cloud with an intention finding theoretical solutions for every single problem and solving the issues with the unique solution obtained by the cloud services [11]. A set of rejected atomic services are the characteristics that are configured for the inconsistency in service composition.

Other services are also linked with the assortment of these services at the time of excluding other services in the cloud. By designing the feature diagrams various papers modeled the configuration variability in the service [12, 13].

Phoronix Test Suite is conducted on the virtual servers by taking as instance for the comparison of Amazon Work Space Elastic Compute Cloud (EC2) server and IBM server. AWS is a cloud service provider which offers cloud servers on demand cloud storage space that has to be paid for resources used by the cloud user [14]. The security evaluation system is designed to enhance the security provided by cloud services by adding some features that supports business activities and open-source scanning models [15].

3 Security Evaluation System

The security evaluation system for enhanced cloud services is represented in Fig. 1 that clearly depicts the quantifiable security evaluation system for various cloud platforms. This architecture includes database vulnerability for security evaluation,

Fig. 1 Cloud computing platform security evaluation system with AWS & IBM Instances

security visualization display module, recovery engine, scanning engine, and a model designed for establishment and maintenance of security evaluation.

This framework also explains about quantifiable evaluation methods and adding security improvement features to the storage security set that consists of network security set, application security, and operation security, etc. This module can be deployed either on single or multiple severs basing on the resource availability.

The real time application dealing with evaluation system consists of security scanning engine which examines the security standards of cloud usage service and security items score is calculated for various evaluation metrics such as network security, storage security, application, and maintenance security. In this architecture serial or parallel scanning of metrics is executed and it generates score for each resource basing on the standards of evaluation model.

Quantifiable security evaluation model is the main theme of this security evaluation system which deals with the modified rules and the means of security evaluation system. The researches lead to the design of this proposed model by the experts by adopting business purpose and open-source scanning engines.

3.1 Security Metrics and Its Collection

The security fields are present in our Security Evaluation Model and the fields involved in security model are given as

$$P = \{P_1, P_2, P_3, P_4, \ldots P_N\}$$

where the security collection for computing is represented as P_i for security collection of storage, maintenance security collection, the network security collection, and security collection on application, etc. Here, P_i of P involves various items such as P_{ij} are involved for examining the security by checking the standards of physical server OS and VM OS. The system container and few other vulnerability items are also involved for checking the security of this system which involves all items that are composed of P_i.

$$P = \{P_{i1}, P_{i2}, P_{i3}, P_{ij}, \ldots P_{iM}\}$$

At each P_{ij} instance the security status of object is scanned by the security scan engine present within the cloud collection of set 'A' cloud users are represented in Eq. (1) as follows:

$$A = \{A_1, A_2, A_3, A_4, \ldots, A_N\} \tag{1}$$

$$A_i = \{A_{i1}, A_{i2}, A_{i3}, A_{ij}, \ldots, A_{iM}\} \tag{2}$$

In this equation the user physical machines and virtual machines are represented with A_i as represented in above equation. The security status is verified by using the object A_{ij}. By considering all these scores after scanning the triple set score is given out as

$$S = [S_{ij}, L_{ij}, O_{ij}].$$

From the above given triple set S_{ij} is used to represent the high score, whereas L_{ij} represents the vulnerability level of cloud security and O_{ij} represents the fixed methods that are used in cloud service. When there is weak security, it is stated as 0 and when there is no security it is taken as 1. Complicated state is denoted as L_{ij}. Here S_i, corresponds to respective P_i to sum of the corresponding total S_i is MAX is given as either 1 or 100 in Eqs. (3) and (4).

$$S_i = \sum_{i=1}^{M} S_{ij} \tag{3}$$

$$MAX = \sum_{i=1}^{M} S_i \tag{4}$$

According computing, storage, and network security weight S_i is taken as a fixed score. For each and every security attribute one score is taken as S_{ij} basing on its specific weight. Here, S_i is considered as dynamic score which may be changed basing on the changing values instantly.

3.2 Security Evaluation Process for Quantifiable Metrics

UP is used for denoting and representing the security evaluation process for quantifiable metrics. This UP_i corresponds to P_i, where P_i indicate the number of scanned items in the set as follows: UP_i.

$$UP = \{UP_1, UP_2, UP_3, UP_4, \ldots, UP_N\}$$

Basing on the security evaluation model the user resource view is scanned by the security scanning engine which is utilized to scan the item P_{ij}, in serial method or parallel means. Thereafter gives an output as US_{ij} which is an actual score.

$$US_i = \sum_{i=1}^{M} US_{ij} \tag{5}$$

Taking into consideration the cloud resource view of various users the actual security score is generated for different kinds of checking attributes. In such case administrator can monitor the entire global security view of the cloud in order to draw the attention of cloud user with regard to cloud security standards. In our proposed architecture to own the critical vulnerability a strategy called CNN-MSVM technique is adopted for obtaining accurate and best result.

3.3 Repairing of Security Vulnerability

The user opts to restore the scanned holes, the repair engine of security will call O_{ij} that corresponds with P_{ij}. Various security holes having repair rules corresponds with the link repair engine in a serial or parallel means.

4 Result and Discussion

The G-Cloud Plat form is implemented within the architecture of our proposed ML-based CNN-MSVM cloud computing security evaluation system. In the current research used two independent cloud platforms out of which one cloud is built basing on G-Cloud and the other is on basis of Open stack platform. Both the cloud platforms can be monitored by using common security API.

Now we discuss about the data collected as a part of our research and the results obtained for tests evaluated basing on security standards and the result are plotted and analyzed from the data collected for obtaining the enhanced performance of basic cloud services. Though conducting research and obtaining test cases is a time taking procedure it is feasible for obtaining the result in this paper. Taking into consideration the respective data from AWS and IBM cloud services the experiments are conducted to depict the security evaluation among the AWS and IBM cloud service providers.

Figure 2 depicts the apache benchmarking results comparison among AWS and IBM req/sec is shown below.

Figure 3 depicts the Dbenchmarking results comparison among AWS and IBM req/sec is shown in below graph.

The proposed system comprises of using machine learning system-based classification to detect any anomalous behavior based on the data analysis across the cloud environment. In this case, convolution neural network (CNN)-based architecture is

Fig. 2 Benchmarking apache result
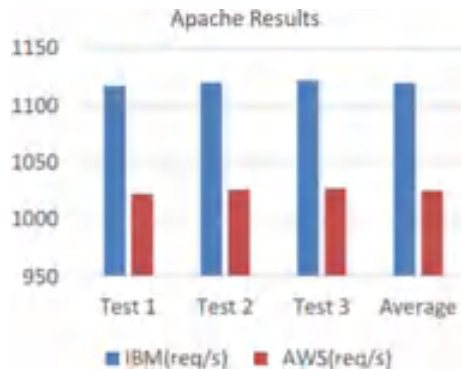
Fig. 3 Dbenchmarking comparison of IBM and AWS
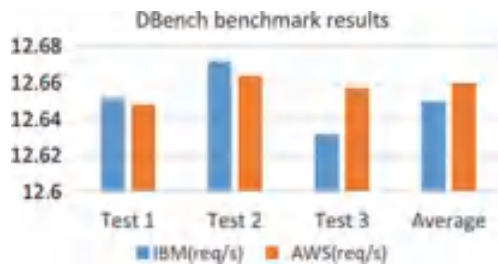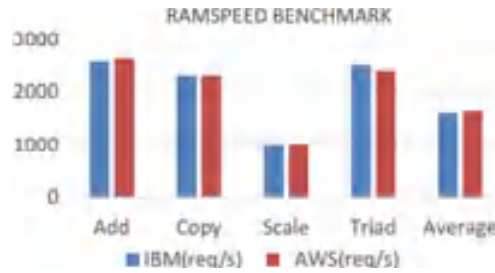
Fig. 4 Comparison of RAM speed among IBM and AWS

used for classification of anomalies. The input information from cloud is fed to the convolution neural network system.

Figure 4 represents the results for comparison of RAM speed among AWS and IBM req/sec as shown in following graph.

Here after the overall experimental results we obtained a scorecard for performance evaluation of cloud services among IBM cloud services and AWS cloud services. The resultant performance is depicted in Fig. 5 in the form of a graph.

From the experimental results, Table 1 gives a brief evaluation of virtual machines on both IBM and AWS cloud service platforms is shown in this table.
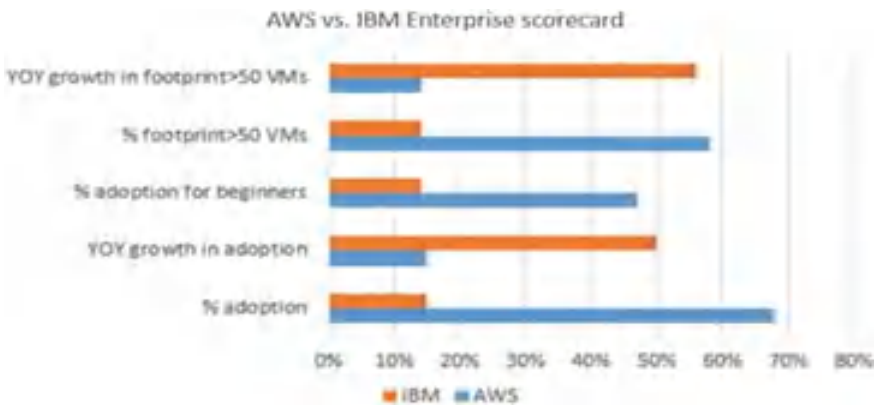
Fig. 5 Overall comparison among IBM and AWS cloud services

Table 1 Comparison among IBM and amazon for virtual machines review

IBM	AMAZON
Hardware: Processor Intel Xeon E526600@2.19 GHz Memory: 512 MB + 256 MB 21 GB V Disk	Processor Intel Xeon E526600@1.80 GHz Memory: 588 MB Disk: 30 GB
Software OS: Ubuntu 1404	Software OS: Ubuntu 1404
File system: ext4	File system: ext4
System layer: IBM Hyper-V	System layer: Xen 3.4.3 amazon hypervisor
Server core count: 1 Thread count: 1	Core count:1 Thread count:1

5 Conclusion

The research conducted in this paper provides the analysis of single or multiple cloud environment for quantifiable security evaluation technique. This research study developed a novel ML-based cloud computing security model called CNN-MSVM method for developing the algorithm to enhance the security standards of a cloud computing channel. The attributes present in the quantifiable security evaluation system consists of scanning engine and recovery engine for evaluating the security standards which also includes Visual Display Module (VDM), etc. Computing, application security, storage, network, and maintenance are the various fields which are the evaluation elements for the security evaluation model. Three tuples measured are vulnerability, score and repair method for each element. ML-based convolutional neural network-multiclass support vector machine (CNN-MSVM) technique is adopted for each field to measure the score and to add this score to the total score summary.

According to the experimental results, better performance of Amazon Work Space (AWS) cloud is observed in aspect of disk performance and RAM Speed when compared with IBM cloud services. This machine learning-based CNN-MSVM method for security of cloud computing services is observed to be a best algorithm when compared with conventional practices. In addition to this AWS also consists of another security feature called RSA better than IBM. When compared with cost also AWS has reasonable and low price while IBM is a costlier service provider. The variation among the two service providing platforms is that web sphere instances of IBM are compatible and it can be accessed on AWS EC2 whereas the vice versa is impossible.

References

1. Jayanthiladevi R, Mohanasundaram A, Keerthana G (2020) Software defined cloud infrastructure. In: Handbook of research on cloud and fog computing infrastructures for data science, pp 108–123
2. John W, Rittinghouse, James Ransome F (2020) Cloud computing: implementation, management, and security
3. Ahmed M, Hossain MA (2019) Cloud computing and security issues in the cloud. Int J Netw Secur Appl 6(1)
4. Choudhury S, Jain T, Kumar P, Rathore YS (2019) Communication by terrorist using access points: cyber security. SmartTechConference. IEEE Explore
5. Curran K, Sean C (2018) Cloud computing security
6. Avuya M, Gerber M, MostertPhipps N (2018) Information security risk measures for Cloud based personal health records. In: 2018 international conference on information society. IEEE
7. Chitra Rajagopal P (2018) Cloud security algorithm to enhance the security of the layers. In: 5th IntrnConf on Sys modeling and advancement in research trends
8. John H (2017) Data security in the world of cloud computing. IEEE
9. Sharma A, Choudhury T (2017) Health monitoring & management using IoT devices in a cloud based framework. In: International conference on advances in computing and communication engineering (ICACCE), IEEE Explore
10. Ali M, Khan SU, Vasilakos AV (2016) Security in cloud computing: opportunities and challenges. Inform Sci Cloud Comput 305:357–383
11. Oscar R (2015) Empirical evaluation of a cloud computing information security governance framework. Inform Softw Technol 44–57
12. Akhil B (2015) An analysis of cloud computing security issues. IEEE
13. Narula S, Jain A, Prachi (2015) Cloud computing security: Amazon Web Service. IEEE
14. Zeng W, Zhao J, Liu M, Zhao J, Liu M (2012) Several public commercial clouds and open-source cloud computing software. In: 7th international conference on CSE (ICCSE) 2012
15. Alexander O (2011) Implementing security rules, safeguards, and IPS tools for private cloud infrastructures. Infrastructure Security as a Service (ISaaS), GROOT

Analysis of Radiation Induced in Multiphase Flow of a Viscous Conducting Heat and Mass Transfer Fluid in a Vertical Porous Medium

Check for updates

Pamula Rajakumari, Jaladi Rajendra Kumar, Polaiah Bojja, C. Madhusudhana Rao, and P. Govardhan

Abstract In this paper, presently a multiphase model discusses the effect of radiation and radiation absorption on free convective heat and mass transfer flow of a viscous conducting fluid through a porous medium in a non-uniformly heated vertical channel. The walls are maintained at a non-uniform temperature, and a uniform concentration is maintained on the walls. Such problems are care characteristically induced multi-physic, with the mix of various fluids with radiation absorption, heat and mass transfer, porous medium, non-uniform temperature. Observing the change of phase, and chemical reactions and their phenomenal influence in the porous medium. In the analysis of various composition fluid systems, the model has proposed to observe on a microscopic scale by applying the attributes of fluid dynamics constrained averaging theory which gives appropriate results that satisfy the applicable law of fluid dynamics for all constituents both at micro- and macro-visual observance levels. In addition to some important observances of fluid dynamics, some of the flow parameters and their attributes while flowing through the porous medium describe the mechanism of multi-fluids transfer and its temperature in the vertical channel walls of the porous medium. The coupled equations governing the flow and heat transfer are solved by using the perturbation technique, and 3D numerical simulations aimed at proving the validity of the proposed adopted method are carried out results, and analysis presented. The expression for the velocity, the temperature, and

P. Rajakumari (✉)
Institute of Aeronautical Engineering, Hyderabad, Telangana, India
e-mail: rajakumari258@gmail.com

J. R. Kumar
Department of Fine Arts-Visual Communication, Koneru Lakshmaiah Education, Foundation Vaddeswaram Guntur, Guntur, Andhra Pradesh, India

P. Bojja
Department of Computer Science and Information and Technology, Institute of Aeronautical Engineering, Hyderabad, Telangana, India

C. Madhusudhana Rao · P. Govardhan
Department of Computer Science Engineering, Institute of Aeronautical Engineering, Hyderabad, Telangana, India

677

rate of heat and mass transfer are derived and are analyzed for different variations of the governing parameters.

Keywords Radiation · Radiation absorption · Heat and mass transfer · Porous medium · Non-uniform temperature · 3D numerical simulations

1 Introduction

As per the importance, visual observation of multiphase flow of a viscous conducting fluid in vertical porous medium described the various phenomenal fluid dynamic motion and characteristics attributes in a porous medium. These visual attributes make us to understand the concepts and phenomena of fluid dynamics, and their role in the modern world's utility as human-centric usage in various fields like medical sciences, where it acts as the main source to run the mechanism of human metabolism. The utilization of various engineering fields is from nanofluids intervention to massive mega-scale structural construction developments. In yielding industrial to domestic power generations through various physical states of transformation [1], fluid has an extensive utilization in agricultural development research in biotechnological services in organic tissue building for the entire human race and as well for the rest of biodiversity developments in the expansion of cattle farming. Some of the industries, like chemical, pharma, textile, infra-constructions for both domestic and industry, thermal power generators, were completely run with the support of fluids (water) [2]. However, 3D numerical simulations and perturbation techniques, the velocity, temperature, and concentration, rate of heat, and mass transfer have been analyzed for different variations in the governing parameters [3].

2 Formulation of the Problem

We consider the motion of the viscous, incompressible fluid through a porous medium in a vertical channel bounded by flat walls as shown in Fig. 1 which is in terms of visualization for the expression of velocity, temperature, rate of heat, and mass transfer are presented and analysis for different variations of the governing parameters [4]. The fluid flows through the rock porous medium, with the gravitational phenomenal influence.
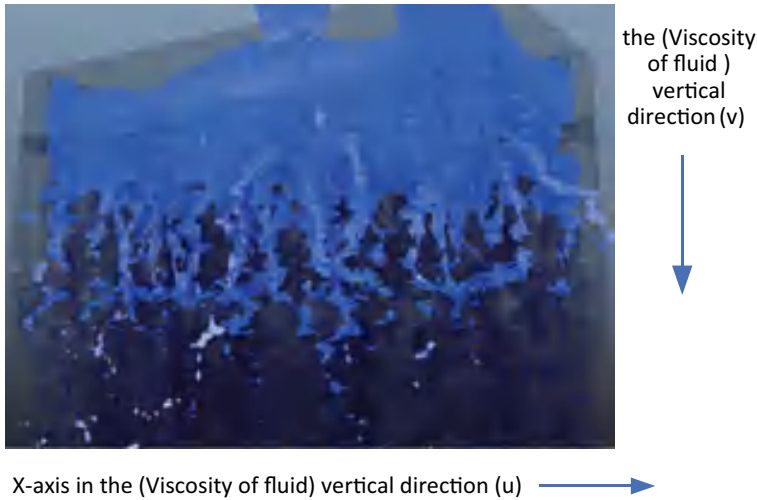
Fig. 1   Variation of velocity (*u*) with gravitational force fluid (*G*)

The observed visual characteristics for the liquid (water) are pressure, temperature, refraction [5], viscosity, and surface tension viscous fluid through a high porous medium bounded by an infinite vertical porous plate [6].

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{1}$$

The Boussinesq approximation is used so that the density variation will be considered only in the buoyancy force. We choose a rectangular Cartesian system (x, y) with an x-axis in the vertical direction and a y-axis normal to the walls. The walls of the channel are at y $= \pm$ L based on Eq. (1).

3   Analysis of the Flow

The visual observance of Fig. 2 shows the pressure applied vertically with the influence of viscosity and surface tension in fluids, in vertical flow through the porous medium. Due to the high medium in density, the flow of the fluid in the porous has much slow. In the presence of high viscosity, the velocity of the fluid flow vertically seems below [6].

the (Viscosity of fluid) vertical direction (v)

X-axis in the (Viscosity of fluid) vertical direction (u)
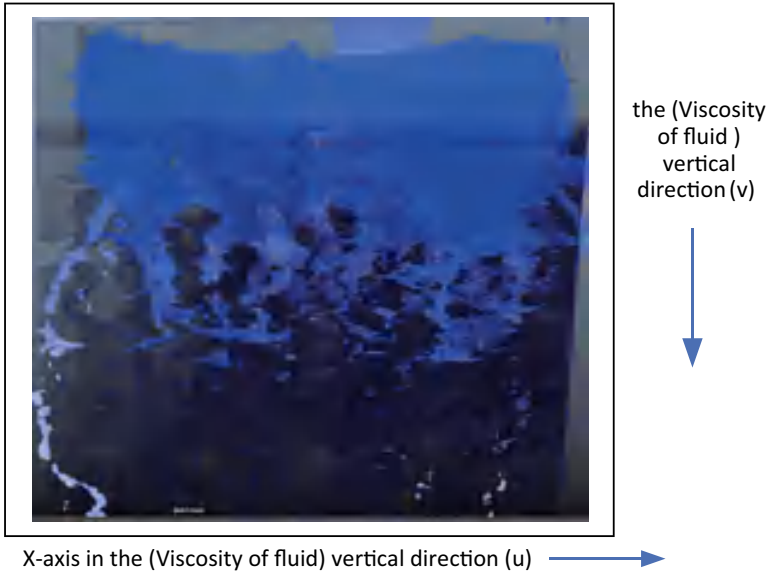
Fig. 2 Variation of *v* with surface tension

As per the visual observance of Fig. 3, it shows low medium density in vertical flow through the porous medium. Due to the low medium in density, the flow of the fluid in the porous has much quicker as shown in Figs. 5 and 6. In the presence of low viscosity, the velocity of the fluid flow vertically seems be high.

Figure 6 results show about the occupancy of the fluid seems too dense in the porous medium. The less in the viscosity, higher rate of flow, and higher in velocity in the vertical flow.

Figure 7 results show about the occupancy of the fluid seems to less in density in the porous medium. The more the viscosity, lesser the flow, and less in velocity in the vertical flow [7].

The fluid is able to display its visually refraction attributes as per the respective fluid's viscosity as shown in Figs. 8 and 9.

Figures 10 and 11 represent the heat transfer of mixed conductive with various fluids, which applies pressure in the vertical flow of the porous medium. The flow of the fluid, a higher rate in velocity due to the vertical flow [8].

$$
\begin{aligned}
\rho_{(e)} C_p (u \partial T / \partial x + v \partial T / \partial y) \\
= \lambda \big((\partial^2 T)/(\partial x^2) + (\partial^2 T)/(\partial y^2) \big) \\
- Q(T - T_e) + Q_1^{\wedge\prime}(C - C_e) - \partial(q_r)/\partial r
\end{aligned}
\tag{2}
$$

the (Viscosity of fluid) vertical direction (v)

X-axis in the (Viscosity of fluid) vertical direction (u)
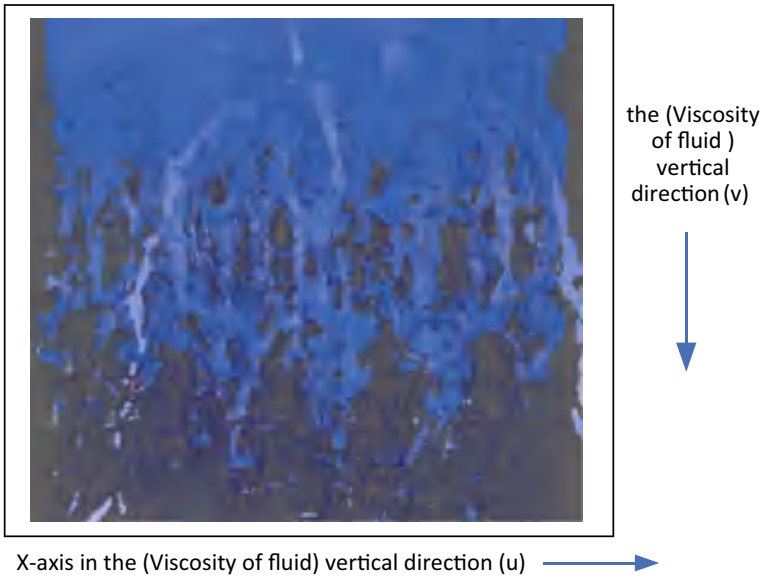
Fig. 3 Variation of *v* with pressure

Figures 10 and 11 represent the heat transfer of the radiation in the mixed conductive with various fluids. Due to the higher density, which causes to apply pressure in the vertical flow in the porous medium [9].

With respect to Figs. 12 and 13 are about the various terms of radiation effective on the flow of fluid motion in the porous medium vertical flow. There can identify the mixed conductive with various liquids and applied with effectiveness as Eq. (3).

$$\left(u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} \right) = D_1 \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right) - k_1 (C - C_e) \tag{3}$$

With respect to Figs. 14 and 15, it is about the various terms of radiation observations in the flow of fluid motion in the porous medium vertical flow as per Eqs. (4) and (5). There it is identified as the nth unit of porous medium mass transfer in the size of volume with the mixed fluid [10] for uniform variation as shown in Figs. 16 and 17.

The visual appeal of the color map in Fig. 17 describes the uniform temperature in the porous medium. Due to the absence of fluid motion in the porous medium, the channel walls are a plain appearance of the uniform temperature image.

$$R\left(\frac{\partial(\psi, \nabla^2 \psi)}{\partial(x, y)} \right) = \nabla^4 \psi + \left(\frac{G}{R} \right)(\theta_y + NC_y) - M^2 \frac{\partial^2 \psi}{\partial y^2} \tag{4}$$

$$RP\left(\frac{\partial\psi}{\partial y}\frac{\partial\theta}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\theta}{\partial y}\right) = \nabla^2\theta - \alpha\theta + Q_1 C \qquad (5)$$

The visual display of the color map in Figs. 18 and 19 visualizes the non-uniform temperature in the porous medium. The flow of the fluid motion in the porous medium that flows through the channel walls seems to be a temperature variation; thus, it appears and displays the non-uniform temperature image from Eqs. (6)–(8)

$$RSc\left(\frac{\partial\psi}{\partial y}\frac{\partial C}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial c}{\partial y}\right) = \nabla^2 C - \gamma C \qquad (6)$$

where

$$R = \frac{qL}{\nu} \quad \text{(Reynolds number)} \quad \gamma = \frac{K_1 L^2}{D_1} \quad \text{(Chemical reaction parameter)}$$

$$G = \frac{\beta g \Delta T_e L^3}{\nu^2} \quad \text{(Grashof number)} \quad M^2 = \frac{\sigma\mu_e^2 H_o^2 L^2}{\nu^2} \quad \text{(Hartmann Number)}$$

$$P = \frac{\mu C_p}{\lambda} \quad \text{(Prandtl number)}, Q_1 = \frac{Q_1'(C_1 - C_2)L^2}{D_1(T - T_e)}$$

(Radiation absorption parameter)

$$Sc = \frac{\nu}{D_1} \quad \text{(Schmidt Number)} \quad \alpha = \frac{QL^2}{\lambda} \quad \text{(Heat source parameter)}$$

$$\mathbf{N_1} = \frac{3\boldsymbol{\beta_R}\boldsymbol{\lambda}}{4\boldsymbol{\sigma}^*\mathbf{T_e^3}} \quad \text{(Radiation absorption parameter)}$$

$$\mathbf{N_2} = \frac{3\mathbf{N_1}}{3\mathbf{N_1} + 4}, \quad \boldsymbol{\alpha_1} = \alpha\mathbf{N_2}, \ \mathbf{P_1} = \mathbf{P}\mathbf{N_2}$$

The corresponding boundary conditions are

$$\psi(+1) - \psi(-1) = 1$$

$$\frac{\partial\psi}{\partial x} = 0, \ \frac{\partial\psi}{\partial y} = 0 \quad \text{at } y = \pm 1 \qquad (7)$$

$$\theta(x, y) = \gamma(\delta x), \quad C = 1 \quad \text{on } y = -1$$
$$\theta(x, y) = \gamma(\delta x), \quad C = 0 \quad \text{on } y = +1$$

$$\frac{\partial\theta}{\partial y} = 0, \ \frac{\partial C}{\partial y} = 0 \quad \text{at } y = 0 \qquad (8)$$

The value of ψ on the boundary assumes the constant volumetric flow in consistent with the hypothesis (Eq. 8). Also, the wall temperature varies in the axial direction in accordance with the prescribed arbitrary function t.

The thermal buoyancy in the flow field is created due to the non-uniform temperature on the walls. $y = \pm L$ while both the walls are maintained at uniform concentration.

4 Discussion of the Results

The result analysis of Figs. 4 and 5 represents with visual observance of Figs. 2 and 3, and it shows the pressure applied vertically with the influence of viscosity and surface tension in fluids and in vertical flow through the porous medium. Similarly, Figs. 10 and 11 are observed in visuals for the heat and radiation resultant as shown in Figs. 12 and 13, respectively. Further, it has been analyzed in terms of visualization performance varies both the uniform and non-uniform vertical channel wall temperature as shown in Figs. 16 and 17. An increase in radiation absorption results from an enhancement in the actual temperature Figs. 18 and 19. Hence, the above analysis is numerical simulated based on Eqs. (6)–(8) represented with chemical reaction variables that are used in the equations.
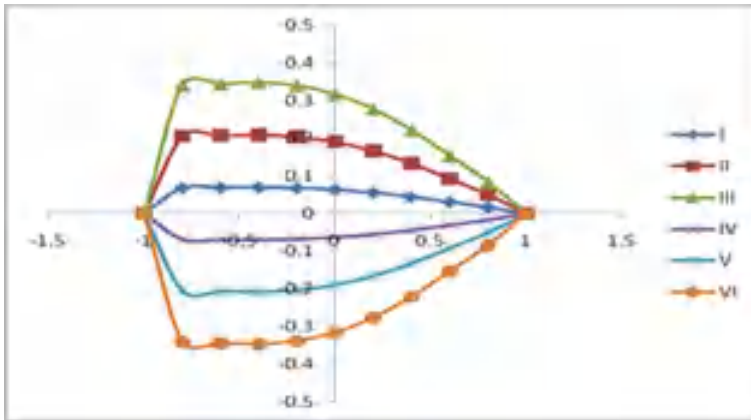
Fig. 4 Variation of u with surface tension
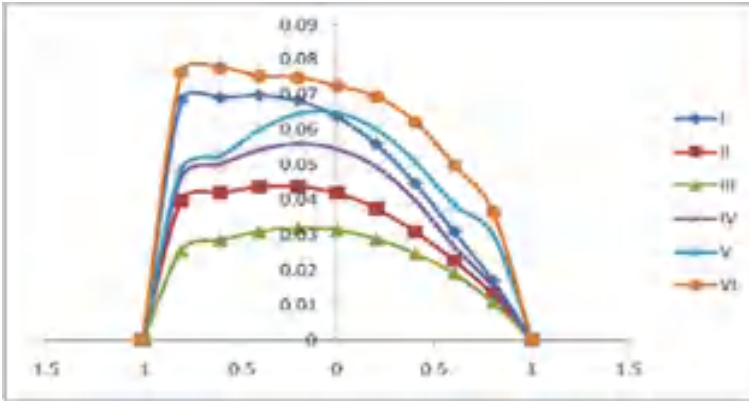
Fig. 5 Variation of *u* with pressure

X-axis in the (Viscosity of fluid) vertical direction (u) ⟶
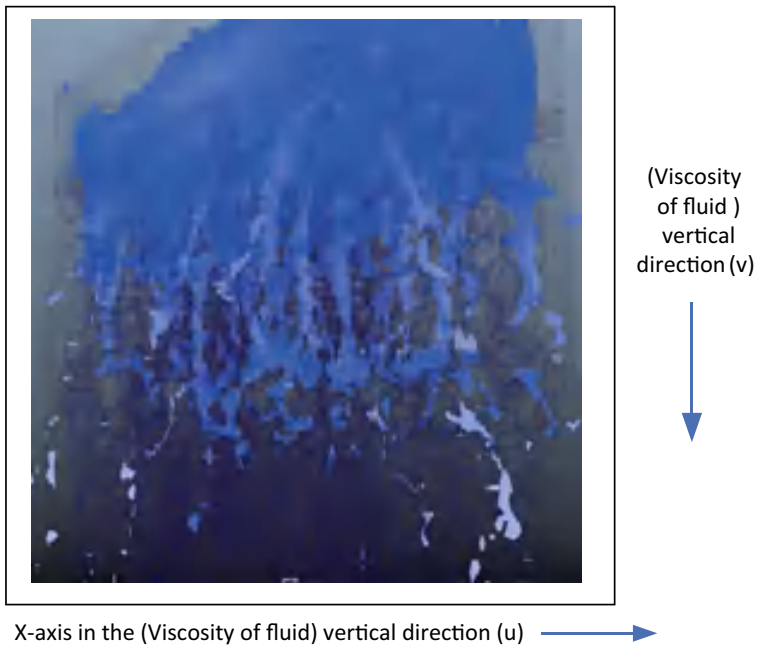
Fig. 6 Variation of *v* with viscosities

(Refraction) vertical direction (v)

X-axis in the (Refraction) vertical direction (u)
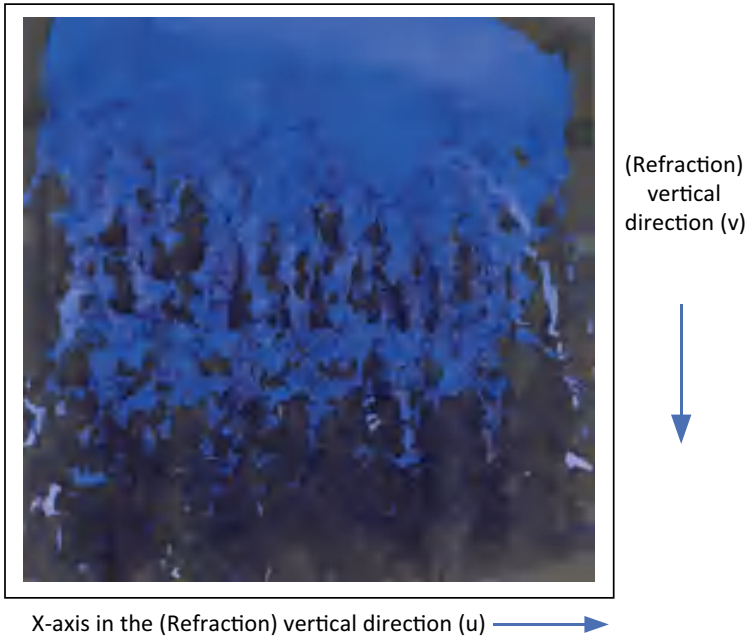
Fig. 7 Variation of *v* with refraction
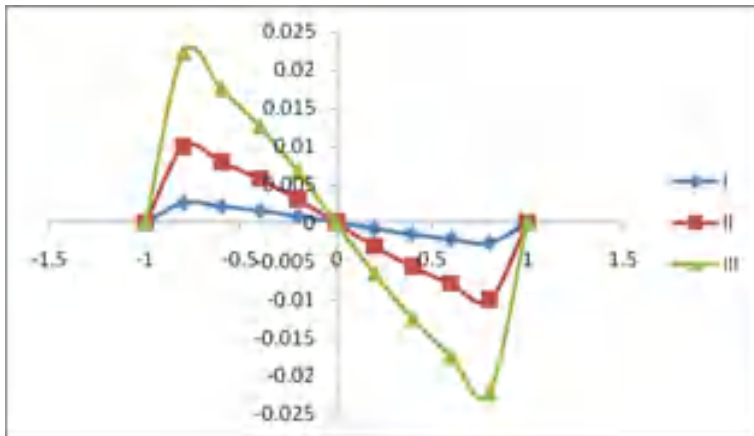
Fig. 8 Variation of *v* with viscosity

Fig. 9 Variation of *v* with refraction

X-axis in the (Heat) vertical direction (u) ⟶
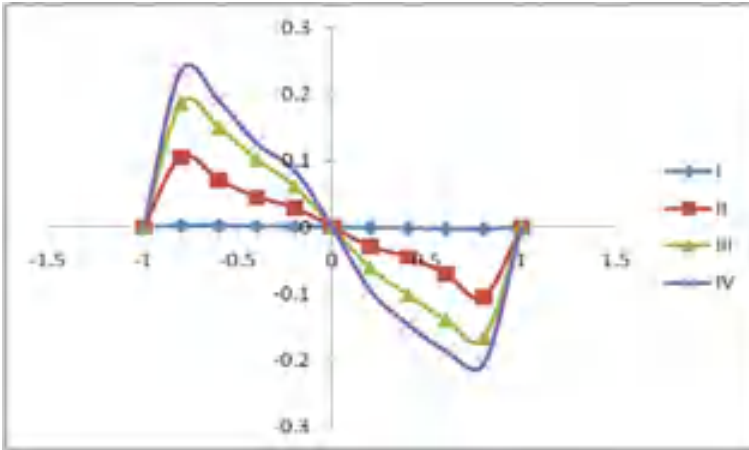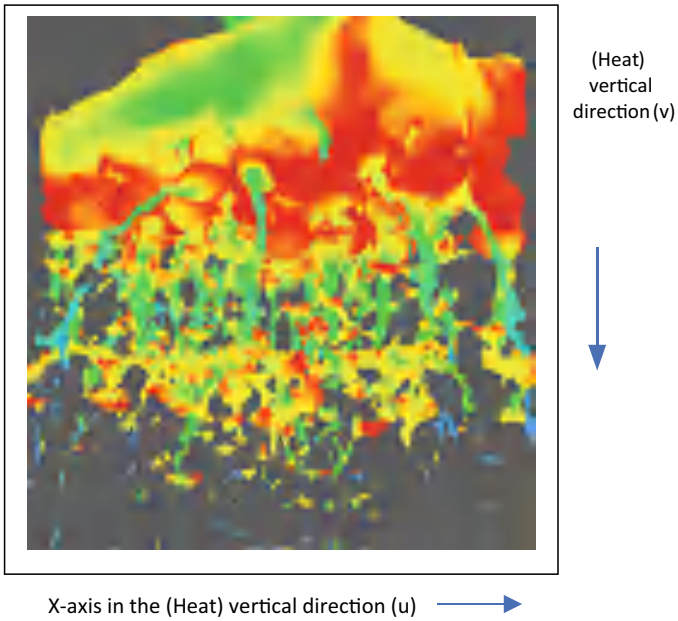
Fig. 10 Variation of *v* with heat

X-axis in the (Radiation) vertical direction (u) ⟶

(Viscosity of fluid) vertical direction (v)

Fig. 11　Variation of *v* with radiation

X-axis in the (**Temperature**) Horizontal direction (u) ⟶

(Temperature) vertical direction (v)

Fig. 12　Variation of *v* with temperature

(Radiation Absorption) vertical direction (v)

X-axis in the (**Radiation Absorption**) Horizontal direction (u) ⟶
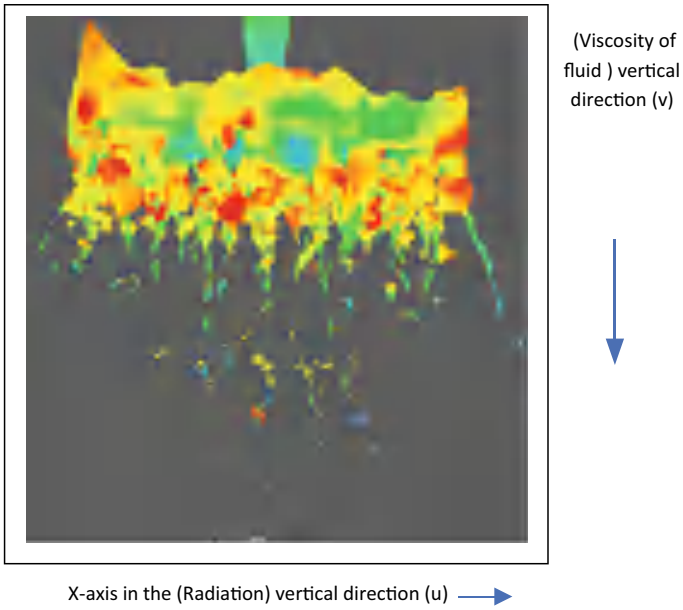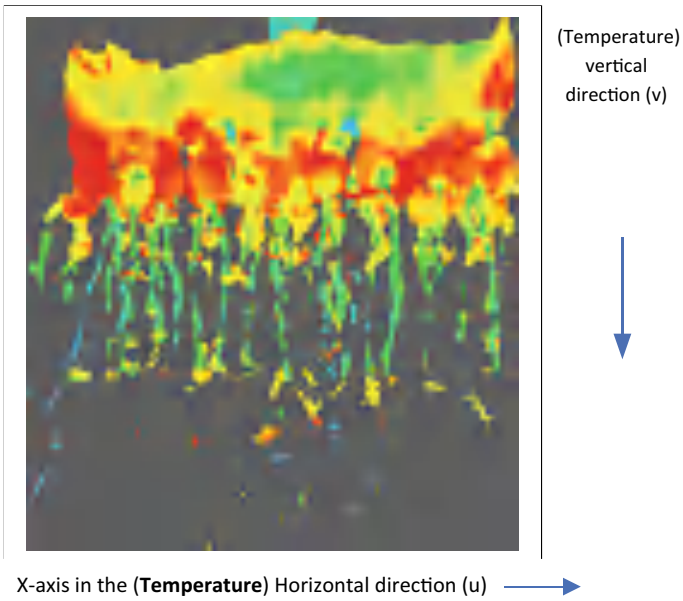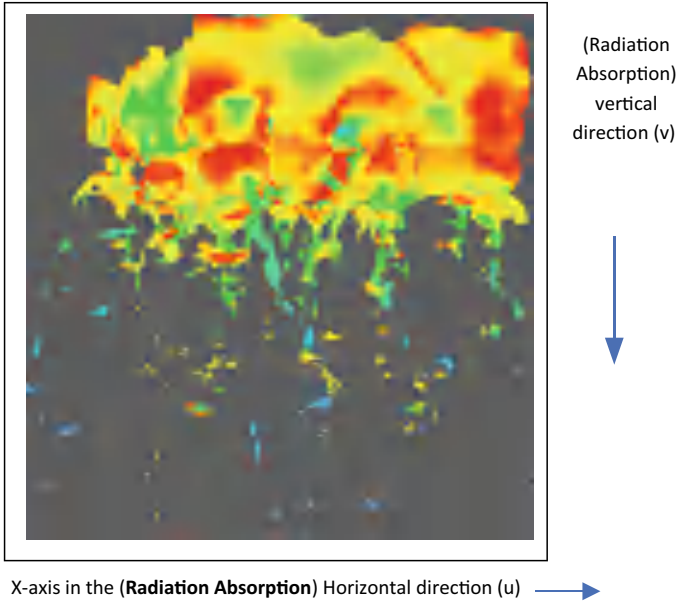
Fig. 13 Variation with radiation absorption
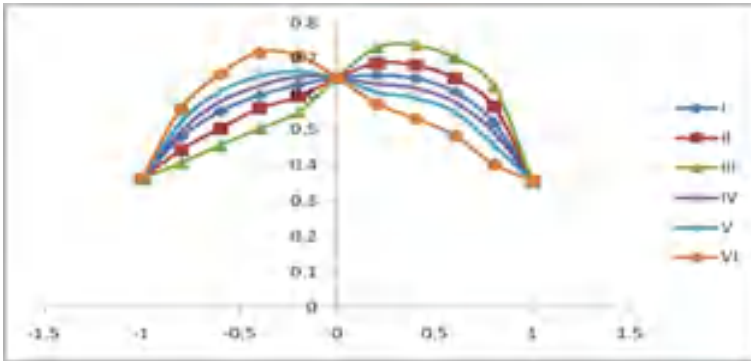
Fig. 14 Variation with temperature
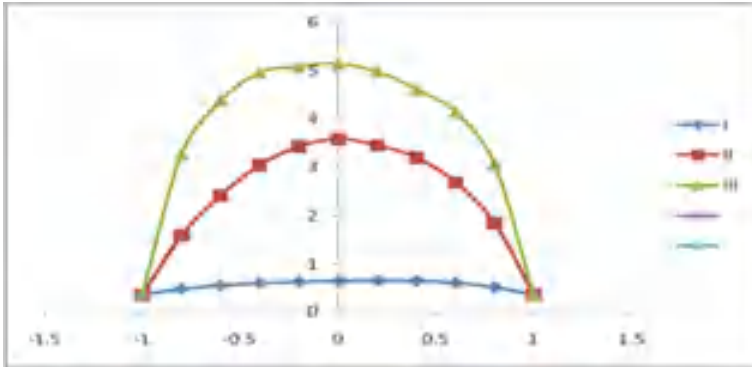
Fig. 15 Variation with radiation absorption
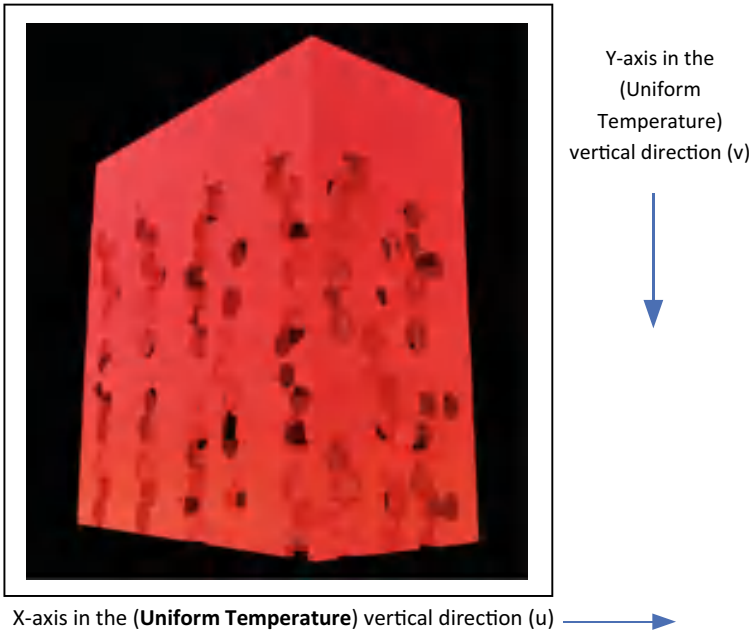
Y-axis in the (Uniform Temperature) vertical direction (v)

X-axis in the (**Uniform Temperature**) vertical direction (u)

Fig. 16 Variation of v with uniform temperature channel in wall temperature

Y-axis in the (Non uniform Temperature) vertical direction (v)

X-axis in the (Non **uniform Temperature**) vertical direction (u) ⟶
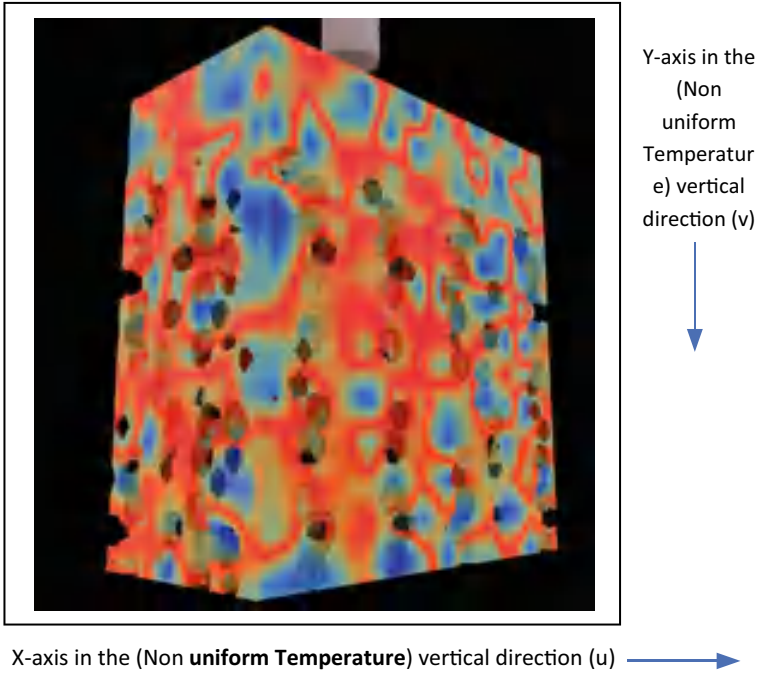
Fig. 17 Variation of *v* with non-uniform temperature channel in wall temperature
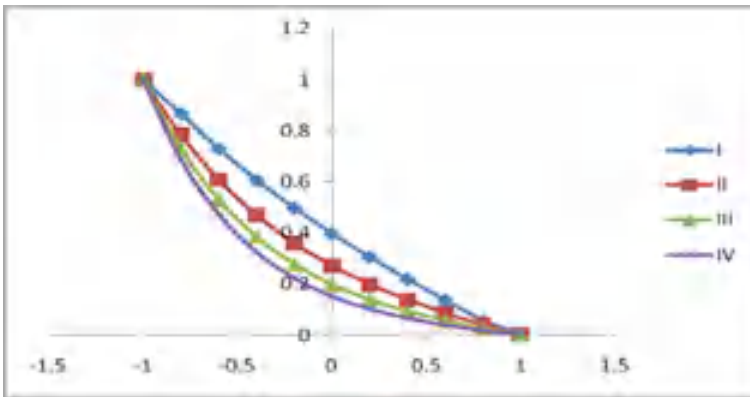
Fig. 18 Variation of *V* with uniform temperature channel in wall temperature
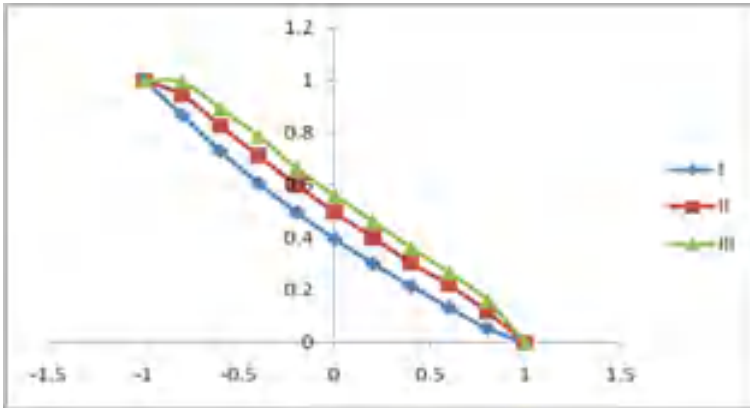
Fig. 19 Variation of *C* with non-uniform temperature channel in wall temperature

5 Conclusion

The analysis of various composition fluid systems is shown in Fig. 1, the model which has been observed in a microscopic scale by applying the attributes of fluid dynamics constrained visualizations, which gives appropriate results from Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19. Therefore, the coupled equations governing the flow and heat transfer were solved by using the perturbation technique, and 3D numerical simulations for the proposed adopted method are carried out results, and analysis presented.

References

1. Francesco P, Schrefler BA, Sciume G (2016) Multiphase flow in deforming porous media: a review. Arch Computat Methods Eng. https://doi.org/10.1007/s11831-016-9171-6 10 March 2016
2. Raja Kumari P et al (2022) Experimental analysis and improvements of a visible spectrophotometer for detection of nano materials. Int J Chem Eng 2022, Article ID 7952197, 10 p https://doi.org/10.1155/2022/7952197
3. Bojja P et al (2021) Anon-linear mathematical model based routing protocol for W ban based health-care systems. Int J Pervasive Comput Commun Int J Pervasive Comput Commun ISSN: 1742-7371 Article publication date: 13 January 2021 202
4. Heinemann ZE, Mittermeir G (2013) Fluid flow in porous media. February 2013
5. Bojja P et al (2022) A dynamic optimization and deep learning technique for detection of lung cancer in CT images and data access through Internet of Things, Springer, Wirel Pers Commun
6. Sreenivasa Reddy B (2006) Thermo–diffusion effect on convective heat and mass transfer through a porous medium, Ph.D. thesis, S.K. University, Anantapur, India. Polaiah
7. Bojja et al (2021) A non-linear mathematical model based routing protocol Wbsn-based health-care system. Int J Pervasive Comput Commun 2021. https://doi.org/10.1108/Ijpcc-09-2000-0138

8. Reddy R (2007) Hydro magnetic convection. Heat and Mass Transfer through a porous medium in channel/pipes with Soret effect, Ph.D. thesis submitted to JNTU, Hyderabad, 2007
9. Vijayabhaskar Reddy P, Effect of non-uniform temperature on convective heart and mass transfer in a vertical channel wirhsoret effect
10. Bojja P (2022) A novel method to solve visual tracking problem: hybrid algorithm of grasshopper optimization algorithm and differential evolution. Springer Link, 15(1):785–822. https://link.springer.com/article/https://doi.org/10.1007/s12065-021-00567-0

Otsu-Based Differential Evolution Method for Image Segmentation

Afreen Shaikh, Botcha Sharmila, Murali Krishna, and Sushil Kumar

Abstract This paper proposes an OTSU-based differential evolution method for satellite image segmentation and compares it with four other methods as Modified Artificial Bee Colony Optimizer (MABC), Artificial Bee Colony (ABC), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) using the objective function proposed by Otsu for optimal multilevel thresholding. The experiments conducted and their results illustrate that our proposed DE+OTSU algorithm segmentation can effectively and precisely segment the input image, close to results obtained by the other methods. In the proposed DE+OTSU algorithm, instead of passing the fitness function variables, the entire image is passed as an input to the DE algorithm after obtaining the threshold values for the input number of levels in the OTSU's algorithm. The image segmentation results are obtained after learning about the image instead of learning about the fitness variables. In comparison to other segmentation methods examined, the proposed DE+OTSU algorithm yields promising results with minimized computational time compared to some algorithms.

1 Introduction

Image segmentation plays a vital part as a pre-processing step in image processing. In particular, it aims at grouping and partitioning pixels within meaningful regions to analyse the image in a detailed manner. Image segmentation applies to varied Computer Vision tasks like Content-based image retrieval, Medical Imag-

A. Shaikh (✉) · B. Sharmila · M. Krishna · S. Kumar
National Institute of Technology, Warangal, India
e-mail: safreen@student.nitw.ac.in

B. Sharmila
e-mail: bsharmila@student.nitw.ac.in

M. Krishna
e-mail: mmuralikrishna@student.nitw.ac.in

S. Kumar
e-mail: kumar.sushil@nitw.ac.in

693

ing, Object detection and identification, and Object recognition tasks (like face and fingerprint), among others. Segmentation methods can be subdivided broadly into (a) edge and line-oriented segmentation methods, (b) region-growing methods, (c) clustering, and (d) region-splitting methods. Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Bacterial Foraging (BF), and Differential Evolution (DE) are regarded as popular choices for the segmentation of complex images due to their exhaustive searching capability. Because of its clearness and stability, thresholding is coined as one of the most popular image segmentation methods. Typically, satellite images have insignificant illumination features because of multiple kinds of environmental distributions. They contain diverse objects (regions), like vegetation, water bodies, and territory. However, these areas lack clear demarcation because of low spatial resolution. Hence, segmenting various land cover parts in these images is a complicated task. One of the key aspects of image segmentation is thresholding [4]. Many algorithms to perform global-level thresholding can be found in the existing literature, which aim at segmenting images and extracting compelling and meaningful patterns (Rosin 2001; Portesde et al. 2004; Zahara et al. 2005).

Otsu, in the year 1979 [5] presented a method for selecting thresholds from grey level histograms. But, inadequate articulation of between-class variance raises the cost of computation of the algorithm, specifically in the selection of the multilevel threshold. Otsu's thresholding aims at automatic threshold selection and region-based segmentation and is one of the most accurate techniques for image thresholding because of its simple calculation. Differential evolution is a reasonably recent population-based evolutionary model which depends on the mutation operation as its prominent step. It efficiently explores large search spaces and exhibits superior results in the following criteria: (1) shorter convergence speed than other evolutionary algorithms, (2) fewer parameter adjustments, making it particularly easy to implement. Tsai [6] uses the moment-preserving principle to determine thresholds of grey level input images—the Tsallis entropy technique, a prominent function used for image thresholding. Kittler et al. (1986) work with the assumption that the grey levels of every entity within an image are usually distributed. Kapur et al. [7] further introduced a robust method for grey level image thresholding by utilizing the histogram entropy [7].

Entropy-based approaches have earned popularity among all the distinctive image thresholding methods. PCO, influenced by social behaviours like flocking of birds or schooling of fishes [8–10], ant colony optimization (ACO) influenced by the pasturing conduct of ant colonies [11, 12] Artificial Bee Colony (ABC) inspired by the pasturing conduct of honey bee swarm [13] (Akay et al. 2013).

Another recent area of research inspired by the coordinated intelligence in a swarm of insects or animals has also emerged called Swarm Intelligence. One of the widespread and recently developed Swarm Intelligence-based techniques is the ABC algorithm, presented by Karaboga [14], that simulates the foraging behaviour of honey bee colonies. Numerous studies have demonstrated that the performance of the Artificial Bee Colony (ABC) method is competitive and comparable with other population-based techniques. The MABC method given by Bhandari et al. [17] introduces an advanced solution search formula (Gao et al. 2012) in charge of

its more promising search solution. Within this search equation, the bee explores only closer to the best solution of the iteration before improving exploitation. The results demonstrate that MABC is best suited for multilevel thresholding of images to attain optimal thresholds for satellite images in comparison to Artificial Bee Colony-based algorithms and Particle Swarm Optimization-based techniques. Evaluating the advantages, such algorithms are the preferred choice for finding the optimum thresholds in simple images. For example, the Genetic Algorithm (GA) and the improved Genetic Algorithm [15] are frequently in use in multilevel thresholding scenarios. The Swarm Intelligence-based computing methods can find efficient and optimal solutions for any objective function and have been widely used with an ability to generate highly accurate results in case of complex problems as well. It has also been found by statistical analysis that the Swarm Intelligence-based algorithms perform well in multilevel thresholding scenarios [16].

2 Brief Explanation of the Algorithms Used in This Study

2.1 Otsu's Thresholding

Bi-level thresholding In Bi-level thresholding, the goal is to find a threshold to minimize the variance between classes in the segmented image. Otsu's algorithm tends to achieve better results when two different peaks are present in the histogram of the original image, one corresponding to the background and the other to the foreground. The entire range of pixels is iterated, and Otsu's threshold is determined when the between-class variances are minimum. Hence, Otsu's threshold tends to be decided by the class with greater variance. Hence, Otsu's method tends to produce sub-optimal results when there is an occurrence of two or more peaks within the histogram of the image or if one of the classes has a significant variance.

The total mean and variance are calculated based on the following formulas: The entire set of pixels is distributed into two classes,

$C_1 \leftarrow$ pixels having grey levels $[1, t]$
$C_2 \leftarrow$ pixels having grey levels $[t + 1, ..., L]$.
where t corresponds to Otsu's threshold

The probability distribution of the two classes is denoted by:

$$C_1 : \frac{p_1}{w_1(t)}, ..., \frac{p_r}{w_1(t)} \tag{1}$$

and

$$C_2 : \frac{p_{r+1}}{w_2(t)}, ..., \frac{p_L}{w_2(t)} \tag{2}$$

where, $w_1(t) = \sum_{i=1}^{r} p_i$ and $w_2(t) = \sum_{i=r+1}^{L} p_i$

The class mean for the two classes μ_1 and μ_2 are defined by:

$$\mu_1 = \sum_{i=1}^{r} \frac{i * p_i}{w_1(t)}; \qquad \mu_2 = \sum_{i=r+1}^{L} \frac{i * p_i}{w_2(t)} \tag{3}$$

Otsu's between-class variance based on discriminate analysis of the threshold image is defined as:

$$\sigma_B^2 = w_1(\mu_1 - \mu_T)^2 + w_2(\mu_2 - \mu_T)^2 \tag{4}$$

For bi-level thresholding, the optimal threshold t^* is chosen to maximize between-class variance n, i.e.

$$t^* = arg_{t<l} max \sigma_B^2(t) \tag{5}$$

Multilevel Thresholding The first step in the algorithm is to obtain the Otsu's threshold and the class means of the two classes divided by the Otsu's threshold. Further, in multilevel thresholding, the pixels of the image are divided into three categories instead of two categories determined by Otsu's threshold. The three categories correspond to the following: (a) the 'foreground' region—a group of the pixel with values $>=$ the larger mean, (b) the 'background' region—a group of pixels with values $<=$ to the smaller mean and (c) the 'to-be-determined (TBD)' region—group of pixels with values between the two class means.

In iteration $i + 1$, the algorithm retains the 'foreground' and 'background' regions from iteration i and re-applies Otsu's method only on the 'to-be-determined' region to further divide it into three classes again. When this iteration stops after satisfying a pre-defined criterion, the final 'to-be-determined (TBD)' region is then divided into two classes—foreground and background instead of three. Lastly, the foreground regions from all the iterations are combined to get the final foreground class, and the final background region is also determined likewise.

For an image represented by L no.of grey levels $0, 1, ..., L − 1$, we could develop the image histogram $H = \{f_0, f_1, ..., f_{L-1}\}$, where f_i is the frequency of grey level i in the image. Let $N = \sum_{i=0}^{L-1} f_i$ determine the total no. of pixels in the image. The occurrence probability of ith grey level is defined by $p_i = \frac{f_i}{N}$.

It can be easily be illustrated that $p_i \geq 0$ and $\sum_{i=0}^{L-1} p_i = 1$.

Otsu's algorithm divides the image into $K + 1$ clusters $\{C0, C1, ..., C_K\}$ using K no. of thresholds chosen from the set $T = \{(t_1, t_2, ..., t_K)|0 < t1 < ... < t_K < L\}$ where C_K is the set of pixels with grey levels $\{t_K, t_K + 1, ..., t_{K+1} − 1\}$. Where $t_0 = 0$ and $t_{K+1} = L$. For every cluster C_K, the cumulative probability w_K and mean grey level μ_K are defined by:

$$w_k = \sum_{i \epsilon C_k}, k\epsilon\{0, 1, 2, ..., K\}; \qquad \mu_k = \sum_{i \epsilon C_k} \frac{i * p_i}{w_k}, k\epsilon\{0, 1, 2, ..., K\} \tag{6}$$

The mean intensity of the whole image μ_T and the between-class variance σ_B^2 are defined by:

$$\mu_T = \sum_{k=0}^{K} w_k * \mu_k = \sum_{i=0}^{L-1} i * p_i; \qquad \sigma_B^2 = \sum_{k=0}^{K} w_k * (\mu_k - \mu_T)^2 = \sum_{k=0}^{K} w_k * \mu_k^2 - \mu_T^2 \quad (7)$$

In Otsu's method, the threshold levels for each cluster are selected on the basis of maximizing the variance among means of the cluster (Huang et al. 2011). The optimal thresholds at maximum between-class variance are determined by:

$$(t_1^*, t_2^*, ..., t_K^*) = \{\sigma_B^2(t_1, t_2, ..., t_K)\} \qquad (8)$$

3 Differential Evolution (DE)

In comparison to other evolutionary algorithms, the differential evolution (DE) algorithm aims to reserve the global search strategy based on population and employs a simple mutation function of the differential and one-on-one competition, reducing the operation's genetic complexity. At the same time, the specific memory ability of DE allows it to dynamically track the current search to adjust its search strategy with robust global convergence. Hence, it is suited for the complex environments of the optimization problem. Fundamental operations like selection, crossover, and mutation are the foundation of the differential evolution algorithm.

3.1 How de Is Used in the Proposed Algorithm?

Initially, the image to be segmented is made to run through OTSU's multilevel thresholding on the basis of the number of levels(L) the image has to be segmented. The output of this step yields us 'L' threshold values which act as partitions for the pixel intensities. These partitions divide the pixel intensity range (0–255) into '$L+1$' clusters. Based on the pixels belonging to the image that fall into a particular cluster, the cluster centre is defined using the mean of the pixels (falling in that cluster).

Now, the image to be segmented is given as input to the DE algorithm. Each of the 'G' generations has Np number of populations, where population size is equal to the count of the image pixels. After each generation, the best population is evaluated based on the least value of the fitness function, which is a measure of the mean square error (MSE) distance of each pixel in the population to its corresponding cluster centre. The next generations are populated using this best population, as described later in the algorithm. The best population of the last generation is again reshaped into image format, which forms the segmented image.

4 Proposed Algorithm

4.1 Algorithm Pre-processing

In our implementation an input image and the number of levels (L) are given as input to the algorithm which would return the cluster partition values (cp_1, cp_2, ..., cp_L). Cluster ranges, $(0, cp_1)$, $(cp_1 + 1, cp_2)$, ..., $(cp_L + 1, 255)$ are computed from the cluster partition values obtained after the thresholding of the image. Cluster centres, (cc_1, cc_2, ..., cc_{L+1}) are calculated as:

$$cc_i = \frac{\sum pix^i}{|pix^i|} \tag{9}$$

where, pix^i = pixels in the cluster range $(0, cp_i)$ and $|pix^i|$ = pixels in the cluster range $(0, cp_i)$

4.2 Differential Evolution

Consider the image has 'm' pixels (pix_1, pix_2, .., pix_m), and algorithm runs for 'G' generations with 'Np' no. of populations per generation. $X_{n,i}^g$: ith pixel value of nth population in gth generation

$$X_n^g : [X_{n,1}^g, X_{n,2}^g, ..., X_{n,m}^g]$$
$$n : 1 \rightarrow Np(population index)$$
$$g : 1 \rightarrow G(generation index) \text{ and}$$
$$i : 1 \rightarrow m(pixel index)$$

5 Results

5.1 Performance Metrics

This section aims at presenting results gathered from various satellite images using the proposed OTSU-based differential evolution algorithm are discussed, and those results are set side by side with the results obtained from the MABC [11], ABC, PSO, and GA algorithms. Along with the quality estimation factor of Peak Signal-to-Noise Ratio (PSNR), algorithm efficiency (based on CPU Timing) and feature assessment is also measured using the Structural Similarity Index Measure (SSIM). Tables 1, 2 and 3 compare the PSNR, SSIM, and CPU time values gathered by applying the proposed DE+OTSU method and set side by side for comparison with the result obtained using MABC, ABC, PSO, and GA methods, respectively.

Algorithm 1 An algorithm with caption

$n = 0$
$i = 0$
while $n \neq 0$ **do**
   while $i \neq P$ **do**
      $[Xlow, Xhigh] \leftarrow cluster\,range\,for\,pixel, i$
      $X_{n,i}^1 \leftarrow X_{low} + rand(0,1) * [X_{high} - X_{low}]$
   end while
end while
$n = 0$
while $n \neq 0$ **do**
   $MSE_{X_n^{g=1}} = \frac{1}{n} \sum_{i=1}^m (X_{n,i}^1 - C_i)^2$
   $X_{best}^1 \leftarrow population\,with\,minimum\,fitness\,value$
end while
Mutant vector :
for each population n in generation g+1
   $V_n^{g+1} \leftarrow X_{best}^g + F * (X_{r1}^g - X_{r2}^g)$
   if $rand(0,1) <= Cr$ **then**
      $U_n^{g+1} \leftarrow V_n^{g+1}$
   else if $rand(0,1) > Cr$ **then**
      $U_n^{g+1} \leftarrow X_n^g$
   end if
   if $fitness(U_n^{g+1} < fitness(X_n^g)$ **then**
      $X_n^{g+1} \leftarrow U_n^{g+1}$
   else if $fitness(U_n^{g+1} > fitness(X_n^g)$ **then**
      $X_n^{g+1} \leftarrow X_n^g$
   end if
    Fitness function and mutant vector calculation will be repeated for G generations. After G generations the X_{best}^G population is considered the solution to image segmentation and the 'm' pixels in X_{best}^G is converted into output image

---

5.2 Input Images and References

See Fig. 1.

5.3 Comparison of Results

Comparison between Modified Artificial Bee Colony Optimizer (MABC), Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA) algorithms with proposed DE+OTSU based on CPU timing in seconds (Columns A), PSNR in dB (Columns B) and SSIM (Columns C)

Table 1 Comparison using Kapur's entropy

Image	Level	DE+OTSU			Kapur's entropy												
					MABC			ABC			PSO			GA			
		A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	
1	2	10.620	21.430	0.770	2.840	24.540	0.900	6.890	24.530	0.870	21.010	24.550	0.940	185.940	24.580	0.940	
	3	10.960	22.130	0.710	2.490	24.760	0.970	6.640	24.790	0.950	22.240	24.740	0.970	181.800	24.750	0.960	
	4	8.260	22.960	0.790	2.530	24.970	0.980	6.640	25.010	0.970	21.850	24.960	0.980	184.050	25.890	0.970	
	5	9.920	23.940	0.790	3.180	25.820	0.990	6.660	25.760	0.980	23.070	25.180	0.980	121.460	25.670	0.980	
2	2	12.570	22.770	0.730	2.950	24.620	0.890	6.510	24.600	0.880	24.010	24.570	0.950	268.020	24.590	0.950	
	3	10.560	22.850	0.790	2.490	24.800	0.960	6.300	24.800	0.950	23.270	24.840	0.970	266.780	24.810	0.970	
	4	8.690	23.720	0.810	2.620	25.020	0.980	7.010	25.000	0.970	23.860	25.000	0.980	268.800	25.970	0.980	
	5	10.650	23.810	0.860	2.730	25.540	0.990	6.570	25.230	0.980	24.950	26.100	0.990	183.320	25.540	0.980	
3	2	9.930	23.360	0.720	2.820	24.640	0.920	8.240	24.650	0.900	25.230	24.640	0.940	436.110	24.690	0.950	
	3	11.360	23.530	0.752	3.287	24.830	0.961	8.076	24.860	0.960	24.809	24.840	0.968	433.874	24.900	0.963	
	4	8.110	23.990	0.773	2.612	25.060	0.978	6.917	25.030	0.962	24.867	25.060	0.979	438.918	25.970	0.976	
	5	10.340	24.140	0.804	2.821	24.540	0.989	8.248	24.540	0.978	25.734	24.550	0.985	296.151	24.580	0.981	

Table 2 Comparison using between-class variance

Image	Level	DE+OTSU			Between class variance													
					MABC			ABC			PSO			GA		
		A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
1	2	10.620	21.430	0.772	2.252	24.564	0.907	2.512	24.580	0.899	13.235	24.546	0.934	121.927	24.553	0.936
	3	10.960	22.130	0.712	2.384	24.780	0.958	2.495	24.760	0.950	4.301	24.732	0.961	118.788	24.764	0.964
	4	8.260	22.960	0.791	2.243	24.981	0.978	2.596	25.007	0.977	4.000	25.012	0.977	119.733	24.874	0.925
	5	9.920	23.940	0.792	2.229	25.628	0.985	2.434	25.473	0.978	4.337	25.198	0.984	119.813	25.005	0.972
2	2	12.570	22.770	0.731	3.245	24.622	0.886	3.504	24.583	0.870	5.748	24.586	0.945	174.337	24.570	0.995
	3	10.560	22.850	0.794	3.437	24.800	0.958	3.589	24.795	0.958	5.723	24.837	0.967	174.190	24.787	0.963
	4	8.690	23.720	0.812	3.440	24.968	0.976	3.745	25.005	0.979	5.671	24.974	0.978	174.576	24.975	0.960
	5	10.650	23.810	0.855	3.407	25.771	0.982	3.733	25.301	0.980	5.793	25.235	0.981	173.696	25.158	0.947
3	2	9.930	23.360	0.721	4.668	24.654	0.921	5.008	24.648	0.904	7.297	24.645	0.942	272.890	24.646	0.943
	3	11.360	23.530	0.752	5.364	24.856	0.964	5.276	24.870	0.969	6.588	24.860	0.969	275.032	24.871	0.963
	4	8.110	23.990	0.773	5.076	25.068	0.971	5.520	25.055	0.973	7.154	25.066	0.978	272.785	26.280	0.955
	5	10.340	24.140	0.804	4.785	25.769	0.982	5.428	25.082	0.968	6.754	25.294	0.982	275.444	25.124	0.956

Table 3 Comparison using Tsallis entropy

| Image | Level | DE+OTSU | | | Tsallis entropy | | | | | | | | | | | |
| | | | | | MABC | | | ABC | | | PSO | | | GA | | |
		A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
1	2	10.620	21.430	0.772	4.192	24.378	0.907	9.295	24.351	0.899	5.286	24.519	0.934	1287.455	24.553	24.105
	3	10.960	22.130	0.712	4.289	24.637	0.958	10.444	24.578	0.950	5.166	24.657	0.961	1243.575	24.764	25.930
	4	8.260	22.960	0.791	5.003	24.961	0.978	9.486	24.752	0.977	5.261	24.982	0.977	129.445	24.874	25.065
	5	9.920	23.940	0.792	5.065	25.449	0.985	8.978	25.554	0.978	5.146	24.861	0.984	122.208	25.005	26.550
2	2	12.570	22.770	0.731	4.687	24.488	0.886	9.904	24.585	0.870	7.548	24.651	0.945	1804.159	24.570	24.748
	3	10.560	22.850	0.794	5.840	24.867	0.958	8.869	24.583	0.958	6.745	24.560	0.967	1813.836	24.787	25.455
	4	8.690	23.720	0.812	5.344	24.964	0.976	9.337	24.783	0.979	6.554	25.054	0.978	178.122	24.975	25.260
	5	10.650	23.810	0.855	6.290	26.764	0.982	9.073	25.037	0.980	7.058	25.151	0.981	180.250	25.158	26.947
3	2	9.930	23.360	0.721	6.721	24.649	0.921	8.936	24.478	0.904	7.839	24.656	0.942	2672.361	24.646	25.288
	3	11.360	23.530	0.752	6.710	24.830	0.964	10.747	24.645	0.969	7.785	24.770	0.969	2644.285	24.871	25.409
	4	8.110	23.990	0.773	6.904	24.865	0.971	10.543	24.871	0.973	7.210	25.057	0.978	297.826	26.280	25.197
	5	10.340	24.140	0.804	7.755	25.093	0.982	10.172	24.859	0.968	7.335	25.065	0.982	263.150	25.124	26.928
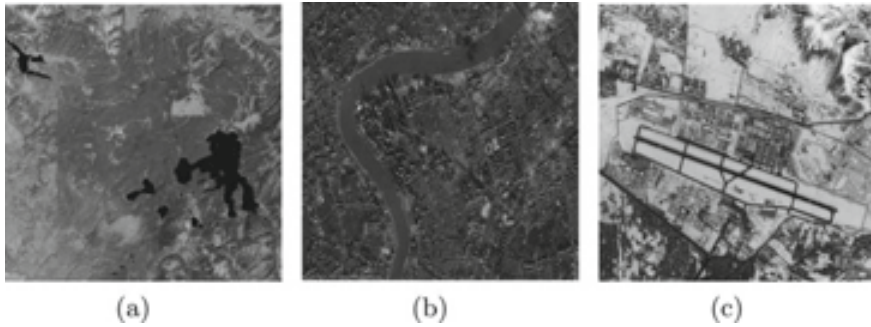
Fig. 1 Satellite images used to perform and compare the experiments **a** NASA Earth Observatory, Yellowstone National Park, June 2009, **b** Pléiades Satellite Image: Shanghai, China, **c** Pléiades Satellite Image: Kabul Airport, Afghanistan

6 Conclusions

In this study, an algorithm for image segmentation tasks based on a combination of differential evolution and Otsu's multilevel thresholding was developed, and the results were experimentally examined against three different satellite images. The study also explored the comparison between the proposed algorithm and the developed MABC, ABC, PSO, and GA using Kapur's entropy, Between-class variance, and Tsallis entropy (one at a time) based on the three satellite images based on PSNR, SSIM, and CPU time parameters. The proposed differential evolution-based Otsu method (DE+OTSU) yields promising results which are close to the results obtained from the other algorithms (MABC, ABC, and PSO), and we observed to be better than the Genetic algorithm (GA). On proper parameter setting, there is a positive probability of increased performance results when using the proposed algorithm.

References

1. Varalakshmi PNK, Uma Devi D (2012) Image segmentation of dermatitis images using differential evolution. Int J Eng Res Technol (IJERT) 01(08)
2. Kachitvichyanukul V (2012) Comparison of three evolutionary algorithms: GA, PSO, and DE. Ind Eng Manag Syst 12:215–223. https://doi.org/10.7232/iems.2012.11.3.215
3. Bhandari AK, Kumar A, Singh GK (2015) Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur's, Otsu and Tsallis functions. Expert Syst Appl 42(3):1573–1601, ISSN 0957-4174. https://doi.org/10.1016/j.eswa.2014.09.049
4. Agrawal S, Panda R, Bhuyan S, Panigrahi BK (2013) Tsallis entropy based optimal multilevel thresholding using cuckoo search algorithm. Swarm Evol Comput 11:16–30, ISSN 2210-6502. https://doi.org/10.1016/j.swevo.2013.02.001

5. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst, Man, Cybern 9(1):62–66. https://doi.org/10.1109/TSMC.1979.4310076. Jan.
6. Tsai W-H (1985) Moment-preserving thresolding: a new approach. Comput Vis, Graph, Image Process 29(3):377–393, ISSN 0734-189X. https://doi.org/10.1016/0734-189X(85)90133-1
7. Kapur JN, Sahoo PK, Wong AKC (1985) A new method for gray-level picture thresholding using the entropy of the histogram. Comput Vis, Graph, Image Process 29(3):273–285, ISSN 0734-189X. https://doi.org/10.1016/0734-189X(85)90125-2
8. Akay B (2013) A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding, Appl Soft Comput 13(6):3066-3091, ISSN 1568-4946. https://doi.org/10.1016/j.asoc.2012.03.072
9. Maitra M, Chatterjee A (2008) A hybrid cooperative-comprehensive learning based PSO algorithm for image segmentation using multilevel thresholding. Expert Syst Appl 34(2):1341–1350, ISSN 0957-4174. https://doi.org/10.1016/j.eswa.2007.01.002
10. Yin P-Y (2007) Multilevel minimum cross entropy threshold selection based on particle swarm optimization. Appl Math Comput 184(2):503–513, ISSN 0096-3003. https://doi.org/10.1016/j.amc.2006.06.057
11. Tao W, Jin H, Liu L (2007) Object segmentation using ant colony optimization algorithm and fuzzy entropy. Pattern Recognit Lett 28(7):788–796, ISSN 0167-8655. https://doi.org/10.1016/j.patrec.2006.11.007
12. Ye Z, Zheng Z, Yu X, Ning X (2005) Automatic threshold selection based on ant colony optimization algorithm. In: IEEE International conference on neural networks and brain, vol 2, pp 728–732
13. Cuevas E, Sencion F, Zaldivar D, Perez-Cisneros M, Sossa H (2012) A multi-threshold segmentation approach based on artificial bee colony optimization. Appl Intell 37(3):321–336
14. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization, Technical Report-TR06. Erciyes University, Kayseri, Turkey, p 200
15. Zhang J, Liu J (2014) Two triggered information transmission algorithms for distributed moving horizon state estimation. Syst Control Lett 65:1–12, ISSN 0167-6911. https://doi.org/10.1016/j.sysconle.2013.12.003
16. Kurban T, Civicioglu T, Kurban R, Besdok E (2014) Comparison of evolutionary and swarm based computational techniques for multilevel color image thresholding. Appl Soft Comput 23:128–143, ISSN 1568-4946. https://doi.org/10.1016/j.asoc.2014.05.037
17. Bhandari AK, Singh VK, Kumar A, Singh GK (2014) Cuckoo search algorithm and wind driven optimization based study of satellite image segmentation for multilevel thresholding using Kapur's entropy. Expert Syst Appl 41(7):3538–3560, ISSN 0957-4174. https://doi.org/10.1016/j.eswa.2013.10.059
18. Cuevas E, Zaldivar D, Cisneros M (2016) Image segmentation based on differential evolution optimization. https://doi.org/10.1007/978-3-319-26462-2-2
19. Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim 11:341–359
20. Soleimani H, Kannan G (2015) A hybrid particle swarm optimization and genetic algorithm for closed-loop supply chain network design in large-scale networks. Appl Math Model 39(14):3990–4012, ISSN 0307-904X. https://doi.org/10.1016/j.apm.2014.12.016
21. Choubey S (2017) Genetic algorithm, particle swam optimization and differential evolution algorithm for optimal hydro generation scheduling. In: International conference on intelligent computing and control systems (ICICCS) 2017, pp 486–491. https://doi.org/10.1109/ICCONS.2017.8250770
22. Xuan Pham T, Siarry P, Oulhadj H (2018) Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation. Appl Soft Comput 65. https://doi.org/10.1016/j.asoc.2018.01.003

Emotion Detection Using Machine Learning and Deep Learning

G. Rajesh Kumar , **D. Srinivasa Rao** , **N. Rajasekhar** ,
Ch. Ramesh Babu , **Ch. Rohini, T. Ravi** , **and N. Mangathayaru**

Abstract The interaction between human and computer for some real application like driver state surveillance, personalized learning, health monitoring, etc. Most reported facial emotion recognition systems, however, are not fully considered subject-independent dynamic features. The main objective of the emotion detection model is to detect the emotions of the people by scanning their pictures in the model. The model detects the emotion of the person in the photograph. The emotions that can be detected in the model proposed by us is sad, angry, happy, shocked, surprised, neutral, disgust, etc. This project is focusing on detecting the emotion based on input human face images. The system can also be integrated as a part of video streaming. In the video stream some random pictures are taken to classify the emotion.

Keywords Emotion detection · Image processing · Classification · Convolutional neural networks (CNN)

1 Introduction

The interaction between people and computers for practical purposes such tracking a driver's state, individualized instruction, health monitoring, etc. However, the majority of reported dynamic features are not entirely regarded as subject-independent dynamic features. Facial expressions serve as a platform for non-verbal commu-

G. Rajesh Kumar (✉) · D. Srinivasa Rao · Ch. Rohini · T. Ravi · N. Mangathayaru
Faculty of Information Technology, VNR Vignana Jyothi Institute of Engineering
and Technology, Hyderabad, India
e-mail: gunupudirajesh@gmail.com
URL: http://vnrvjiet.ac.in/itdept/faculty/8.pdf

D. Srinivasa Rao
e-mail: srinivasarao_d@vnrvjiet.in

N. Rajasekhar
Department of IT, GRIET, Hyderabad, India

Ch. Ramesh Babu
Department of IT, MEC, Hyderabad, India

nication in human-computer interactions. Therefore, it is quite difficult to identify emotions in real-world applications. A facial expression recognition system must get past the human face's numerous variations in colour, orientation, expression, posture, texture, and other aspects.

The internet offers a huge range of possibilities, making it time-consuming to find the most pertinent information while searching. This proposal system's capabilities include the capacity to work with various types of data with ease.

There are many algorithms for classifying photos and identifying them. The estimate of facial landmarks using support vector machines (SVM) [1], K-nearest neighbours (KNN) [2], logistic regression, etc. The Local Binary Pattern (LBP) is a technique that thresholds the area around each pixel in a picture, labels the pixels, and interprets the output as a binary integer. Because of illumination and feature extraction, pre-processing of the images is necessary for any strategy to be successful. Many CNN models have already been developed, however there is still room to increase accuracy and make the algorithm mobile-friendly.

1. Preprocessing
2. Facial Image Preprocessing
3. Feature Extraction Techniques.

1.1 Preprocessing

The image is subjected to any pre-image processing activities in order to create a reliable and quick performance through CNN. The pre-processing techniques listed below [3, 4] are employed in picture processing: Data augmentation, image sharpening, histogram scaling, grayscale and redimensioning, face detection and cropping, etc.

Normalization: An image is normalized in order to remove discrepancies and create a better representation of the face.

Gray Scaling: Gray scaling is a technique for transforming a picture to a pixel density based on the amount of light present in the image.

Re-dimensioning: Re-dimensioning involves enlarging the image to cut out any extraneous details. Unquestionably, this lowers the memory requirement and speeds the calculation.

1.2 Facial Image Preprocessing

There are many factors that can affect CNN performance such as dense background, brightness, and deviation. Applying pre-filter filters may result in improved precision in classifying facial conditions. For example, sharpening images can enhance the edges of important features such as the mouth and eyes. These conclusions are important in predicting facial expressions. Histogram measurement helps to distinguish the front and back when both colours are the same [5–7].

Data Augmentation: We used Keras Image Data Generator to optimize data. Produces 32 advanced images in a single image by rotating, browsing, and applying other specific techniques as shown in Fig. 1.

Image Sharpening: The Engagear Mask produces an original picture mask using a blurred or negative image. Combining a genuine image with a blurred image results in a lovely "slightly blurred" final image. The ability to regulate the sharpening process makes utilizing Unsharp Mask superior to other sharp filters like Gaussian High Pass. Adjustable parameters are available with Unsharp Mask. By examining and comprehending the FER-2013 photographs [8], we discovered that they were a little hazy, and that sharpening in these images helped to define the edges of the dominant features, such as the eyes and mouth, which are suitable for identifying human emotions as displayed [9] in Fig. 2.
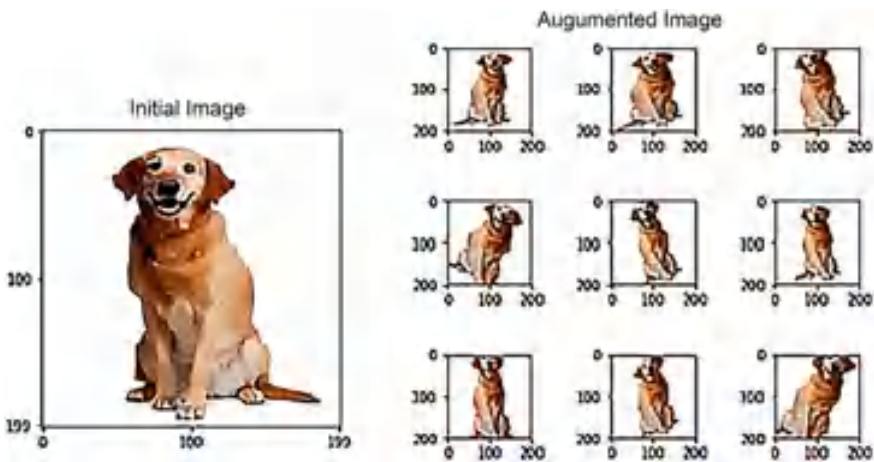
Fig. 1 Data augmentation
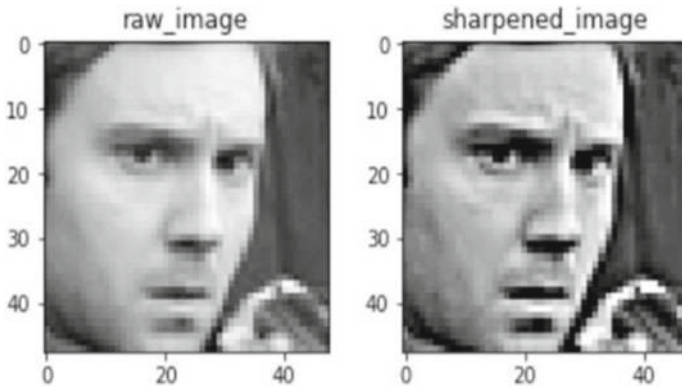
Fig. 2 Image sharpening

1.3   Feature Extraction

The first batch of raw data is separated and dispersed into control groups as part of the size reduction procedure, which also includes feature extraction. As a result, processing will be simple. The fact that these big datasets are extremely flexible is their most crucial quality. Processing this variable uses a lot of computing power. By choosing and incorporating dynamics into features, feature extraction successfully reduces the amount of data while obtaining the optimal feature from those enormous datasets. These features can accurately and precisely describe an actual set of data while still being simple to analyse.

2   Literature Survey

The process of recognizing someone's emotions is known as emotional recognition, and it is closely related to a number of fields, including human-computer interaction, the processing of emotions in people, irrational analysis, medical diagnosis, data-driven animations, human-robot communication, and much more.

The centre of the mind is another name for the face. The face can convey a number of diverse messages, much like a range of facial touches. When translating computers, these advantageous characteristics can increase the security and compatibility of [10, 11] human-machine interactions. Facial expressions were discussed as a good source of information for arranging a person's genuine feelings1. One of the most crucial non-verbal processes is the recognition of facial expression (FER) [12], which enables human-machine interfaces (HMI) [12, 13] to comprehend a person's emotions and intentions. This software utilizes split functions.

Fig. 3 The basic framework of applications for facial expression analysis

Face detection involves numerous steps, including face recognition, face image alignment, facial feature retrieval, preview processing, and image detection. Essentially, there are two different sorts of output elements: a method that concentrates on entire mathematical properties and one that employs the geometric attribute domain. The geometrical-based technique is frequently used to describe the placement of face characteristics as segregation determinants shown in Fig. 3.

This study's major goal was to assess and comprehend the advantages of adopting convolutional neural network models over other in-depth study models. Using a webcam and the described model, it is possible to categories people's faces in real time.

2.1 Analysis of Facial Expression

Many fields, including clinical psychology, psychiatry, neurology, pain evaluation, incorrect diagnosis, intelligent settings, and multimodal human-computer interfaces can benefit from the application of automatic facial analysis (AFEA) (HCI) [14–19]. In a typical AFEA, there are three steps: face collecting, facial data extraction and representation, and facial recognition. Fundamentally, there are two different approaches to extracting facial features: geometric or feature-based guessing methods, and visual-based methods. In this article, the writers employ techniques based on the look of statistics shown in Fig. 4.

2.2 Facial Emotion Recognition (FER)

Using EEG: Earlier some of the similar systems were built for human emotion detection. Some of the systems used Electroencephalography (EEG) to capture the
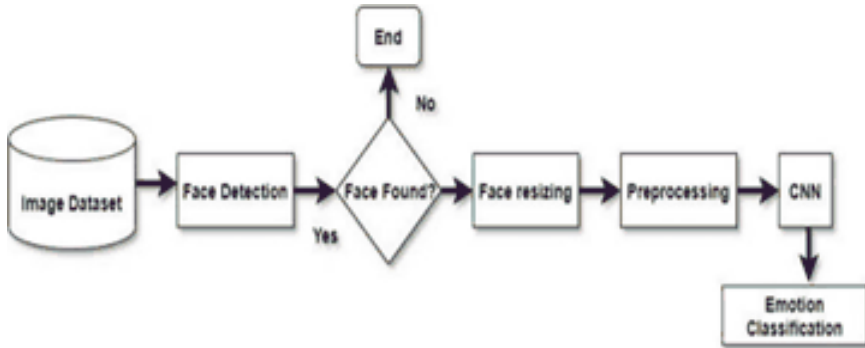
Fig. 4 Flowchart for the emotion detection classifier

human brain activity and predict their emotions but this was not used later because it was complex, needed costly equipment, and was efficient to predict only some of the emotions. So, this method was ignored [?]. Algorithms used are:

1. K-Nearest Neighbours
2. Multilayer Perception
3. Support Vector Machines

Using IRTI images: Some people also used images from Infrared Thermal Imaging Cameras to develop systems that can use predict the emotion of a person. This gave better results, but the system is so costly and can't be used in smaller organizations. the classification of expressions becomes a tough job because of daily conditions like a shadow, refection, and darkness (or low light). Databases used:

1. Japanese female facial expression (JAFFE) [20]
2. Cohn cade database

Different algorithms used in this system are:

1. The facial features were extracted using a histogram of oriented gradients (HOG), and then a sparse representation classifier (SRC) was used to classify the facial expressions Combination of Adaboost and support vector machine (SVM) to recognize facial expressions [20–24].

2.3 Bi-linear Convolution Neural Network

Dammavalam et al. [25] proposed method with are threefold. Firstly, a bi-linear convolution neural network (Bi-CNNs) for plant leaf disease identification and classification is projected. Then, we fine-tune VGG and pruned ResNets and exploit them as feature extractors and unite them to totally connected dense networks. The

hyperparameters are regulated to reach quicker convergence and attain improved simplification for the duration of stochastic optimization of Bi-CNN(s). Lastly, the projected model is planned to influence scalability by entailing the Bi-CNN model into a real-world application and announcement it as an open-source.

DWT, DCT, and SWT image fusion approaches are proposed and implemented for image fusion. Fused outcomes are assessed through varous evaluation factors. Because of the basic possessions and potentiality of the SWT centred fused results outdoes DWT and DCT-based fused methods. Subsequently SWT fused outcomes DCT variance and CV with DCT variance-based fused results upgraded fused image content [26].

The projected work is the comparison between neuro fuzzy-based image fusion and iterative neuro fuzzy fusion methods along with quality assessment metrics for image fusion; image quality index, mutual information measure, fusion factor, fusion symmetry, fusion index, root mean square error, peak signal to noise ratio, entropy, correlation coefficient, and spatial frequency. Investigational outcomes attained from projected technique demonstrate that the use of the iterative neuro fuzzy fusion method can professionally reserve the spectral evidence while enhancing the spatial perseverance of the remote sensing and medical imaging [27].

3 Methodology

The one versus all (OVA) method will be used to train and predict the even fundamental emotions in the detection and recognition implementation provided here (anger, contempt, disgust, fear, happiness, sadness, and surprise). A Viola-Jones course object face finder is used to extract the general face from the image initially. By using simple highlights known as Haar-like elements, the Viola-Jones identification system seeks to differentiate faces or components of a face (or distinct articles).

In order to interact, one must ignore highlight boxes in a picture and determine the difference in additional pixel values between adjacent areas. The important factor is then put in contrast with an edge that shows whether or not an article is thought to be distinct. This calls for having pre-planned limitations for various element boxes and highlights. The use of explicit element boxes for facial elements is predicated on the idea that most faces and the highlights included inside them will adhere to general guidelines. Basically, it will generally hold that some areas will be lighter or hazier than surrounding areas in a component district of interest on the face.

For example, the nose is generally more dazzling than the sides of the face that are directly next, or more brilliant than the upper lip and nose span region. A nose can therefore be identified if a suitable Haar-like component, as those shown in, is used and the difference in pixel total for the nose and the surrounding areas outperforms the edge. It should be noted that Haar-like highlights are quite simple, making them weak classifiers that demand multiple runs. The faces are then removed and scaled to a preset dimensional standard after being found.
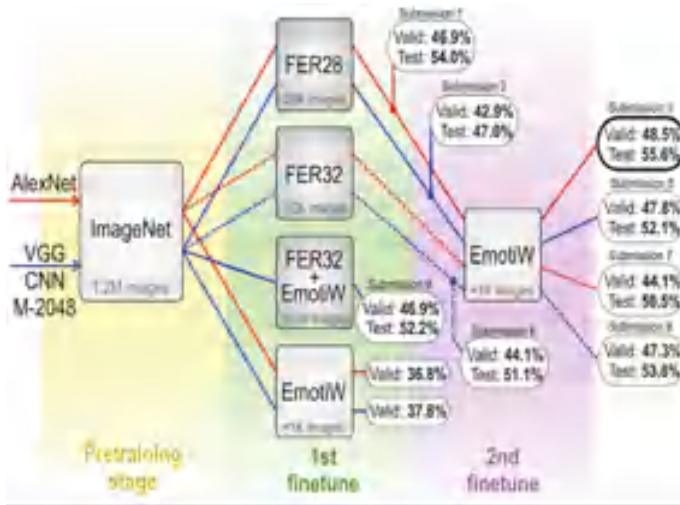
Fig. 5 Hybrid CNN by combining different image datasets

We will scale the retrieved faces to 100×100 pixels since Zhang has demonstrated that lesser resolution (64×64) is sufficient. This will make doing the additional analysis with less computing burden. The mean image for each training face will then be determined. Faces from the Extended Cohn-Kanade [3] dataset, which include those that reflect the fundamental emotions, make up the entirety of the training set. The remaining images in the preparation set are subtracted from the mean image.

The dissipate grid S is then formed using the mean-deducted preparation set. The objective is to choose a change in assumptions that will allow us to transmit our facial information in a more simplified manner. By doing so, it will be possible to keep the majority of the data in its original form as a direct fusion of the smaller aspect set. PCA accomplishes this by attempting to increase the difference between the original data and the new premise. To avoid processing difficulties, we first determine the eigenvalues and eigenvectors of the grid SHS before doing PCA on the data using the Sirovich and Kirby technique [23].

Then, by extending the eigenvector grid by S, the eigenvectors of the dissipate framework, denoted as SSH, may be recovered. We can successfully reduce dimensionality while successfully retaining most of the information by projecting our training data onto the top eigenfaces, in this case the 100 associated with the highest eigenvalues, by keeping the top eigenvectors, also known in this context as eigenfaces. In a reduced dimensionality, this enables us to move on to the Fisher linear discriminant analysis (LDA). To achieve a clear class separation between the class of interest and the other classes, we will execute Fisher LDA for each emotion for which we aim to train a predictor. The objective function that minimizes within class variance and maximizes between class variance will be optimized shown in Fig. 5.

4 Architecture

4.1 LeNet5

The organization has 5 layers with learnable boundaries and subsequently named Lenet-5. It has three arrangements of convolution layers with a blend of normal pooling. AlexNet: The AlexNet has eight layers with learnable boundaries. The model comprises of five layers with a blend of max pooling followed by 3 completely associated layers and they use Relu initiation in every one of these layers aside from the result layer.

4.2 Approaches Used

When it comes to recognizing photographs, there are a slew of image classification methods to choose from. Principal component analysis (PCA), support vector machines (SVM), facial landmark estimation, K-nearest neighbours (KNN), and logistic regression are some of the techniques available. The Local Binary Pattern (LBP) is a pattern that names an image's pixels and thresholds each pixel's neighbourhood before converting the output to a binary integer. Because of illumination and feature extraction, the success of each technique is dependent on image preprocessing. Many CNN models have already been developed, but there is still room to improve accuracy and make the algorithm mobile-friendly [24].

This work is an applied, profound learning model. Profound learning is an all around set model in the example acknowledgement space. It utilizes a Convolutional Neural Network (CNN) calculation utilizing Keras library. CNN is a particular kind of counterfeit neural organization that utilizes an AI unit. CNN applies to objects identifications, face acknowledgement, picture handling, and so on Profound convolutional neural organization (DCNN) creation of numerous neural organization layers, which is additionally can have the option to separate huge elements from the information. To manage with the images, we will be using technique of DATA AUGMENTATION. To extract the features automatically advanced neural networks like Convolutional Neural Networks are used. Table 1 presents the accuracies obtained by our experimentation

5 Conclusions and Future Work

To develop the Emotion Detection System with better precision and accuracy. Explore more advanced deep neural networks apart from Convolutional Neural Networks like Deep Convolutional Neural Networks (DCNN), Recurrent Neural Network (RNN). Hybrid versions of different trending Neural Networks can be experimented by trying

Table 1 Accuracy table

Algorithms	Accuracy in %
LeNet5	80.43
AlexNet	87.16
ResNet	86.9
VGG	77.28
FERC	78.47

different permutations and combinations. If we possible to build as a multi-platform Application. To develop the Emotion Detection System with better precision and accuracy. Explore more advanced deep neural networks apart from Convolutional Neural Networks like Deep Convolutional Neural Networks (DCNN), Recurrent Neural Network (RNN). Hybrid versions of different trending Neural Networks can be experimented by trying different permutations and combinations. If we possible to build as a multi-platform application.

References

1. Lee TS (1996) Image representation using 2D Gabor wavelets. IEEE Trans Pattern Anal Mach Intell 18(10)
2. Shen L, Bai L (2006) A review of Gabor wavelets for face recognition. Patt Anal Appl 9:273–292
3. Lisowska A, Porwik P (2004) New extended wavelet method of 2D signal decomposition based on Haar transform. Mathem Comput Simul Elsevier J
4. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24:971-987; Li SZ, Jain AK (eds) In: Handbook of face recognition. New York, Springer
5. Zhao W, Chellappa R, Phillips PJ, Rosenfeld A (2003) Face recognition: a literature survey. ACM Comput Surv 34(4):399–458
6. Ahonen T, Hadid A, Pietikainen M (2004) Face recognition with loca binary patterns. In: 8th European conference on computer vision, May 2004, pp 469–481
7. Babaii Rizvandi N, Pizurica A, Philips W (2008) 2008/01/01, SP - 539, EP-542, T1–Active appearance model (AAM)–from theory to implementation
8. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: Computer vision-ECCV 2014, Springer, pp 818–833
9. Yacoob Y, Davis L (1994) Computing spatio-temporal representations of human faces. In: Computer vision and pattern recognition. Proceedings CVPR'94, 1994 IEEE computer society conference on, IEEE, pp 70–75
10. Wang Jun, Yin Lijun (2007) Static topographic modeling for facial expression recognition and analysis. Comput Vision Image Understanding 108(1):19–34
11. Dettmers T (2015) Understanding convolution in deep learning. http://timdettmers.com/2015/03/26/convolution-deep-learning/. [Online Accessed 24 May 2016]
12. Lv J-J, Cheng C, Tian G-D, Zhou X-D, Zhou X (2016) Landmark perturbation-based data augmentation for unconstrained face recognition. Signal Process: Image Commun

13. Kotsia Irene, Buciu Ioan, Pitas Ioannis (2008) An analysis of facial expression recognition under partial facial image occlusion. Image and Vision Comput 26(7):1052–1067
14. Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis. In: Proceedings of the fourth IEEE international conference on automatic face and gesture recognition (FG'00), Grenoble, France, pp 46–53
15. Howard AG (2013) Some improvements on deep convolutional neural network based image classification. arXiv preprint arXiv:1312.5402
16. Cui X, Goel V, Kingsbury B (2015) Data augmentation for deep neural network acoustic modeling. IEEE/ACM Trans Audio Speech and Lang Proc 23(9):1469-1477
17. Coates A, Huval B, Wang T, Wu D, Catanzaro B, Andrew N (2013) Deep learning with cots hpc systems. In: Proceedings of the 30th international conference on machine learning, pp 1337–1345
18. Dugad R, Desai UB (1996) A tutorial on hidden markov models. Technical Report No. SPANN-96.1, Bombay Powai, India, May, 1996
19. Li SZ (2009) Markov random field models in computer vision. In: ICCV 1994 20. R. Rojias, AdaBoost and the super bowl of classifiers a tutorial introduction to adaptive boostring technical report
20. Li X, Wang L, Sung E (2005) A study of AdaBoost with SVM based weak learners proceedings of international joint conference on neural network
21. Daniel DM, David D, Marc S (2000) Hidden markov models for images
22. Gupta A, Gupta M (2020) Transfer learning for small and different datasets: fine-tuning a pre-trained model affects performance. J Emerg Investigators
23. Luttrell J, Zhou Z, Zhang C, Gong P, Zhang Y (2017) Facial recognition via transfer learning: fine- tuning Keras vggface. In: 2017 International conference on computational science and computational intelligence (CSCI), Las Vegas, NV, USA, pp.576–579. https://doi.org/10.1109/CSCI.2017.98
24. Shin M, Kim M, Kwon D (2016) Baseline CNN structure analysis for facial expression recognition. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN), New York, NY, pp 724–729. https://doi.org/10.1109/ROMAN.2016.7745199
25. Srinivasa Rao D, Ch RB, Sravan Kiran V, Rajasekhar N, Srinivas K et al. (2022) Plant disease classification using deep bilinear cnn. Intell Autom Soft Comput 31(1):161–176
26. Babu CR, Rao DS (2017) Comparison of discrete wavelet transform (DWT), discrete cosine transform (DCT) and stationary wavelet transform (SWT) based satellite image fusion techniques. Int J Cur Res Rev 9(49):2017
27. Srinivasa Rao D, Seetha M, Hazarath M (2012) Iterative image fusion using neuro fuzzy logic and applications. In: International conference on machine vision and image processing (MVIP), pp 121–124

Object Detection Using Deep Learning Approaches

G. Rajesh Kumar, **D. Srinivasa Rao**, **N. Rajasekhar**,
Ch. Ramesh Babu, **K. Renuka, M. Koteswara Rao, A. Revathi,**
and N. Mangathayaru

Abstract Computer vision is the field of science that studies how computers and software can recognise and understand images and scenes. Computer vision has many different parts, such as recognising images, finding objects, making images, making images bigger, and more. Object detection in real time is a big, busy, and hard area of computer vision. Object detection is when more than one object needs to be found in an image. Image localization is when there is only one object to find in an image. This finds the objects of a class that make sense in digital pictures and videos. Real-time object detection can be used for many things, such as tracking objects, video surveillance, recognising pedestrians, counting people, self-driving cars, recognising faces, following a ball in sports, and many more. As deep learning grows quickly, more powerful tools that can learn semantic, high-level, and deeper features are made available to solve problems in traditional architectures. The network architecture, training strategy, and optimization function of these models are all different. Convolution neural network is an example of a deep learning tool that can be used to find objects with OpenCV, which is a library of programming functions that are mostly used for real-time computer vision. Most accidents happen because of things that get in the way, like other people, cars, people, fire hydrants, traffic signs, and so on. This is mostly about finding these kinds of problems that can lead to disaster.

G. Rajesh Kumar (✉) · D. Srinivasa Rao · K. Renuka · M. Koteswara Rao · A. Revathi · N. Mangathayaru
Faculty of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
e-mail: gunupudirajesh@gmail.com
URL: http://vnrvjiet.ac.in/itdept/faculty/8.pdf

D. Srinivasa Rao
e-mail: srinivasarao_d@vnrvjiet.in

N. Rajasekhar
Department of IT, GRIET, Hyderabad, India
e-mail: drnrajasekhar@gmail.com

Ch. Ramesh Babu
Department of IT, MEC, Hyderabad, India

1 Introduction

Over the years, the population of the country has been rapidly increasing. Progressing population growth is demanding many requirements. Requirements include food to live, vehicles for commuting, building materials for constructions, few other industries to support mankind, and especially the Internet. Over the process of globalisation and modernisation, technology vastly underwent many changes to meet the expectations of the growing population.

Earlier travelling on foot or using certain animals like horses, bulls, and elephants was the means of transport. Later after the wheels were invented, man started to connect the wheels with axle and made carts out of it. In recent ages, when engines were developed and few other inventions like gasoline engine and gearbox were invented, they rapidly revolutionised the world to develop cars, motorcycles, trains, trucks, buses, and aircraft. Revolution was not brought in a single day. It progressed day after another. After the invention of cars, first accident led to invention of brakes to stop the vehicle, second accident led to invention of gear box to control the power transmission to the wheels, and few others led to invention of clutch, steering mechanisms, suspension systems, and reinforced chassis. On the other hand, safety systems like airbags, anti-lock braking system (ABS), and electronic brake force distribution (EBD) were developed to prevent accidents and destabilisation of vehicle.

The problem we focused more here is obstacle detection so that it could alert the driver or the vehicle in advance. Obstacles of multiple types are detected. Existing systems use devices such as ultrasonic sensors and range finder sensors to detect the proximity of an obstacle and do not even consider concave objects such as trenches and pits as obstacles. So, we have come up with a solution which uses image processing and object detection techniques to identify the obstacle and also to classify. Such simulations could be helpful in detecting objects and preventing accidents.

2 Literature Survey

Object detection is a branch of computer vision that has been widely implemented in people's lives, which is used as security monitoring, autonomous driving, person detection, and so on, with the goal of locating instances of items of a specific class. The performance of object detector models has substantially improved because of the rapid development of deep learning networks for detection tasks and visualisation tasks.

S. N. Duraimurugan, S. P. Chokkalingam, and Geethapriya are the authors of Real-Time Object Detection using YOLO. Their goal was to use the YOLO method to recognise several objects belonging to a specific class from an image [1]. Ren [2] is the author of Faster R-CNN: Towards Real-Time Object Detection. They have said

that the bases of the first-place winning entries in numerous tracks are faster R-CNN and RPN. Technology enables a unified, deep learning-based object identification system to operate at frame rates that are close to real time. The learnt RPN also enhances the quality of region proposals, which improves the total object detection model's accuracy [3]. Joseph Redmon is the author of You Only Look Once: Unified, Real-Time Object Detection. Object detection model is explained as a regression problem in his study, and the classifier is repurposed using the YOLO technique. Sandeep Kumar, Aman Balyan, and Manvi Chawla, who are the authors, worked on object detection and recognition in images. In their research, the EasyNet model was used to recognise photos and distinguish objects in photographs, such as bicycles, fruits, animals, and buildings [4]. Mohana and H. V. Ravish Aradhya published Object Detection and Classification Algorithms Using Deep Learning for Video Surveillance Applications. The classification of objects in photos and video is the subject of this paper's previous work, which employs the YOLOv2 technique. Noise, blurring, and rotating jitter, among other issues with real-world photos or images, have a significant impact on object detection application. Using You Only Look Once (YOLO), CNN, R-CNN, Fast R-CNN, and Faster R-CNN, a convolutional neural network-based techniques, where the items may be recognised in real time or real life.

Object detection is crucial in computer vision, automated vehicles (self-driving vehicles), and industrial automation, among other applications. Object detection in real time is a very difficult task. Deep learning outperforms the traditional target detection in object detection. Region proposal object detection techniques are the deep learning approaches that construct region proposal networks and subsequently classify them. SPPNet, region-based convolutional neural networks, Faster R-CNN, Fast R-CNN, and others are examples. SSD and YOLO are regression object identification techniques that produce a region suggestion network while also classifying them as objects [5].

Object detection and tracking of objects have become a critical and important component of modern technologies.

2.1 Object Tracking

Object tracking is the process of utilising a camera to monitor a moving object over time. Due to the amount of data contained in video, object tracking can be a time-consuming procedure. Splitting a video stream into many sub-frames allows for object tracking.

2.2 Object Detection

Object detection refers to the process of searching for physical things in the world, such as people, buildings, bicycles, and more. To detect an object is to locate it, so

that a bounding box may be drawn around it. This process is frequently referred to as "localising" the object. There are many applications for object detection, and as technology improves, it has attracted a lot of interest in recent years.

Most cutting-edge object detectors rely on deep learning networks as their detection network and central processing unit (CPU) to extract features from images and videos, label them, and pinpoint their locations.

2.3   Existing Datasets

In order to train and verify computer vision algorithms, datasets play a crucial role in driving research forward. Microsoft Common Objects in Context (COCO) includes 328,000 images and 91 distinct object kinds. Approximately, 2.5 million of these pictures have been tagged. The Microsoft COCO Dataset includes characteristics like as object segmentation, context-aware recognition, multiple items per image, and five descriptions per image [6].

About 79 million coloured images with a 32×32 pixel resolution can be found within the 80 million little image dataset. There is a link to the original image and some text that describes the label roughly beside each one [7].

Subsets of the CIFAR 80 million small picture dataset with labelled annotations are the CIFAR 10 and CIFAR 100 datasets. There are 60,000 colour photos with a resolution of 3232 in the CIFAR 10 dataset, split among ten item categories with 6000 images each. In total, there are 50,000 images used for training and 10,000 for testing [8].

Caltech-UCSD Birds-200-2011 (CUB-200-2011) is a newer version of the CUB-200 dataset [9], which originally included 200 different bird species. There are a total of 11,788 pictures, each of which has a single bounding box, 15 component locations, and 312 binary attributes.

The ImageNet Huge Scale Visual Recognition Challenge (ILSVRC) has been held annually since 2010 for the detection and categorization of objects in millions of images over a wide variety of item categories, much as the PASCAL VOC challenges.

Information collected for the PASCAL VOC challenge: To stay abreast of emerging technologies and to evaluate approaches to challenging problems in computer vision, the CV community regularly releases new problems for public review. Most object detection research articles utilise PASCAL VOC challenges to compare their proposed system against industry-standard datasets [9].

Analysis of these datasets has led us to the conclusion that COCO Dataset is superior to PASCAL VOC. COCO Dataset only offers 20 item categories, whereas COCO Dataset has 80, giving us additional options for how to classify objects in our work [10].

2.4 ImageNet and COCO Dataset

ImageNet has over 14 million images which is very much higher than COCO Dataset but for object detection projects which we mainly concentrate on the road obstacle does not need those many images, so COCO Dataset is best for our project.

The COCO Dataset depicted in the figure, which includes example objects such as a bottle, a sofa, a chair, a motorcycle, and a car, gives a picture for detection and categorization.

COCO Dataset was introduced by Microsoft. It is mainly used for image classification and detection. Below is the picture related to some of the object categories in COCO Dataset [10].

DWT, DCT, and SWT image fusion approaches are proposed and implemented for image fusion. Fused outcomes are assessed through various evaluation factors. Because of the basic possessions and potentiality of the SWT centred fused results outdoes DWT and DCT-based fused methods. Subsequently, SWT fused outcomes DCT variance and CV with DCT variance-based fused results upgraded fused image content [11–17].

2.5 Existing Methods

ResNet: We employ the same strategy as DSSD to better train the network model (the performance of the residual network is better than that of the VGG network). Improve your skills in order to achieve success. The initial phase in the transition was to switch from the VGG network utilised by the original SSD to ResNet. Several convolution feature layers will be added to the end of the base network [6].

Using a filtered search, Ross Girshick and his team at R-CNN were able to narrow an image down to just 2000 regions. That was referred to as a "region proposal". They avoided having to select a large number of regions in this way. You do not need to figure out how to categorise every one of those regions, instead, focus on just 2000 of them. These 2000 proposed regions are generated using the selective search algorithm described below.

1. Create the initial subdivision. Many potential regions are created by us.
2. Repeatedly merge adjacent sections using the greedy approach to create larger ones.
3. Use the proposed regions based on the created regions.

A convolutional neural network is used to take the 2000 candidate regions as input and produce a 4096-dimensional feature vector. In this role, the CNN learns to recognise patterns inside an image and uses those patterns to populate the dense layer of the output. SVM is used to determine whether or not the object is in the proposed candidate region based on the image features [7].

3   Proposed System

To overcome the false positives and true negatives and expense issue in the existing system, the proposed system is built using image processing and machine learning to detect the objects with a decent confidence score. This project adopts FR-CNN algorithm and custom object detection models to detect the possible obstacles. Even the drawback of expensiveness is covered by using image processing. FR-CNN is an object detecting algorithm which uses the concept of anchor boxes and scales down the ratio to get the outline of the object.

After techniques like non-max suppression and intersection over union are used, the final image will be shown with a bound box and a confidence score. By replacing selective search with the region proposal network, Faster R-CNN fixes the problem with selective search (RPN). First, we use ConvNet to get feature maps from the image, and then we send those maps through an RPN, which gives us object proposals. The maps are then put into groups, and the bounding boxes are guessed. Every time we talk about the idea for a new app, people do not really understand how they would use it or what it would do. A use case diagram is very helpful in this kind of scenario. The simplest way to describe a use case diagram is that it shows a system or application, how people, groups, or other systems interact with it, and how the system or application works in general [8].

As was said above, you have to upload an image. The image will then be preprocessed to find the bounding boxes of obstacles. In the next step, features and outlines of the object are extracted to put it into one of the several classes as shown in Fig. 1.

Along with above steps as shown in above Fig. 1, we need to give image dataset which consists of configuration file, data file, and the classes of the objects to be

Fig. 1 System architecture

detected. Once dataset is supplied, it undergoes the process of training and testing, and through rigorous training, objects are detected. Once road obstacles are detected, they are classified as car, person, stop sign, etc.

4 Faster Region-Based Convolutional Neural Networks (FR-CNN)

Compared with the previous R-CNN and Fast R-CNN, Faster R-CNN [5] model is more efficient as it implements the regional proposal network which utilises neural network to take care of the generating bounding box process. Figure 2 presents proposed model.

4.1 Terms in FR-CNN

IoU: Crossing over intersection union is a metric for evaluating how effectively you can locate a matching object inside a given dataset. Numerous object detection competitions showcase the standard's versatility. The majority of our work with this technique is dedicated to evaluating its efficacy in convolutional neural network detectors. It is easy to use IoU as a unit of measurement. The IoU formula equals the result obtained by dividing the overlapping region by the combined region size. Any IoU value above 0.5 is considered satisfactory.
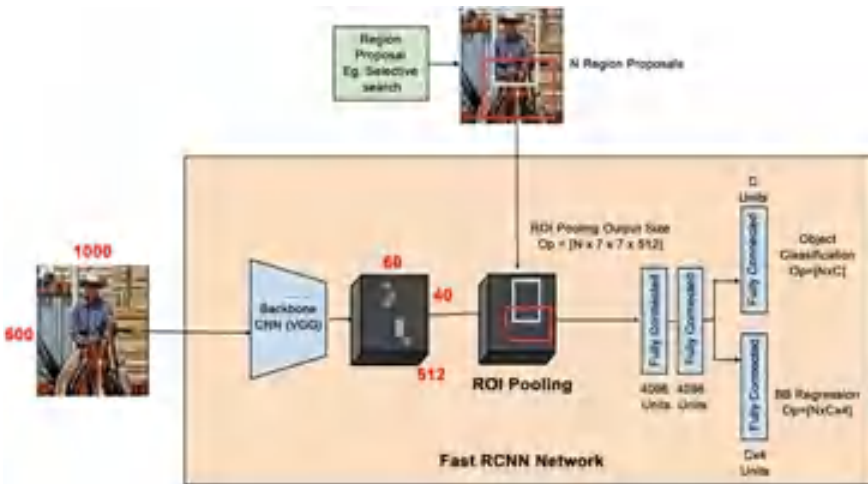
Fig. 2 Proposed model for FR-CNN
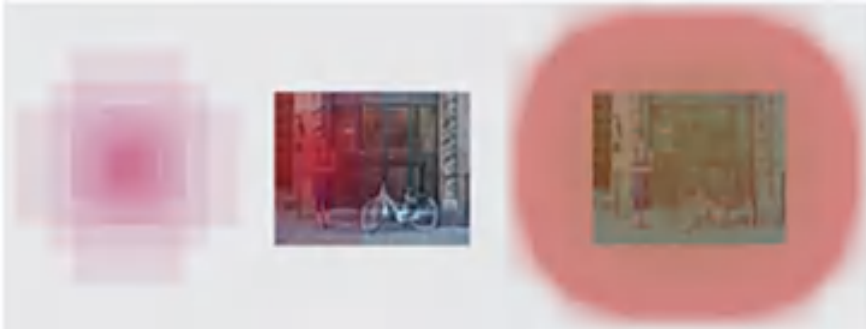
Fig. 3 Anchor based on sliding window method

$$IOU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{1}$$

Anchor: In order to consider all of the potential region ideas, Faster R-CNN employs a sliding window method dubbed the anchor-based method. The sliding windows have been relocated from the image to a feature map presented in Fig. 3.

An anchor with two characteristics, scale and aspect ratio, is positioned in the centre of the sliding window. Three different sizes (8, 16, 32) and three different ratios (0.5, 1, 2) at each anchor point allow us to construct a feasible bounding box. There are W H anchors and 33 bounding boxes for each anchor in a W H feature map 9 W H.

4.2 Model

These four components make up the bulk of Faster R-CNN:

1. Faster R-CNN, a CNN network target detection approach, begins with a foundation of standard Conv + ReLU + pooling layers to derive feature maps from input images. a. Conv layers following RPN layers also make use of the feature maps, as do fully connected layers.
2. To propose proposals for regions, you can use the region proposal network (RPN). The anchor positions are determined by the layer using softmax. Anchor errors are fixed via bounding box regression, allowing for more precise proposal generation.
3. Return on investment (ROI) pooling: This layer receives the feature maps and proposals, mixes the data, selects the appropriate feature maps, and passes them on to the next fully linked layer where the desired category can be located.
4. Finally, the proposal's feature maps are used to determine its classification, and, with the use of bounding boxes, the precise location of the detection frame is determined.

Each successive convolution layer abstracts the information from the preceding layer. Typically, the first layer acquires knowledge about edge edges, and the second layer learns pattern patterns in edge edges in order to comprehend more complicated forms and data. The convolutional features are the last to be discovered. A lot of the room has been taken out, but the depth has been increased. The feature map's dimensions are shrunk using a pooling layer placed between convolutional ones. The convolutional layer's ability to learn more filters increases the map's depth.

The conventional detection method generates a very time-consuming detection frame. For instance, R-CNN generates a detection frame using the selective search technique. Faster R-CNN does not use the tried-and-true SS technique or sliding window. It is the RPN itself that is employed to construct the detection frame. The Faster R-CNN has this advantage as well, because it can reduce the time required to create detection frames. It turns out that the RPN network is actually made up of two lines: one for the foreground and background by the softmax classification anchors, and another for calculating the bounding box offset for the anchors in order to acquire the precise proposal. Proposals are combined with foreground anchors and bounding box traction offsets by the final proposal layer, which is also responsible for discarding proposals that are too tiny or fall outside the bounds. In this case, the equivalent of the target positioning function has been completed by the entire network up to the proposal layer.

4.3 Post-Processing

It is a lot like RPN in that we are left with a bunch of classed objects that need some further attention before we can return them. Considering the most probable class of ideas is necessary for updating the bounding box. The "ignore likelihood" is the highest for the background class proposal. Class-based NMS is utilised during final object discovery, when the expected background is disregarded. The items are first sorted by class and then by likelihood before being processed through NMS and finally being combined. Moreover, we can restrict the total number of objects in each category and establish a minimum probability requirement for the finalised list.

4.4 Training

The authors of the research report that they trained Faster R-CNN in stages, first individually and then combining the learned weights. It was later discovered that end-to-end training is more effective when carried out in tandem. To use the whole model, you will need to choose between four distinct losses: two for RPN and two for R-CNN.

The training method for the base network is determined by the resources at hand and the information that has to be absorbed by the network. It is not necessary to

retrain the base basic network if the new data is very close to the original dataset used during training. In order for the network to be able to calculate the gradient, significant time and resources must be invested in its training. Four distinct losses are added together using a weighted sum to illustrate the full picture. We can adjust the relative importance of the classification loss and the regression loss or the weights used by R-CNNs and RPNs. Similar to the detection network, Conv layers can be used to extract feature maps.

5   Conclusions and Future Work

We use Faster R-CNN to find objects in this project. Faster R-CNN is a convolutional neural networks architecture for spotting objects that works with region proposal network (RPN). RPN takes a feature map as input and makes a list of possible outcomes, along with a score for how likely each one is to happen. We have used the COCO Dataset to find things. The proposed method should be good at finding things on the road.

References

1. Bourdev LD, Malik J (2009) Poselets: body part detectors trained using 3D human pose annotations. IEEE 12th international conference on computer vision, ICCV 2009, Kyoto, Japan, September 27-October 4, 2009. IEEE, Kyoto, pp 1365–1372
2. Cadena C, Dick A, Reid I (2015) A fast, modular scene understanding sys-tem using context-aware object detection. In: 2015 IEEE international conference on robotics and automation (ICRA), Seattle, WA
3. Correa M, Hermosilla G, Verschae R, Ruiz-del-Solar J (2012) Humandetection and identification by robots using thermal and visual information indomestic environments. J Intell Robot Syst 66:223–243. https://doi.org/10.1007/s10846-011-9612-2
4. Dalal N, Triggs B (2005) Histograms of oriented gradients for humandetection. In: IEEE computer society conference on computer vision and pattern recognition. CVPR 2005, vol 1, pp 886–893. IEEE, San Diego, CA. https://doi.org/10.1109/CVPR.2005.177
5. Erhan D, Szegedy C, Toshev A, Anguelov D (2014) Scalable object detec-tion using deep neural networks. In: Computer vision and pattern recognition frontiers in robotics and AI. www.frontiersin.org
6. Shrivastava A, Gupta A, Girshick R (2016) Training region-based object detectors with online hard example mining. CVPR
7. Lin T-Y, Dollar P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. CVPR
8. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition. CVPR 2005, vol 1, pp 886–893. IEEE, San Diego, CA. https://doi.org/10.1109/CVPR.2005.177
9. Erhan D, Szegedy C, Toshev A, Anguelov D (2014) Scalable object detection using deep neural networks. In: Computer vision and pattern recognition frontiers in robotics and AI. www.frontiersin.org
10. Girshick R, He K, Zhang X, Ren S, Sun J (2016) Fast R-CNN. ICCV 2015. Deep residual learning for image recognition. CVPR

11. Srinivasa Rao D, Babu ChR, Sravan Kiran V, Rajasekhar N, Srinivas K et al (2022) Plant disease classification using deep bilinear cnn. Intell Autom Soft Comput 31(1):161–176
12. Babu CR, Rao DS (2017) Comparison of discrete wavelet transform (DWT), discrete cosine transform (DCT) and stationary wavelet transform (SWT) based satellite image fusion techniques. Int J Cur Res Rev 9:49
13. Srinivasa Rao D, Seetha M, Hazarath M (2012) Iterative image fusion using neuro fuzzy logic and applications. In: International conference on machine vision and image processing (MVIP), pp 121–124
14. Rajesh Kumar G, Gugulothu N, Nimmala M (2019) An evolutionary feature clustering approach for anomaly detection using improved fuzzy membership function: feature clustering approach for anomaly detection. Int J Inform Technol Web Eng (IJITWE) IGI Glob 14(4):19–49
15. Radhakrishna V, Rajesh Kumar G, Aljawarneh S (2015) Strategic application of software process model to optimize business intelligence results. Proceedings of the the international conference on engineering & MIS 2015:1–6
16. Mangathayaru N, Rajesh Kumar G, Narsimha G (2016) Text mining based approach for intrusion detection. In: International conference on engineering & MIS (ICEMIS), pp 1–5. IEEE
17. Aljawarneh S, Radhakrishna V, Rajesh Kumar G (2019) A recent survey on challenges in security and privacy in internet of things. In: Proceedings of the 5th international conference on engineering and MIS, pp 1–9

Telugu Tweets Sentiment Analysis Based on Ordinal Regression

G. Balakrishna Priya and M. Usha Rani

Abstract Nowadays, twitter is a platform where users can share their opinion on specific topic that occurring in the society. To analyze their opinions, sentiment analysis is a technique to exact the user's sentiment on a particular topic. Much research has been done on machine learning and deep learning algorithms for analyzing the tweets. In world, human tweets are in different languages like Persian, Urdu, Hindi, Telugu, and Tamil. The current work is focused on analysis of Telugu language tweets. Here the comprehensive dataset that consists of Telugu tweets has been prepared for sentiment analysis by using deep learning approaches. For estimating the sentiment polarity and emotions, the deep long short-term memory modules were used and the findings show the best results in terms of accuracy.

Keywords Twitter · Sentiment analysis · Ordinal regression · Telugu · Deep learning

1 Introduction

After the development of the microblogging websites, the people got a source to share their opinion on the particular topics, and one of the major microblogging is the twitter. It is a major platform where the researchers will get huge data and can have a more scope to analyze the sentiments of a people on particular topic [1–3]. The twitter is a platform to do the market analysis; here, the customer will post the opinion on the particular product or any event, and with this, the selling or organizing companies will get the idea of the customers of the product or any relevant item, in terms of the negative, positive, and neutral opinion [4, 5].

The sentiment analysis is one of the major applications of the natural language processing, and in India, it has major scope as in India many regional languages exist.

G. Balakrishna Priya (✉) · M. Usha Rani
Department of Computer Science, Sri Padmavathi Mahila Viswavidyalayam, Tirupathi, Andhra Pradesh, India
e-mail: gbkrishnapriya18@yahoo.com

It will able to know the behavior of the people on particular topic/incident/event/any product. Nowadays, the sentiment analysis work is going in different languages in India like Tamil, Malayalam, Bengali, and Urdu. This current work is focusing on the Telugu language. In India the Telugu has 16th rank among all languages in the category of most speaking language in worldwide [8].

The current work is on the analysis of the sentiment on the Telugu tweets by using deep learning algorithm deep long short-term memory modules. The remainder of the next sections are in Sect. 2 discussed the literature survey, in Sect. 3 proposed system, in Sect. 4 the implementation, in Sect. 5 the results and discussion, and in Sec. 6 the conclusion and future work.

2 Literature Survey

In [6], the authors have done work about the tweets in the Arabic language for the sentiment analysis. Here instead of the natural neural word embeddings, sentiment-specific word embeddings of the Arabic language were used, and the BERT has been used for the representation of the Arabic language. The results were shown efficiency in terms of the F-Score.

In [8], the authors have proposed SentiPhaseNet algorithm for sentiment analysis of the Telugu tweets, this system will overcome the problems in the SWNet algorithm, this has been compared with the other machine learning algorithms, and this has shown effectiveness in terms of the performance.

In [9], the work is on comparative analysis of the advanced learning ensemble methods and identified the gaps to improvise the algorithms and proposed a sentiment analysis technique for the analysis of the Telugu tweets. The results were shown good in terms of the precision and F-Score.

In [10], the authors have done sentiment analysis on the Bengali language with the convolution neural networks algorithm and has been classified the tweets as positive, negative, and neutral. The work can be applied for other languages as well in India.

3 Proposed System

Here for analyzing and doing the sentiment analysis for the Telugu tweets the framework (TWLD) [7] that is specifically designed for Telugu tweets sentiment analysis is used. The framework is shown in Fig. 1. The algorithm is shown in Algorithm 1.

In the current system, the deep long short-term memory deep learning algorithm is used for estimating the sentiment analysis on the topic of the pandemic situation occurred by the COVID-19. Here in the first step, the tweets are collected and the preprocessing has been done. The twitters related to the COVID-19 are collected with the twitter API and in the preprocessing step, the noise data, misspellings, duplicates, and related to the unwanted data are processed and then applied the word embedded
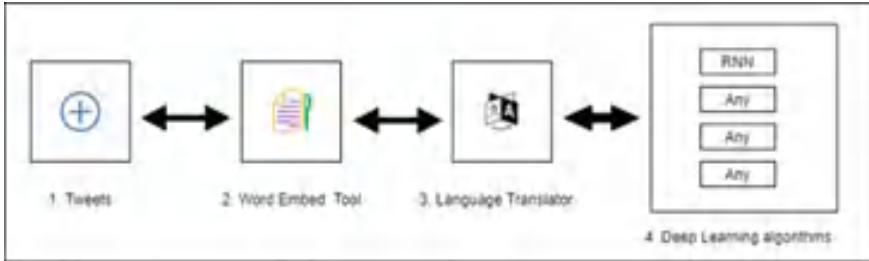
Fig. 1 Framework for sentiment analysis of the Telugu tweets

Table 1 Dataset summary

Dataset	Positive	Negative	Total
Training	4300	4300	8600
Testing	1800	1800	3600

tool for getting the meaning of the tweets and then with the language translators that are converted into the English language and finally applied the deep learning algorithm for the sentiment analysis. The used dataset is shown in Table 1. The implementation part and the results and discussion were done in the next sections.

Algorithm 1 Sentiment Analysis of the Telugu Tweets
Begin
    Procedure collection of tweets:
      By using the twitter API
    End Procedure
    Procedure Pre- processing:
      Remove duplicates
      Remove noise data
      Remove unwanted data
    End Procedure
    Procedure Word Embed Tool
      Tokenization
      Tagging
    End Procedure
    Procedure Language Translator
      Convert to English
    End Procedure
    Procedure Sentiment analysis
      Feature extractions
      Training
      Testing
    End Procedure
  End Until

4 Implementation

The proposed system is implemented with the dataset shown in Table 1 the Natural
Language Toolkit (NLTK) is used for tokenizing and tagging the data. With the help
of the Python programming language in the Anaconda framework, the current work
implements and the obtained results were shown and discussed in the next section.

5 Results and Discussion

After implementing the proposed system on the dataset, the obtained results were
shown in Table 2. The results were shown in terms of the precision, recall, F1-Score,
and the support. The tweets are based on the training dataset and by the deep learning
algorithm the testing data shows the results efficiently (Fig. 2).

Table 2 Results

	Precision	Recall	F1-Score	Support
Positive	0.62	0.76	0.68	435
Negative	0.67	0.67	0.67	456
Neutral	0.00	0.00	0.00	122
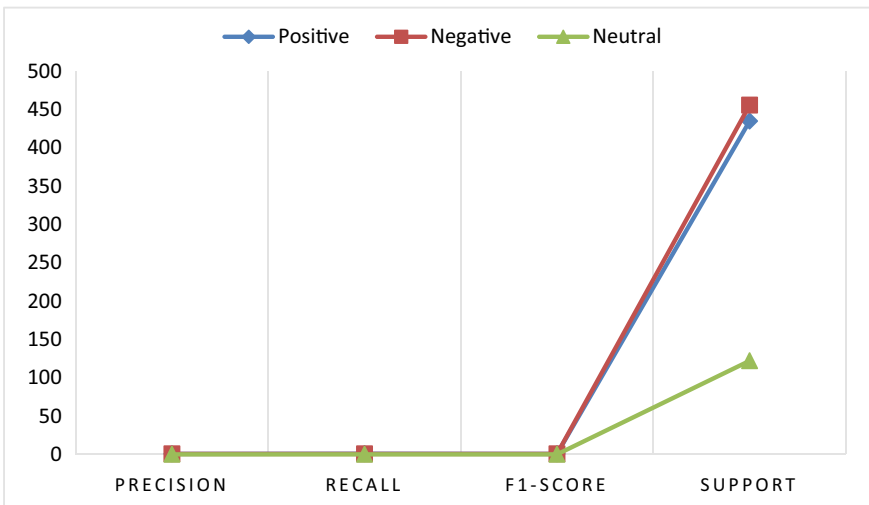
Fig. 2 Results of proposed system

6 Conclusion

The deep learning is an emerging technology and has applications in almost all the fields, and this work is on sentiment analysis on the COVID-19; here the deep learning algorithm deep long short-term memory is used along with the framework that is specifically designed to do sentiment analysis on the Telugu tweets. The results were shown effectiveness in terms of the F-Score. This can be improvised with the help of the any other deep learning approach. In the future work by using this framework, work can be carried by any other deep learning algorithms.

References

1. Saad SE, Yang J (2019) Twitter sentiment analysis based on ordinal regression. IEEE Access 7:163677–163685
2. O'Connor B, Balasubramanyan R, Routledge BR, Smith NA (2010) From tweets to polls: linking text sentiment to public opinion time series. Proc ICWSM 11(122–129):1–2
3. M. A. Cabanlit and K. J. Espinosa, "Optimizing N-gram based text feature selection in sentiment analysis for commercial products in Twitter through polarity lexicons," in Proc. 5th Int. Conf. Inf., Intell., Syst. Appl. (IISA), Jul. 2014, pp. 94–97.
4. S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in Proc. 20th Int. Conf. Comput. Linguistics, Aug. 2004, p. 1367.
5. Whitelaw C, Garg N, Argamon S (2005) Using appraisal groups for sentiment analysis. In: Proceedings of the 14th ACM Int. Conf. Inf. Knowl. Manage., Oct/Nov 2005, pp 625–631
6. Al-Twairesh N, Al-Negheimish H (2019) Surface and deep features ensemble for sentiment analysis of Arabic tweets. IEEE Access 7:84122–84131
7. Priya, G. Balakrishna, Usha Rani M (2020) A framework for sentiment analysis of Telugu tweets. Int J Eng Adv Technol (IJEAT) 9(6)
8. Naidu R, Bharti SK, Babu KS (2018) Building SentiPhraseNet for Sentiment Analysis in Telugu. In: 2018 15th IEEE India Council international conference (INDICON). IEEE
9. Boddupalli S et al (2019) Sentiment analysis of Telugu data and comparing advanced ensemble techniques using different text processing methods. In: 2019 5th International conference on computing, communication, control and automation (ICCUBEA). IEEE
10. Shalini K et al (2018) Sentiment analysis of Indian languages using convolutional neural networks. In: 2018 International conference on computer communication and informatics (ICCCI). IEEE

Image Style Transferred to Graphical User Interfaces

Karim Hammoudi, Adnane Cabani, Halim Benhabiles, and Mahmoud Melkemi

Abstract This paper presents an approach for permitting the restyling of an application by automatic analysis of a considered image. The approach relies on the analysis of color distribution from a query image and a color-to-component mapping in order to perform its style transfer to the GUI. No deep learning technique is used, e.g., making the approach applicable without investigating image style learning through training stages. We show how our approach can (i) be used to directly and dynamically restyle GUI of applications, (ii) be adopted as an end user functionality toward the restyling of mobile app, desktop app, web app, and so on, (iii) serve to developers, e.g., for facilitating the selection of appropriate reference graphic charter during the design stage of their applications by exploring GUI appearances from a list of considered query images. Experimental results show the efficiency of the approach and its high potential of generalization.

Keywords Image style transfer · Visual computing · GUI · Graphic charter · Design pattern · UX · HCI

K. Hammoudi (✉) · M. Melkemi
Department of Computer Science, IRIMAS, Universiéde Haute-Alsace, Mulhouse F-68100, France
e-mail: mahmoud.melkemi@uha.fr

Université de Strasbourg, Strasbourg, France
e-mail: karim.hammoudi@uha.fr

A. Cabani
Normandie Univ, UNIROUEN, ESIGELEC, IRSEEM, Rouen 76000, France
e-mail: adnane.cabani@esigelec.fr

H. Benhabiles
JUNIA, CNRS-IEMN, Universite de Lille, Lille F-59000, France
e-mail: halim.benhabiles@junia.com

1 Introduction and Motivation

Nowadays, the graphical user interface of majority of applications deployed over the world is static in the sense that the default designs cannot be changed or can sometimes be modified according to a predefined list of design templates. In this context, we propose a restyling approach which exploits an input image of our choice (already stored image, taken photo) for dynamically colorizing a graphical user interface.

The style transfer is a particularly active topic which is more and more investigated for automatically transferring style characteristics of a data [1–3] or a drawing technique [4–6] to another one toward facilitating styling tasks for a large spectrum of applications.

To the past, e.g., in [7], efficient statistical approaches have been developed to successfully transfer color characteristics from one image to another one.

In [8], authors propose an approach which can change the appearance of photos to a different time of day. This approach exploits a database of time-lapse videos. By this way, a photo taken in daylight can be transformed into a realistic photo taken at night.

Recently, authors of [9] proposed an approach for permitting a photorealistic style transfer while limiting spatial distortions or unrealistic artifacts which should not happen in real photographs. The key ingredient of their approach is the use of wavelet transforms that naturally fits in deep networks. In [10], authors propose an approach which permits to insert the image of an object in the image of a painting and to automatically harmonize the overall. To this end, the style of the painting is transferred to the image of the inserted object making the resulting global composition appearing as it was originally painted. Besides, the deep learning approach which is presented in [11] can transfer in high quality the style of a cartoon face, a caricature face, or an anime face to a real face picture. In [12], authors propose a framework that permits a style transfer without applying a direct use of an image, but only with a text description of the desired style. By this way, associating the term "fire" to an image permits to apply on a fire style on this latter.

Most of the recent style transfer approaches currently rely on the use of neural-related techniques but not all [13]; e.g., in reason of applicative contexts for which data training stages, data-to-data matching stages or deep architecture parametrization cannot be easily operated.

To the best of our knowledge, very few approaches have been proposed for transferring the style of an image to a graphical user interface. In [14], a deep learning restyling approach is presented to globally transfer the style of an image to the image of a considered graphical user interface. The obtained results are very interesting and promising. However, it seems that their quality of results is related to a well-choice of the images. On the other side, suggested training stages could be time-consuming which may be fastidious for the users. In the next section, we propose an approach which could be less time-consuming since no machine learning technique is used. On the other hand, our method evaluates how aesthetic and usable could be the interface restyled from a query image.

2 Proposed Approach

Figure 1 shows the global dataflow diagram of the color-based approach which is proposed for restyling the graphical user interface of an application from a query image. This restyling workflow is applied to the GUI of a mobile application, but other types of application such as web application or desktop application can be considered. In our process, the proposed application, which implements our approach, includes functionalities for loading a stored image or for taking a picture from a camera for its exploitation toward restyling the graphical user interface. Once the image loaded or taken, dominant colors are identified. To this end, a histogram of color distribution is computed, and a classification of color intensities is operated [15] with respect to the number of component category which is present in the considered graphical user interface (e.g., text, buttons, background). Then, the default color of graphical components is replaced by the identified dominant colors of the considered image. Additionally to the transfer of color, the analyzed image can also be used for texturing the background of the application according to the expectation of the user. The output shows an application which is restyled with the appearance of the considered image.

More specifically, the modification of default colors associated with the graphical components is performed from an image by previously producing a component adjacency graph. Indeed, considering the graphical user interface of an application (e.g., Fig. 2a), the corresponding blueprint mode is activated. This permits to obtain a simplified and wireframe view of the GUI (see Fig. 2b). The blueprint mode provides a wireframe and complementary view of the GUI which permits to focus on components of the design without the distraction of content.

Both conventional design and blueprint views are exploited in order to define an adjacency graph (see Fig. 2c) where each region corresponds to a node [16]. This latter graph can then be represented as a component adjacency graph such as shown in Fig. 2d. We can observe that the text of the header (font) is connected to the filled header. This latter is connected to the background which is connected to the border of buttons with a single edge as well as to the central image. Since this latter corresponds to a loaded picture which is static in this design, this component is not included in
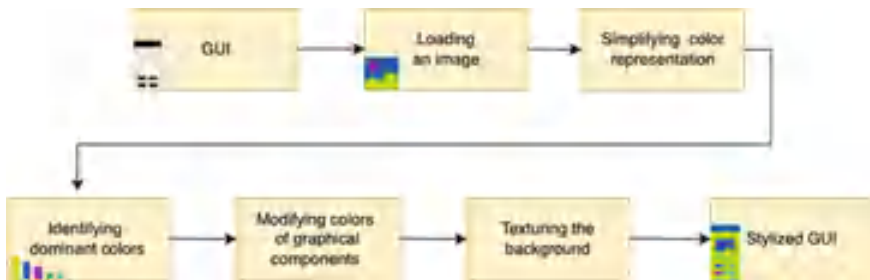
Fig. 1 Global dataflow diagram of the proposed approach for dynamically restyling a GUI of an application from an image

(a) Default application   (b) Blueprint view   (c) Region adjacency graph

(d) Component adjacency graph
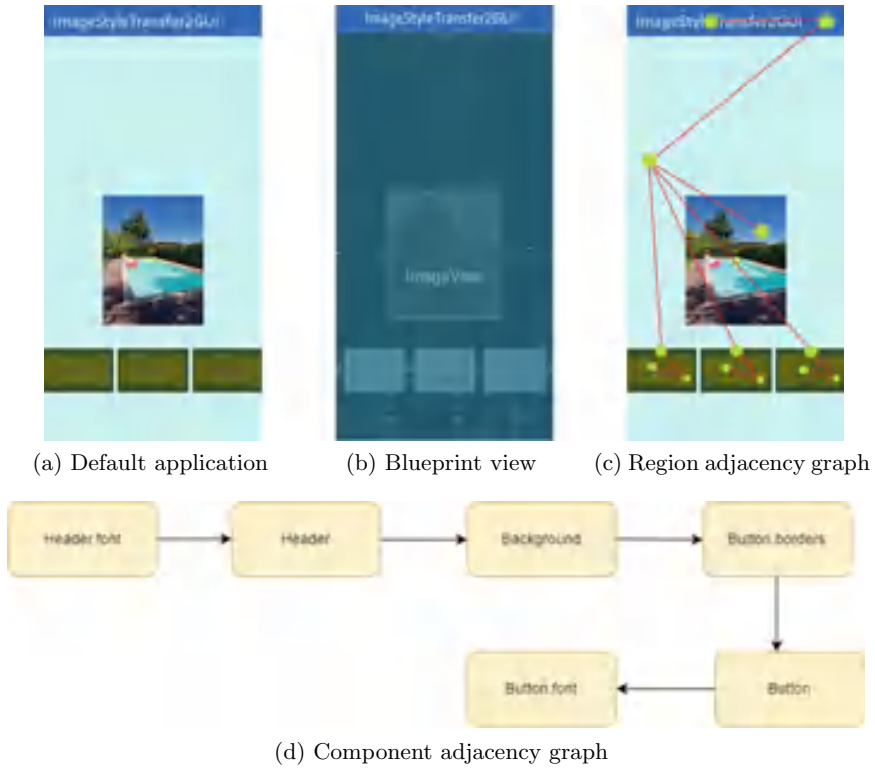
Fig. 2 Workflow applied to produce a component adjacency graph

the graph. Button borders are then connected to filled buttons which are themselves connected to the button text.

More precisely, this defined component adjacency graph is then used as a fixed data structure toward simultaneously evaluating the aesthetic quality and the usability of the restyled graphical user interface. More precisely, a metric has been defined in order to globally measure the contrast in between components of the defined adjacency graph. It is assumed that a high contrast value will result in a visually pleasant GUI with a satisfactory level in component usability.

The components of the defined adjacency graph are encapsulated into one ordered vector $\mathbf{v}$.

$$\mathbf{v} = (C_1, ..., C_n)^T, \tag{1}$$

where $C_{i_{(1 \leq i \leq n)}}$ corresponds to there retained adjacency components ($n \geq 2$). The constrast ratio is defined as follows:

$$r = \frac{\sum_{j=1}^{m} \Delta(C_j, C_{j+1})}{m \times \Delta_{\max}}, \qquad (2)$$

where m represents the number of edges in between components; i.e, $m = n - 1$. $\Delta(C_j, C_{j+1})$ corresponds to the Euclidean distance in between dominant color values which are respectively assigned to the components C_j and C_{j+1}. $\Delta_{\max}$ corresponds to the highest contrasting value being which can be obtained, e.g., when adjacent components are black and white, respectively. In this case, the contrast ratio reaches its largest value $r = 1$.

The number of identified dominant colors is n since it corresponds to the number of adjacency components (i.e., number of components of **v**). These retained dominant colors are stored in a vector $\mathbf{w} = (col_1, ..., col_n)$. Therefore, the largest contrast $\mathbf{r}^\star$ can be reached by determining the optimal color distribution $\mathbf{w}^\star$. Specifically, we seek a permutation σ such that $\mathbf{w}^\star = (col_{\sigma(1)}, ..., col_{\sigma(n)})$ maximizes the contrast ratio r. The maximum is reached after at most n! steps ($n!$ is the number of permutations of $\{1, ...n\}$). The computation of $\mathbf{w}^\star$ is not time-consuming when n is not large as it is the case in our application. Otherwise, the search space can be reduced by randomly selecting a fixed number of permutation. From these retained permutations, we keep the one which provides the best contrast ratio.

$$\mathbf{r}^\star = arg \max_{\mathbf{w}} c \qquad (3)$$

3 Experimental Results and Evaluation

Figure 3 displays a dynamic restyling of a default application by successively taking selfie photos in the context of try-on of clothes. This study case can be seen as a simulation of a virtual clothes try-on system for which the wearing of a virtual garment accordingly involves the instantaneous harmonization of the GUI.

We can observe that dominants colors from image objects have been transferred to the graphical components of the interface. In particular, we remark that the photo backgrounds have colored the backgrounds of the application in Fig. (3a, c). In Fig. 3a, buttons seem colored from the shirt color, button borders seem colored from the skin color. In Fig. 3b, buttons seem colored from the skin color, the header seem colored from the T-shirt color. In Fig. 3c, buttons seem colored from the T-shirt color, button borders seem colored from the skin color. Visually, outputs of user tests 1 and 2 can be satisfying. However, the output of the user test 3 is not satisfying in reason of the color of the buttons and their inner texts which is relatively close. This makes the corresponding texts unreadable and this application unusable.

Figure 4 shows the application styled by loading images of different natures. Quality of the applications styled from the loaded images is displayed as a score ranging in [0,1] which has been obtained by computing the contrast metric c. Fig. 4a, f are ordered by their score from lowest to highest style transfer quality.

(a) User test 1 (b) User test 2 (c) User test 3
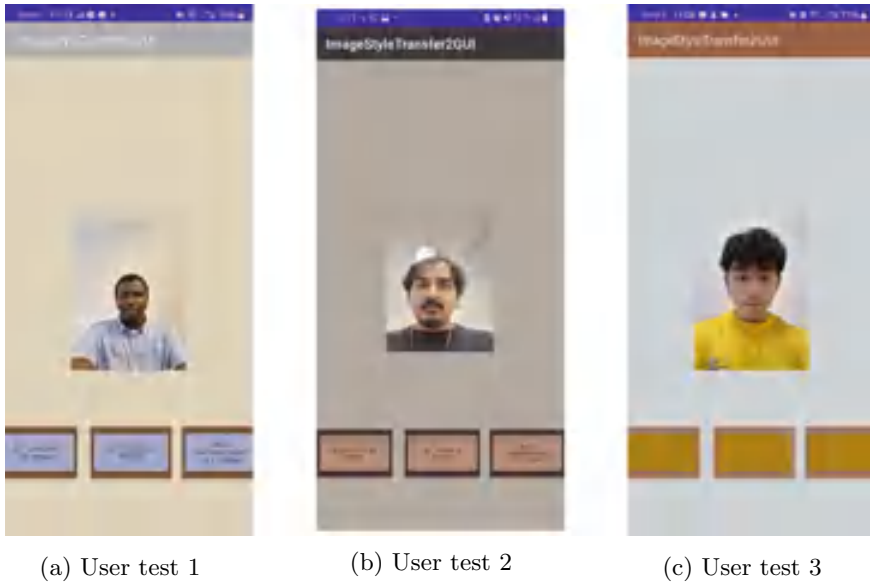
Fig. 3 Application dynamically styled by taking successive selfie photos in the context of try-on of clothes

Figure 4a displays the loading of a kind of chessboard image for which a single dominant color has voluntarily been retained. The resulting design of the application is homogeneous making the background, button borders, buttons, and writings appearing as a whole. The obtained null score value corresponds to an application which is neither aesthetic nor usable.

Figure 4b and c show outdoor photos. In Fig. 4b, we can observe that the color of the vegetation seems reported to the background, button borders, and writings. The color of the inner buttons seems to come from the background. The color of cars, although visually consistent, is not used for coloring components since not dominant. The quality of the image transfer is low (score value 0.21) in reason of the low contrast in between components related to the buttons which also makes the interface complex to use. In Fig. 4c, the sky color is reported to the header, the remaining components got close colors from the bricks of the building facades. Quality of the styled application is similar to the one of the Fig. 4b. No color optimization processes have been operated to Figs. 4a–c.

Figure 4d depicts an indoor photo. The colors, although relatively limited in term of variety, are distributed to components with a satisfactory way. We observe a high contrast, distinguishable components, readable writings, and usable button functionality. We also observe the same results by using artistic images which have a variety of colors, e.g., see Figs. 4e and f. It seems that the quality of the styled application is satisfying in term of aesthetic (graphical coherency) and usage when the contrast score is superior or equal to 0.44.
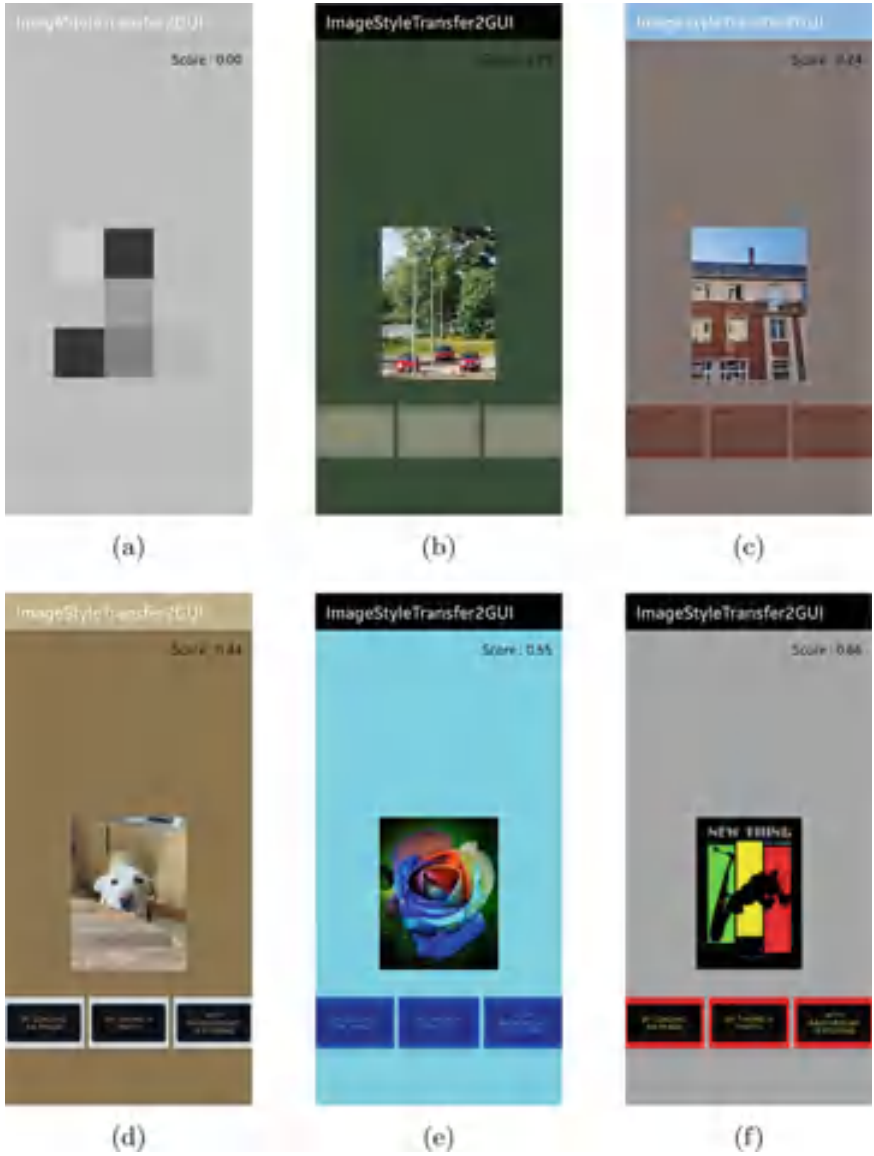
Fig. 4 Application styled by loading images of different natures. Quality of the applications styled from the loaded images is displayed as a score ranging in [0,1]

4   Conclusion

An approach has been proposed for restyling the graphical user interface of applications from an image. Such approach can be integrated as a complementary tool for application development cycles in order to facilitate the creation of designs. Such approach can also be included in a large spectrum of applications toward permitting to end users the self-changing of the design through an image loading or taking functionality. The transfer of colors is performed without invoking neural analysis techniques which can be time-consuming in training stages. The employed graph design and optimization process provide a direct and promising solution for restyling applications from query images.

References

1. Grinstein E, Duong NQK, Ozerov A, Prez P (2018) Audio style transfer. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 586–590
2. Yin K, Gao J, Shugrina M, Khamis S, Fidler S (2021) 3dstylenet: creating 3d shapes with geometric and texture style variations. In: Proceedings of international conference on computer vision (ICCV)
3. Cygert S, Czyzewski A (2019) Style transfer for detecting vehicles with thermal camera. In: 2019 Signal processing: algorithms, architectures, arrangements, and applications (SPA), pp 218–222
4. Texler O, Fišer J, Lukáč M, Lu J, Shechtman E, Sýkora D (2019) Enhancing neural style transfer using patch-based synthesis. In: Proceedings of the 8th ACM/eurographics expressive symposium on computational aesthetics and sketch based interfaces and modeling and non-photorealistic animation and rendering. Expressive '19, (Goslar, DEU), p 4350, Eurographics Association
5. Kotovenko D, Wright M, Heimbrecht A, Ommer B (2021) Rethinking style transfer: From pixels to parameterized brushstrokes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 12196–12205
6. Liu X, Wu W, Wu H, Wen Z (2021) Deep style transfer for line drawings. Proceedings of the AAAI conference on artificial intelligence 35:353–361 May
7. Reinhard E, Adhikhmin M, Gooch B, Shirley P (2001) Color transfer between images. IEEE Comput Graph Appl 21(5):34–41
8. Shih Y, Paris S, Durand F, Freeman W (2013) Data-driven hallucination of different times of day from a single outdoor photo. ACM Trans Graphics (TOG) 32(11)
9. Yoo J, Uh Y, Chun S, Kang B, Ha J-W (2019) Photorealistic style transfer via wavelet transforms. In: 2019 IEEE/CVF international conference on computer vision (ICCV), pp 9035–9044
10. Luan F, Paris S, Shechtman E, Bala K (2018) Deep painterly harmonization. arXiv preprint. arXiv:1804.03189

11. Yang S, Jiang L, Liu Z, Loy CC (2022) Pastiche master: Exemplar-based high-resolution portrait style transfer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 7693–7702
12. Kwon G, Ye JC (2022) Clipstyler: image style transfer with a single text condition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 18062–18071
13. Jing Y, Yang Y, Feng Z, Ye J, Yu Y, Song M (2020) Neural style transfer: a review. IEEE Trans Vis Comput Graph 26:3365–3385
14. Fischer MH, Yang RR, Lam MS (2020) Imaginenet: restyling apps using neural style transfer
15. Zheng Q, Lu M, Wu S, Hu R, Lanir J, Huang H (2022) Image-guided color mapping for categorical data visualization. Comput Vis Media
16. Tremeau A, Colantoni P (2000) Regions adjacency graph applied to color image segmentation. IEEE Trans Image Process 9(4):735–744

Author Index